HK J Paediatr (new series) 2026;**31**:26-35

Original Article

The Reliability of Neuromotor Assessment Using Test of Infant Motor Performance Screening Items in Infants

J **YOON**, DC **SHIN**

Abstract ***Purpose:*** To examine the Test-retest and Inter-rater reliability of the Test of Infant Motor Performance Screening Items (TIMPSI) for identifying infants who require comprehensive Test of Infant Motor Performance (TIMP) assessment. ***Methods:*** Sixteen infants were evaluated using the TIMPSI by three experienced physical therapists. Test-retest reliability was assessed by comparing the first and second assessments, while Inter-rater reliability was determined by comparing results from different testers. Video recordings were used to mitigate stress among the infants. ***Results:*** The study found that TIMPSI significantly reduced the assessment duration compared to TIMP and demonstrated excellent reliability in both Test-retest and Inter-rater evaluations. ***Conclusions:*** TIMPSI, with its brief duration and reduced questionnaire, effectively alleviates the physical burden on infants. Its high reliability supports its use in early identification of motor development issues, potentially mitigating long-term neurodevelopmental damage in high-risk infants.

Key words Developmental disabilities; Early diagnosis; Premature birth

Introduction

Neonatal intensive care advancements have significantly enhanced the survival rates of high-risk newborns, such as premature and extremely underweight infants. With a global incidence of 15 million cases, preterm birth stands as a primary contributor to neonatal mortality and morbidity, imposing a substantial social and economic burden.[1] Prematurely born infants face a heightened risk of lifelong neurological disorders due to brain immaturity.[2]

Department of Physical Therapy, Graduate School of Kyungnam University, Republic of Korea
J **YOON** *PT, MSc*

Department of Physical Therapy, College of Health and Welfare, Sahmyook University, Republic of Korea
DC **SHIN** *PT, PhD*

Correspondence to: Dr DC **SHIN**
Email: icandox@syu.ac.kr

Received October 22, 2024

As they progress in age, approximately 50% of extreme preterm infants experience neurological deficits, encompassing issues like motor coordination, cognitive impairment, and attention deficits.[3] Despite the improved survival rates, long-term neurodevelopmental prognosis has not seen commensurate advancements.[4] Additionally, the increasing population of children with developmental motor disorders, ranging from developmental delays to cerebral palsy, underscores the imperative for early identification and intervention.[5]

As survival rates rise, the prevalence of developmental disorders increases, especially among high-risk infants. This underscores the heightened risk of brain damage and abnormal development. External factors, such as the neonatal intensive care environment, can further exacerbate these challenges. Factors like reduced antioxidant capacity in early infants and exposure to ventilator therapy-generated active oxygen may contribute to heightened risks of brain damage.[6] Additionally, the impact of environmental stimuli in the neonatal intensive care unit, such as loud sounds and bright lights, is known to influence neuropathic development in newborns, with decreased auditory stimuli notably affecting speech and language development.[7]

Given these complexities, early identification of infants at risk for neuromotor developmental disorders is crucial for timely intervention.[8-10] The American Pediatric Association recommends screening high-risk infants for neurodevelopmental disorders before the age of 9 months to enable prompt therapeutic interventions and maximise neuroplasticity.[9]

The International Clinical Practice Guidelines for Early Diagnosis of Cerebral Palsy published in 2017 recommend algorithms for early diagnosis using methods proven through a systematic review. To diagnose cerebral palsy in infants younger than 5 months, the first option involves combining medical history taking, brain imaging techniques such as magnetic resonance imaging (MRI), and General Movement assessments (GM), Hammersmith Infant Neurological Examination (HINE), and the Test of Infant Motor Performance (TIMP).[11] If MRI tests are likely to pose a risk to infants' health or are not feasible in low- and medium-income countries, it is recommended to diagnose cerebral palsy using the second option, which combines the results of the HINE[12] and the TIMP.[13]

To assess neuromotor development, various tools, including TIMP, have been employed. The TIMP is widely used globally to assess developmental delays in infants with a Corrected Age (CA) of 4 months from 34 weeks of Gestational Age (GA).[14-16] Additionally, it is utilised to evaluate the prognosis of early intervention,[17] provide parental education on infant motor development,[18] and can predict and diagnose neuromotor development delays.[19, 20]

The TIMP Screening Items (TIMPSI) Test, a screening test of the TIMP, was developed for infants who are highly sensitive to external stimuli and for those who have difficulty undergoing the full TIMP assessment. This process entailed selecting and conducting psychometric analysis on three subsets from the entire TIMP, ultimately comprising 29 items. The selection was based on Rasch model analysis of 990 infants out of the 42 TIMP items. This procedure outlines the creation of TIMPSI by carefully choosing 29 items from the broader set of 42 items in the TIMP. TIMPSI serves as an early screening tool for infants suspected of developmental disabilities.[21] The full version of the TIMP takes 33 minutes (times range from about 21-45 minutes), while the average time required for TIMPSI is 22 minutes (times range from about 12-32 minutes), providing an advantage in evaluating early infancy.

Various studies related to reliability and validity must be conducted for the test tool to be used in clinical practice.[22] However, Research on the reliability and validity of the TIMPSI is relatively limited.[18] Investigated test-retest reliability only, while study focused on infants with spinal muscular atrophy type I.[23] To determine how effective TIMPSI is in screening neuromotor development among premature infants, we need to conduct reliability and validity studies designed specifically for this population. Thus, the primary aim of this study is to examine the test-retest and inter-rater reliability of the TIMPSI in high-risk infants under 4 months of age who are susceptible to neuromotor developmental challenges.

Methods and Participants

This methodological study aims to investigate the Test-Retest Reliability and Inter-rater Reliability of The Test of Infant Motor Performance Screening Items in High-risk Infants aged under 4 months with Neurodevelopmental Impairments.

A total of eighteen subjects were initially recruited for this study, with sixteen subjects meeting the inclusion criteria and participating in the research. Exclusion criteria led to the removal of one individual based on Prechtl's Behavioral States and another who couldn't undergo re-testing within three days. Recruitment was conducted through a hospital bulletin board notice targeting patients attending Hongik Rehabilitation Hospital in Changwon, Korea.

Inclusion criteria is 1) under 4 months CA; 2) premature or diagnosed with brain damage; 3) Precht's Behavioral States Level 3 and 4.[24] Exclusion criteria is 1) visual or auditory defection; 2) hereditary metabolic disease or an underlying disease; 3) congenital deformity or congenital heart disease; 4) unstable vital sign.[18, 25, 26]

The legal guardians of the subjects were comprehensively briefed about the study's procedures, objectives, potential benefits, risks, and all relevant aspects by the researcher. The study was conducted with the signed consent of the legal guardians. Ethical approval for all stages of the research was obtained from the Clinical Research Information System (KCT0005640).

Procedure

In this study, to minimise stress for infants in potentially challenging environments, video recordings, known to reduce stress and enhance accuracy, were employed during the test (Ko & Kim, 2012). An assistant

recorded the examination process conducted by Tester A using a Galaxy Note 20 (Samsung, Seoul, Korea). Tester A initially performed the TIMPSI and repeated the test within three days.

Tester B and Tester C performed while watching recorded videos in separate spaces. The testers provided no information regarding the age and medical history of the infants. Standardised materials and protocols were used. Infants, dressed only in diapers, were placed on a firm surface (rubber mat) at room temperature (25-29℃). If signs of stress were observed, the intervention was halted, and measures such as soothing or providing a pacifier were taken before resuming.

Throughout the test, signs of stress, including increased respiratory rate, changes in muscle tension, and cyanosis, were monitored. An oximeter and a sputum suction device were used if needed. Consistent with the Test User's manual Version 3, a shiny red ball (approximately 5 cm) provided visual stimulation, and a plastic rattle (approximately 10-12 cm) offered auditory stimulation. If two tests were conducted on the same day, a sufficient time gap ensured the infant's rest and suitable readiness for the subsequent test.

To evaluate Test-retest reliability, Testers B and C each independently scored both the first and second TIMPSI video sessions recorded by Tester A. The results from the first and second tests by all three testers were used. Inter-rater reliability was assessed using the results of the first TIMPSI test conducted by all three testers.

Test of Infant Motor Performance Screening Items (TIMPSI)

The TIMPSI serves as an early screening tool for infants suspected of developmental disabilities.[21] The items for TIMPSI were selected based on specific criteria: 1) they cover a range of difficulty levels determined through Rasch item analysis, 2) collectively, they assess the performance of all body parts, including the head, trunk, arms, and legs, and 3) they exhibit strong psychometric characteristics, demonstrating a good fit to the Rasch model and high item-to-total test score correlations. TIMPSI by extracting 29 items from the full 42-item TIMP using Rasch model analysis, based on a sample of 990 low birth weight infants born in the United States. In their study, they investigated the concurrent validity between the TIMP and the TIMPSI. When using a −0.25 SD cut-off score on the TIMPSI to predict performance on the full TIMP, they reported a sensitivity of 72%, specificity of 84%, positive predictive value of 63%, and negative predictive value of 89%, with a kappa coefficient of 0.54 ($p<0.0001$).

TIMPSI comprises three subsets: the Screening set, Easy set, and Hard set. The Screening set of TIMPSI utilises the best 11 items identified through Rasch psychometric analysis from the entire TIMP test to screen infants of any age.

The TIMPSI, in total, consists of 29 items, and the total score is derived from the sum of scores across all subsets. The highest achievable score is 98 points, indicating better exercise performance. According to the original validation study, the average duration of the test is 22 minutes, falling within a time range of 12 to 32 minutes (Figure 1).

Based on the raw score of the screening items, a second set is administered, consisting of either 10 easier items or 8 harder items, contributing to a comprehensive assessment of performance in various tasks involving postural control (Figure 2). The Screening set includes 11 items with 5-to-7-point rating scales (ranging from 0 to 51 points). The Easy set comprises 10 items, with 4 items dichotomously scored and 6 items with a 5 or 6 point rating scale (ranging from 0 to 31 points). The Hard set consists of 8 items, with 5 items dichotomously scored and 3 items on a 5-6 point evaluation scale (ranging from 0 to 17 points). The Easy and Hard sets are administered adaptively based on the infant's performance on the Screening set. Infants who score lower on the Screening set are subsequently evaluated using the Easy set, which includes simpler motor control tasks such as head alignment and limb movement in supported positions. In contrast, infants who demonstrate higher scores on the Screening set are tested using the Hard set, which contains more complex postural and antigravity control tasks, such as reaching or maintaining head and trunk stability against gravity. This adaptive design minimises fatigue and stress in lower-performing infants while preserving the test's sensitivity for detecting advanced motor control in higher-performing infants.

The total score for TIMPSI is calculated by combining the score of the Screening set with the score of either the Easy or Hard set. If the Hard set is administered, an additional 31 points are added to the recorded score. TIMPSI results indicate suspected developmental delay if the total score falls outside the average range (mean±1SD) for the same age group. In cases of suspected developmental delay, further tests are recommended. The test tool incorporates black and white photographs as a

Figure 1 Test of infant motor performance screening items.

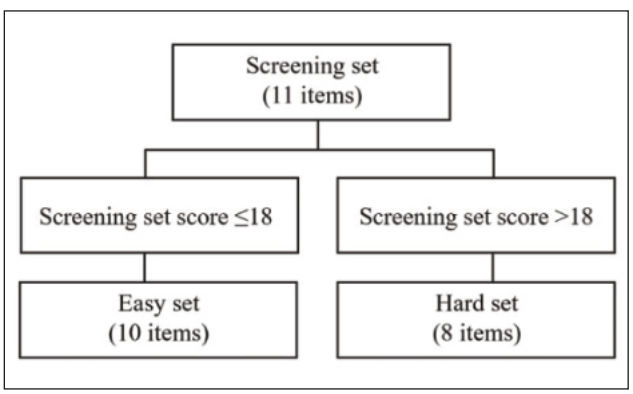

Figure 2 Flow chart of test of motor performance screening items.

qualitative criterion to confirm the tester's judgement on the infant's response level. TIMPSI version 1.1 was utilised. Compared with the original 42-item TIMP, the TIMPSI was designed as a concise screening version by omitting 13 items that were found to be redundant or provided limited additional information in the Rasch model analysis. These omitted items primarily involved transitional or mid-level movements that overlapped with adjacent items assessing similar motor control components. Consequently, the remaining 29 items of the TIMPSI comprehensively represent head, trunk, and limb control across a full range of task difficulty while maintaining strong psychometric properties. This

refinement allows the TIMPSI to retain the discriminative capacity of the full TIMP while substantially reducing administration time and physical burden on infants.

Statistical Analysis

In this study, statistical analysis was conducted using SPSS 25.0 for Windows. Normality of the data was assessed using the Shapiro-Wilk test. General characteristics of the subjects were analysed through frequency analysis and descriptive statistics. The investigation of reliability employed the Intraclass Correlation Coefficient (ICC).

ICC is a widely used indicator of repeatability and reproducibility, representing the proportion of total variation in measured values attributed to variation between individuals.[27] For this study, the two-way batch random effect model ICC (2.1) and its 95% confidence interval (CI), as well as the two-way batch mixed effect model ICC (3.1) with its 95% CI, were utilised.

The ICC analysis followed criteria outlined by Portney and Watkins.[28] Where ICC values of 0.90 or higher are considered very high, 0.75-0.90 as high, 0.50-0.75 as moderate, and less than 0.50 as low. The Standard Error of Measurement (SEM) was calculated to estimate the error in the unit of measurement, providing an expected error in the individual clinical experience of the tester.[29,30]

To assess the consistency of measured values, the Bland-Altman plot was employed. This graphical representation displays the difference between the measured values against the mean difference between the two test results, allowing for a visual evaluation of score distribution and potential measurement bias.[31] The significance level (α) for all statistical tests was set at 0.05 or less.

Result

In the study, there were sixteen participants, including six males and ten females. The average age of the infants was 59.12 days±47.34 days (Postmenstrual Age (PMA) 38 weeks – CA 16 weeks). The average gestation period was 198.5 days±29.23 days (28 weeks and 4 days), and the average birth weight was 1.330 kg±61g. All subjects had a history of neonatal intensive care unit admission, with an average hospitalisation period of 58.94 days±48.08 days. The general characteristics are as follows (Table 1).

Fifteen infants among the subjects were diagnosed with brain damage using diagnostic imaging Tools (one by brain ultrasound and fourteen by brain MRI), while the remaining one was a premature baby with no confirmed brain damage. The prevalent types of brain damage were Periventricular leukomalacia (PVL) in six infants, Interventricular haemorrhage (IVH) in four, Subarachnoid haemorrhages (SAH) in two, Cerebellar haemorrhages in two, and Cerebrum atrophy in one case. The average test duration was 17 minutes ± 3 minutes and 43 seconds, with an average interval of 2.56 days±0.72 days between the first and second tests.

Test-retest reliability of TIMPSI

Among the total of sixteen subjects, thirteen infants performed the Easy set with a score of 18 or less on the Screening set, while 3 subjects underwent the Hard set with a score of 19 or more. In the first test conducted by tester A, the average score for the screening set was 13.63±9.19, the Easy/Hard set was 13.81±11.98, and the total score averaged 27.44±20.58. The average scores for the screening set in the second test were 14.44±9.28, for the Easy/Hard set were 14.38±12.22, and for the total score were 28.81±21.05. The Test-retest reliability of the screening set was ICC=0.994 (0.984-0.998) with SEM 0.70. The Test-retest reliability of the Easy/Hard set was ICC=0.996 (0.990-0.999) with SEM 0.75, and for the total score, it was ICC=0.998 (0.994-0.999) with SEM 0.91. These results

Table 1 General characteristics of subjects (N=16)

	N=16
Variable	**Frequency (%) or Mean (SD)**
Gender	Male 6 (37.5%)
	Female 10 (62.5%)
Age (day)	59.12 (47.34)
Gestational period (day)	198.5 (29.23)
NICU care (day)	58.94 (48.08)
Birth weight (g)	1330 (61)
Brain damage	Periventricular leukomalacia 6 (37.5%)
	Interventricular haemorrhage 4 (25%)
	Subarachnoid haemorrhage 2 (12.5%)
	Cerebellar haemorrhages 2 (12.5%)
	Cerebrum atrophy 1 (6.25%)
	None 1 (6.25%)
Test time (minutes)	17 minutes (3 minutes 43 seconds)

NICU: Neonatal intensive care unit

demonstrated very high Test-retest reliability above ICC=0.90 for all test sets, and SEM was also less than 10% of the average score in all test sets ($p<0.05$).

In the first test conducted by tester B, the average scores for the screening set were 13.37±8.92, for the Easy/Hard set were 13.63±12.17, and for the total score were 27±20.63. In the second test, the average scores for the screening set were 14.06±8.73, for the Easy/Hard set were 14±12.44, and for the total score were 28.06±20.86. The Test-retest reliability of the screening set was ICC=0.991 (0.974-0.997) with SEM 0.82. The Test-retest reliability of the Easy/Hard set was ICC=0.997 (0.991-0.999) with SEM 0.66, and for the total score, it was ICC=0.997 (0.993-0.999) with SEM 1.11. These findings showed very high Test-retest reliability above ICC=0.90 for all test sets, and SEM was also less than 10% of the average score in all test sets ($p<0.05$) (Table 2).

In the first test conducted by Tester C, the average scores for the screening set were 13.44±8.86, for the Easy/Hard set were 14.06±12.01, and for the total score were 27.5±20.46. In the second test, the average scores for the screening set were 14.75±8.73, for the Easy/Hard set were 14.63±12.01, and for the total score were 29.83±20.38. The Test-retest reliability of the screening set was ICC=0.995 (0.986-0.998) with SEM 0.61. The Test-retest reliability of the Easy/Hard set was ICC=0.998 (0.993-0.999) with SEM 0.52, and for the total score, it was ICC=0.998 (0.995-0.999) with SEM 0.90. These results indicated very high Test-retest reliability above ICC=0.90 for all test sets, and SEM was also less than 10% of the average score in all test sets ($p<0.05$).

Following the Test-retest reliability assessment, we conducted a Bland-Altman analysis to examine the agreement between the scores obtained by each tester (Tester A, Tester B, Tester C). The Bland-Altman plot provides a visual representation of the differences in scores, allowing us to assess the consistency of measurements across the raters.

The plot illustrates the mean difference between the scores of two testers on the y-axis against the average of their scores on the X-axis. Additionally, 95% limits of agreement are displayed, representing the range within which 95% of the score differences fall.

Upon careful examination of the Bland-Altman plot (Figure 3), it was noted that only one case exceeded the 95% limits of agreement. However, the overall pattern indicates good consistency in the assessments conducted by different testers, with the majority of score differences falling within an acceptable range.

To assess the reliability of the test-retest, we compared the total scores from the first and second tests. The Bland-Altman plot for Test-retest reliability indicated an even score distribution within the 95% confidence limit, except for one subject tested by C. Interestingly, the SEM values for Testers B and C, who conducted video-based assessments, were slightly lower than those of Tester A. This may be due to the standardised and repeatable conditions of video observation, which eliminate variations in infant behavior or environmental factors across sessions. In contrast, live retesting by Tester A inherently involves real-time variability, which may increase measurement error.

This outlier, tested at PMA 38 weeks, exhibited consistently low scores, suggesting lower Test-retest reliability consistency in subjects with very low scores. While this outlier case warrants attention, the majority of the data points demonstrate reliable consistency among testers, supporting the robustness of the TIMPSI across different raters.

Table 2 Test-retest reliability of TIMPSI

		Test 1	Test 2	ICC (95% CI)	SEM
		Mean (SD)			
Tester A	Screening set	13.63 (9.19)	14.44 (9.28)	0.994 (0.984-0.998)	0.70
	Easy/Hard set	13.81 (11.98)	14.38 (12.22)	0.996 (0.990-0.999)	0.75
	Total	27.44 (20.58)	28.81 (21.05)	0.998 (0.994-0.999)	0.91
Tester B	Screening set	13.37 (8.92)	14.06 (8.73)	0.991 (0.974-0.997)	0.82
	Easy/Hard set	13.63 (12.17)	14.00 (12.44)	0.997 (0.991-0.999)	0.66
	Total	27.00 (20.63)	28.06 (20.86)	0.997 (0.993-0.999)	1.11
Tester C	Screening set	13.44 (8.86)	14.75 (8.73)	0.995 (0.986-0.998)	0.61
	Easy/Hard set	14.06 (12.01)	14.63 (12.01)	0.998 (0.993-0.999)	0.52
	Total	27.50 (20.46)	29.38 (20.38)	0.998 (0.995-0.999)	0.90

ICC: Intraclass correlation coefficient; CI: Confidence interval; SEM: Standard error of measurement

Inter-rater Reliability of TIMPSI

The Inter-rater reliability was analysed using the first test results conducted by three testers. The average score of tester A for the screening set was 13.63±9.19, for the Easy/Hard set was 13.81±11.98, and for the total score was 27.44±20.58. The average score of tester B for the screening set was 13.38±8.92, for the Easy/Hard set was 13.63±12.17, and for the total score was 27±0.63. The average score of Tester C for the screening set was 13.44±8.86, for the Easy/Hard set was 14.06±12.01, and for the total score was 27.5±20.46.

The Inter-rater reliability of the Screening set was ICC=0.982 (0.960-0.993). The Easy/Hard set had a slightly higher ICC=0.996 (0.992-0.999) than the Screening set, and the total score had an ICC=0.994 (0.986-0.998). It demonstrated very high Inter-rater reliability above ICC=0.90 for all test sets. SEM was less than 10% of the average score in all test sets, with the Screening set at 1.18, Easy/Hard set at 0.74, and total score at 1.55 ($p<0.05$) (Table 3).

Discussion

The TIMPSI, designed for assessing neuromotor performance levels in infants under 4 months, serves as an efficient and minimally burdensome motor performance test. This study aimed to explore the Test-retest and Inter-rater reliability of TIMPSI in High-risk Infants under 4 months with Neurodevelopmental Impairments. The findings demonstrate strong Test-retest reliability across all sets (ICC>0.991), with the total score exhibiting the highest reliability. Inter-rater reliability also proved to be excellent, surpassing ICC>0.982 across all sets.

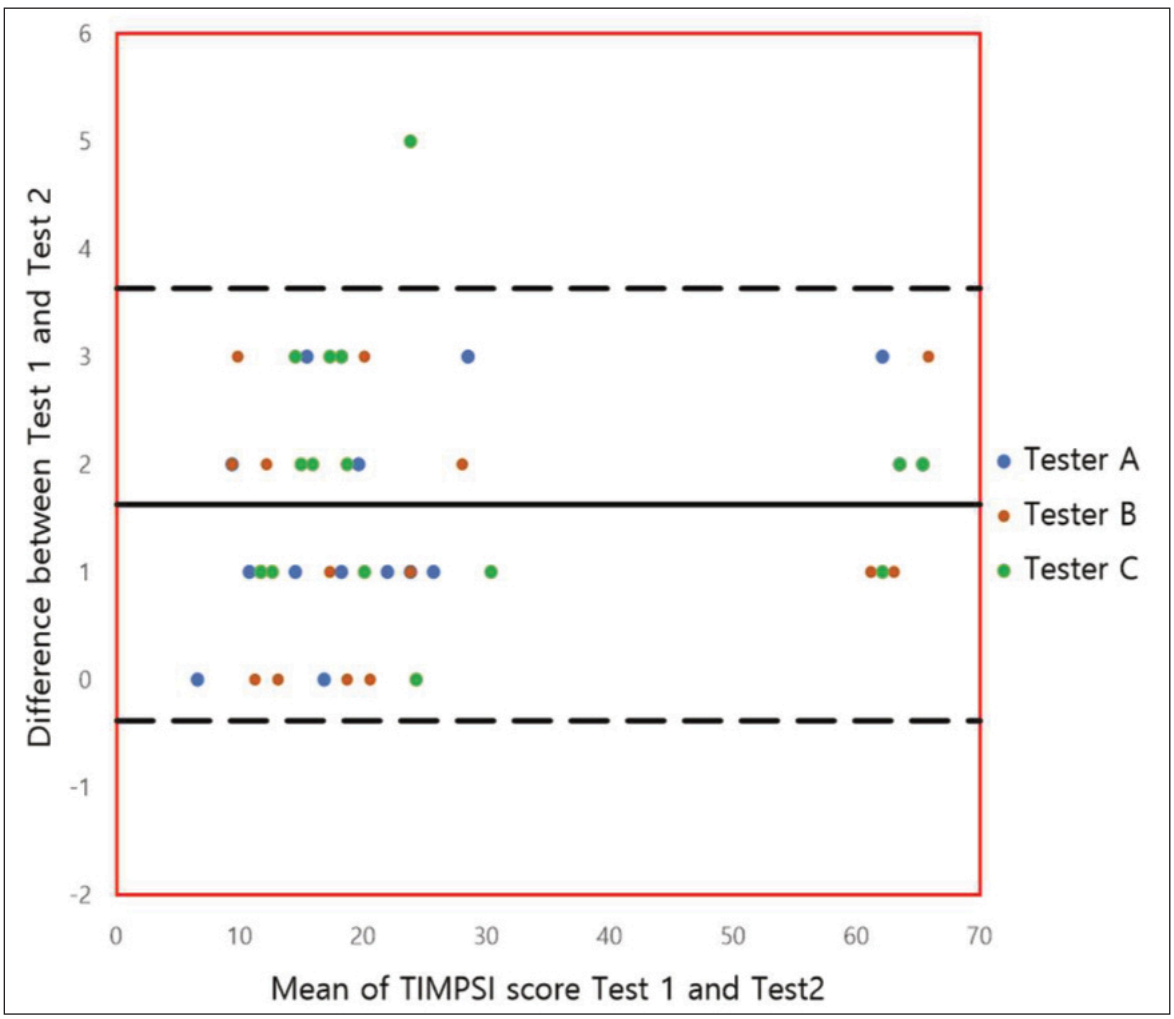

Figure 3 Bland-Altman Plot: Agreement between Test 1 and Test 2 for test-retest reliability.

Among the infants, thirteen underwent the Easy set and three underwent the Hard set. The Bland-Altman plot for Test-retest reliability indicated even score distribution within the 95% confidence limit, except for one subject tested by C. This outlier, tested at PMA 38 weeks, exhibited consistently low scores, suggesting lower Test-retest reliability consistency in subjects with very low scores.

This outlier case in the Bland-Altman plot warrants attention, indicating potential challenges in maintaining reliability for infants with extremely low scores, especially when tested at an earlier age. This observation aligns with Ustad, Evensen study,[18] highlighting the need for caution and further investigation when assessing infants with developmental concerns at an earlier developmental stage.

This is consistent with a study in which the TIMPSI of Ustad, Evensen[18] shows Test-retest reliability with a Bland-Altman plot. In his study, three of the 51 subjects were out of range of 95%, two of whom were infants tested at 36-37 weeks of age, showing low Test-retest reliability consistency in low-scoring subjects, but could not be generalised due to a small number of subjects.[18]

Inter-rater reliability analysis for the first test by three testers showed ICC=0.982 for the Screening set, ICC=0.996 for the Easy/Hard set, and ICC=0.994 for the total score. The Easy/Hard set displayed the highest reliability, possibly due to its correlation with posture control levels and responses to stimuli and posture changes from the Screening set.

Comparing with Krosschell, Maczulski study,[23] which analysed the reliability and validity of the TIMPSI in infants with SMA type I,[23] our results align. In their study, the test was conducted by 9 evaluators on 38 infants to assess Test-retest reliability, and the results, expressed as the Pearson correlation coefficient, demonstrated excellence in the total score (p=0.35), screening set (p=0.45), and easy set (p=0.29). Inter-rater reliability was also evaluated by 12 evaluators on 4 infants, and the reported results indicated excellent reliability. Our study similarly found excellent Test-retest and Inter-rater reliability for TIMPSI.

In this study, Test-retest reliability for TIMPSI total score was ICC=0.997-0.998, and Inter-rater reliability was ICC=0.994, both demonstrating excellent reliability suitable for clinical application. ICC=0.90 or higher is deemed desirable, and our results meet this criterion. ICC=0.75-0.90 is considered highly reliable, but an ICC of 0.90 or higher is deemed desirable for clinical application.[31] Both Test-retest reliability and Inter-rater reliability investigated in this study demonstrated ICC values of 0.90 or higher, making them suitable for clinical application. In a study by Campbell, Swanlund,[21] the TIMPSI, comprising 29 items, exhibited robust psychometric characteristics among TIMP's items, showcasing high reliability and validity as neuromotor development tests for infants under 4 months. Although this study did not compute internal consistency coefficients such as Cronbach's α, it is worth noting that the TIMPSI items were originally selected through rigorous Rasch analysis based on a large-scale infant dataset (n=0.990), which ensures structural coherence and item homogeneity across subsets.[21] Therefore, the internal consistency of the instrument has been previously validated during its development. The Test-retest reliability in our study further supports the high reliability of TIMPSI, which, along with its validated item structure from Rasch analysis, confirms its internal consistency and clinical applicability.

Limitations include small sample sizes from specific geographic regions and a focus on high-risk infants. Additionally, videotaped testing outcomes may vary depending on the tester's proficiency. Future research needs to target more diverse populations, taking into account ethnic, social, and cultural characteristics. While the TIMPSI demonstrates strong reliability as a rapid screening tool for early identification of infants at risk for neurodevelopmental delays, its ability to predict or prevent long-term impairment remains unclear, and it should be considered a preliminary tool to guide referrals for comprehensive evaluation rather than a standalone predictor of outcomes.

Table 3 Inter-rater reliability of TIMPSI

	Tester A	Tester B	Tester C	ICC (95% CI)	SEM
		Mean (SD)			
Screening set	13.63 (9.19)	13.38 (8.92)	13.44 (8.86)	0.982 (0.960-.993)	1.18
Easy/Hard set	13.81 (11.98)	13.63 (12.17)	14.06 (12.01)	0.996 (0.992-0.999)	0.74
Total	27.44 (20.58)	27.00 (20.63)	27.50 (20.46)	0.994 (0.986-0.998)	1.55

ICC: Intraclass correlation coefficient; CI: Confidence interval; SEM: Standard error of measurement

Conclusion

This study analysed the test-retest reliability (ICC=0.997-0.998) and inter-rater reliability (ICC=0.994) of the TIMPSI in a sample of 16 high-risk infants under four months of age. TIMPSI was identified as a rapid and economical screening tool that facilitates early assessment and therapeutic intervention. In particular, compared to the full TIMP, the TIMPSI requires a significantly shorter administration time (an average of 17 minutes) and fewer test items, thereby minimising stress on the infant. Therefore, in situations where the administration of the full TIMP is challenging, the TIMPSI may serve as an effective alternative for assessing the neurodevelopmental status of infants under four months of age. However, although the TIMPSI is a highly reliable tool for the early screening of neurodevelopmental disorders, its ability to predict or prevent long-term neurodevelopmental impairments remains unclear. Consequently, the TIMPSI should be considered an initial screening tool within a broader, comprehensive assessment framework rather than an independent predictor of long-term developmental outcomes. Further research is warranted to clarify its prognostic utility.

Conflict of Interest

All authors declare no conflict of interest.

References

1. Ophelders D, Gussenhoven R, Klein L, et al. Preterm Brain Injury, Antenatal Triggers, and Therapeutics: Timing Is Key. Cells 2020;9:1871.
2. Jarjour IT. Neurodevelopmental outcome after extreme prematurity: a review of the literature. Pediatric Neurol 2015;52:143-52.
3. Anderson P, Doyle LW; Victorian Infant Collaborative Study Group. Neurobehavioral outcomes of school-age children born extremely low birth weight or very preterm in the 1990s. JAMA 2003;289:3264-72.
4. Ryan MA, Murray DM, Dempsey EM, Mathieson SR, Livingstone V, Boylan GB. Neurodevelopmental outcome of low-risk moderate to late preterm infants at 18 months. Front Pediatr 2023;11:1256872.
5. Benzies KM, Magill-Evans JE, Hayden KA, Ballantyne M. Key components of early intervention programs for preterm infants and their parents: a systematic review and meta-analysis. BMC Pregnancy Childbirth 2013;13 Suppl 1:S10.
6. Perrone S, Negro S, Tataranno ML, Buonocore G. Oxidative stress and antioxidant strategies in newborns. J Matern Fetal Neonatal Med 2010; Suppl 3:63-5.
7. Pineda R, Durant P, Mathur A, Inder T, Wallendorf M, Schlaggar BL. Auditory Exposure in the Neonatal Intensive Care Unit: Room Type and Other Predictors. J Pediatr 2017:183:56-66.
8. Spittle AJ, Doyle LW, Boyd RN. A systematic review of the clinimetric properties of neuromotor assessments for preterm infants during the first year of life. Dev Med Child Neurol 2008;50:254-66.
9. Blencowe H, Cousens S, Chou D, et al. Born too soon: the global epidemiology of 15 million preterm births. Reprod Health 2013;10:S2.
10. Luu TM, Rehman Mian MO, Nuyt AM. Long-Term Impact of Preterm Birth: Neurodevelopmental and Physical Health Outcomes. Clin Perinatol 2017;44:305-14.
11. Bosanquet M, Copeland L, Ware R, Boyd R. A systematic review of tests to predict cerebral palsy in young children. Dev Med Child Neurol 2013;55:418-26.
12. Romeo DM, Ricci D, Brogna C, Mercuri E. Use of the Hammersmith Infant Neurological Examination in infants with cerebral palsy: a critical review of the literature. Dev Med Child Neurol 2016;58:240-5.
13. Novak I, Morgan C, Adde L, et al. Early, Accurate Diagnosis and Early Intervention in Cerebral Palsy: Advances in Diagnosis and Treatment. JAMA Pediatr 2017;171:897-907.
14. Barbosa VM, Campbell SK, Berbaum M. Discriminating infants from different developmental outcome groups using the Test of Infant Motor Performance (TIMP) item responses. Pediatr Phys Ther 2007;19:28-39.
15. Dos Santos Chiquetti EM, Valentini NC. Test of Infant Motor Performance for Infants in Brazil: Unidimensional Model, Item Difficulty, and Motor Function. Pediatr Phys Ther 2020;32:390-7.
16. Lee EJ, Han JT, Lee JH. Risk factors affecting Tests of Infant Motor Performance (TIMP) in pre-term infants at post-conceptional age of 40 weeks. Dev Neurorehabil 2012;15:79-83.
17. Oberg GK, Campbell SK, Girolami GL, Ustad T, Jørgensen L, Kaaresen PI. Study protocol: an early intervention program to improve motor outcome in preterm infants: a randomized controlled trial and a qualitative study of physiotherapy performance and parental experiences. BMC Pediatr 2012;12:15.
18. Ustad T, Evensen KA, Campbell SK, et al. Early Parent-Administered Physical Therapy for Preterm Infants: A Randomized Controlled Trial. Pediatrics 2016;138:e20160271.
19. Campbell SK, Kolobe TH, Wright BD, Linacre JM. Validity of the Test of Infant Motor Performance for prediction of 6-, 9- and 12-month scores on the Alberta Infant Motor Scale. Dev Med Child Neurol 2002;44:263-72.
20. Kolobe TH, Bulanda M, Susman L. Predicting motor outcome at preschool age for infants tested at 7, 30, 60, and 90 days after term age using the Test of Infant Motor Performance. Phys Ther 2004;84:1144-56.
21. Campbell SK, Swanlund A, Smith E, Liao PJ, Zawacki L. Validity of the TIMPSI for estimating concurrent performance on the test of infant motor performance. Pediatr Phys Ther 2008;20:3-10.
22. Rothstein JM, Campbell SK, Echternach JL, et al. Standards for Tests and Measurements in Physical Therapy Practice. Physical Therapy 1991;71:589-622.
23. Krosschell KJ, Maczulski JA, Scott C, et al. Reliability and validity of the TIMPSI for infants with spinal muscular atrophy type I. Pediatr Phys Ther 2013;25:140-8.

24. Prechtl HF. The behavioural states of the newborn infant (a review). Brain Res 1974;76:185-212.
25. Chiquetti EMDS, Valentini NC, Saccani R. Validation and Reliability of the Test of Infant Motor Performance for Brazilian Infants. Phys Occup Ther Pediatr 2020;40:470-85.
26. Song YH, Chang HJ, Shin YB, et al. The Validity of Two Neuromotor Assessments for Predicting Motor Performance at 12 Months in Preterm Infants. Ann Rehabil Med 2018;42:296-304.
27. Szklo M, Nieto FJ, editors. Epidemiology: Beyond the Basics 1999.
28. Portney LG, Watkins MP, editors. Foundations of Clinical Research: Applications to Practice 2015.
29. Lahey MA, Downey RG, Saal FE. Intraclass correlations: There's more there than meets the eye. Psychological Bulletin 1983;93:586-95.
30. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-8.
31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.