



ISSN: 2651-4451 • e-ISSN: 2651-446X

Turkish Journal of Physiotherapy and Rehabilitation

2020 31(3)240-246

Emine Handan TÜZÜN, PhD, PT¹
Levent EKER, MD¹
Selma UZUNER, MSc, PT²
Buse SEZEREL, MSc, PT³
Hayriye TOMAÇ, MSc, PT¹
Serpil MIHÇIOĞLU, MSc, PT¹
Wayne DITE, MAppSci(Research),
BAppSci(PE)⁴

- 1 Eastern Mediterranean University, Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Famagusta, Northern Cyprus, 10 Mersin, Turkey.
- 2 University of Kyrenia, Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Kyrenia, Northern Cyprus, 10 Mersin, Turkey.
- 3 Kyrenia Dr. Akçiçek Hospital, Department of Physical Therapy and Rehabilitation, Kyrenia, Northern Cyprus, 10 Mersin, Turkey.
- 4 Royal Talbot Rehabilitation Centre, Austin Health, Kew, Vic., Australia.

Correspondence (İletişim):

Emine Handan TÜZÜN, PhD, PT
Eastern Mediterranean University
Health Sciences Faculty,
Department of Physiotherapy and Rehabilitation,
Famagusta, Northern Cyprus, 10 Mersin, Turkey.
Phone: +90 392 630 13 70
E-mail: handan.tuzun@gmail.com
ORCID: 0000-0001-6989-6675

Levent EKER
E-mail: ekerlevent95@gmail.com
ORCID: 0000-0001-7301-4073

Selma UZUNER
E-mail: selma_uzuner_92@hotmail.com
ORCID: 0000-0002-4199-0093

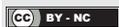
Buse SEZEREL
E-mail: busesezerel@hotmail.com
ORCID: 0000-0002-7265-0657

Hayriye TOMAÇ
E-mail: hayriye.tomac123@hotmail.com
ORCID: 0000-0002-8245-9995

Serpil MIHÇIOĞLU
E-mail: srplmhcgil@gmail.com
ORCID: 0000-0001-7186-324X

Wayne DITE
E-mail: wayne.dite@austin.org.au
ORCID: 0000-0002-6568-7363

Received: 08.08.2019 (Geliş Tarihi)
Accepted: 03.02.2020 (Kabul Tarihi)



Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

VALIDITY AND RELIABILITY OF THE FOUR SQUARE STEP TEST IN TYPICALLY DEVELOPED CHILDREN

ORIGINAL ARTICLE

ABSTRACT

Purpose: The purpose of this study was to evaluate the reliability and the validity of the four square step test (FSST), and level of agreement between the FSST and the timed up and go test (TUG) in typically developed children.

Methods: Two raters assessed 150 children aged 5 to 10 years in two consecutive days. Concurrent and discriminative validities were determined. To evaluate the intra- and inter-rater reliability, the intra-class correlation coefficient (ICC) was used. The Bland-Altman plot and analysis were performed to evaluate the agreement between the measurements obtained by two raters and by the FSST and TUG.

Results: The FSST and TUG scores were strongly correlated ($r=0.685$, $p=0.001$). The FSST was able to distinguish between children by age groups ($p=0.001$). The ICC2.1 and ICC2.2 estimates were 0.86 and 0.81, respectively. For the assessments of two raters, the Bland-Altman plot showed that all values fell within 95% limits of agreement. Bias was calculated as -0.12 ± 0.50 . The 96% of the data points were within the ± 1.96 SD of the mean difference for the FSST and TUG scores. Bias was calculated as 0.47 ± 0.80 .

Conclusion: The FSST is a valid and reliable test to assess the dynamic balance of typically developed children. However, the FSST cannot be substituted for the TUG.

Key Words: Balance; Children; Reliability; Validity.

TİPİK GELİŞİM GÖSTEREN ÇOCUKLARDA DÖRT KARE ADIM TESTİNİN GEÇERLİK VE GÜVENİRLİĞİ

ARAŞTIRMA MAKALESİ

ÖZ

Amaç: Bu çalışmanın amacı, tipik gelişim gösteren çocuklarda dört kare adım testinin (FSST) güvenilirlik ve geçerliliğini ve FSST ve süreli kalk yürü testi (TUG) ile arasındaki uyum düzeyini değerlendirmektir.

Yöntem: 5-10 yaş arası 150 çocuk iki değerlendirici tarafından ardışık iki günde değerlendirildi. Eş zamanlı ve ayırt edici geçerlikler belirlendi. Değerlendirici içi ve değerlendiriciler arası güvenilirliği değerlendirmek için, sınıf-içi korelasyon katsayısı (ICC) kullanıldı. İki değerlendirici tarafından yapılan ölçümler ve FSST ve TUG arasındaki uyumu değerlendirmek için Bland-Altman grafiği ve analizi yapıldı.

Sonuçlar: FSST ve TUG kuvvetli şekilde ilişkiliydi ($r=0,685$, $p=0,001$). FSST, çocukları yaş gruplarına göre ayırt edebildi ($p=0,001$). ICC2.1 ve ICC2.2 tahminleri sırasıyla 0,86 ve 0,81'di. İki değerlendiricinin değerlendirmeleri için, Bland-Altman grafiği, tüm değerlerin % 95'lik uyum limitleri içinde olduğunu gösterdi. Yanlılık $-0,12 \pm 0,50$ olarak hesaplandı. Veri noktalarının % 96'sı FSST ve TUG skorları için ortalama farkın $\pm 1,96$ standart sapması içindeydi. Yanlılık $0,47 \pm 0,80$ olarak hesaplandı.

Tartışma: FSST, tipik gelişim gösteren çocuklarda dinamik dengeyi değerlendirmek için geçerli ve güvenilir bir testtir. Bununla birlikte, FSST, TUG testinin yerine kullanılamaz.

Anahtar Kelimeler: Denge; Çocuk; Güvenirlik; Geçerlik.

INTRODUCTION

Balance is a complex motor characteristic that has physiological, biomechanical, and psychological components, and it may be static or dynamic. Both the static or dynamic balance is essential for accomplishing in functional tasks involved in activities of daily living (1,2). Static balance refers to maintain the center of mass within the base of support while the body is not moving. In contrast, dynamic balance refers to anticipating and reacting to changes in balance as the body moves through space (3). Recent research has shown that independent mobility is linked to emotional, social, motor, language, and other developmental benefits in children. Independent and safe mobility allows children to increase their interaction with the peers (4). Therefore, assessing children's balance using valid and reliable instruments is vital for ensuring the development of safe mobility and independence with functional tasks. Currently, dynamic balance could be evaluated using clinical or laboratory tests. However, clinical balance tests are more practical and cheaper than laboratory-based tests (5). In clinical settings, both single balance measures and batteries of balance tests can be used for evaluating stability during locomotion in typically developed children. The timed up & go (tug) test, pediatric balance scale, berg balance scale, functional walking test, and timed up and down stairs are commonly used tests and have evidence for good validity and reliability (6). Among these tests, no single measure has been reported as the gold standard in the assessment of dynamic balance in children. Therefore, the use of tests, in combination, could substantially enhance the prediction of overall dynamic balance (6). In recent years, in addition to these dynamic balance tests, the four square step test (FSST), which was originally designed for the elders (7), has been the subject of research for dynamic balance assessment of pediatric populations. It was shown that FSST was a valid and reliable measure for assessing the dynamic balance in children with cerebral palsy and Down syndrome (8,9). In a pilot study conducted in 15 typically developed children, Salis and Samuel have found that the FSST had good inter- and intra-rater reliability. The authors have noted the need for further researches in large

sample size (10). In another study conducted in 30 typically developed children, Leizerowitz and Katz-Leurer have reported that the lenient version of the FSST was a feasible, moderately reliable, and valid test to assess dynamic balance in this population (11). Based on the current literature, it appears that there is insufficient evidence to support the use of the FSST in typically developed children. The purpose of this study was to evaluate the concurrent and discriminative validity and reliability of FSST, and the level of agreement between the raters for, TUG and FSST in typically developed children.

METHODS

Study Design and Participants

This observational study was carried out at Cihangir-Düzova Elementary School in Famagusta, North Cyprus, between October 2018 and January 2019 after the approval of the Eastern Mediterranean University BAYEK Health Ethics Subcommittee. Before the study, the Directorate of North Cyprus Primary Education Department was informed for the study and the permission was obtained. Written informed consent was obtained from all participants' primary caregivers and assent. A priori sample size calculation was performed using the PASS (2005), taking into account the statistical tests to be used in the analyses (12). A sample size of 146 subjects with four observations per subject achieves 80% power to detect an ICC of 0.60 under the alternative hypothesis when the ICC under the null hypothesis is 0.50 using an F-test with a significance level of 0.05. Considering the drop-out risk of the participants due to various reasons, the initial sample size was increased to 150 subjects. Inclusion criteria used for selecting the subjects were (1) aged 5 to 10 years, (2) absence of musculoskeletal or neurological disorders, (3) no history of any orthopedic surgery, and (4) not uncorrected visual acuity.

Assessors and Procedures

Two physiotherapists with four years of experience in physiotherapy and rehabilitation units assessed participants on two separate days. Before the trial, testing instructions to be delivered to participants (7,13) were translated and adapted into Turkish.

In order to maximize potential agreement and to improve rating accuracy, reliability, and validity, assessment skills of the raters were refreshed through training.

The administration of the tests was performed in a corridor of the school. Assessments were performed in a quiet and adequately lighted environment to minimize the distraction. Participants were tested with their regular footwear. The demographic information of participants was recorded. The dynamic balance performances on the FSST were assessed by Rater A and B independently, and this assessment repeated 24-h later. In order to avoid any discrimination between children participating in the study, the TUG test also was repeated by two raters with a 24-hour interval. The participants were allowed to rest up to five minutes between the sessions. A stopwatch was used to measure time to complete the FSST and TUG tests.

FSST Assessment Protocol

A square was formed by using four 90 cm long canes placed on the ground. The sequence was explained and demonstrated to the participants. They were allowed one practice trial and, subsequently, two trials. Instructions were given and encouraged to complete the sequence as fast as possible without touching the canes. The sequence of trial was 1,2,3,4,1,4,3,2,1. If a participant touched canes or performed test sequences inaccurately, or both feet of the participant did not make contact with the floor in each square, the trial was considered as unsuccessful, and the test was repeated. One unsuccessful trial was allowed. The best time was taken as the score of FSST, which was completed successfully (10,11).

TUG Assessment Protocol

A three-meter pathway on the testing floor was marked with red-colored masking tape. The TUG

was administered using one standard chair. The TUG was explained and demonstrated to the participants. They were allowed one practice trial and, subsequently, two trials. If the participant ran on the pathway, or stopped at the turn, or did not sit down right away, the trial was considered as unsuccessful, and the test was repeated. One unsuccessful trial was allowed. The best time was taken as the score of TUG, which was completed successfully (6).

Statistical Analysis

All statistical analyses were carried out using the IBM SPSS Statistics software version 22.0 (IBM Corporation, Armonk, NY, USA). Before statistical tests were conducted, we checked potential outliers and missing data. Normal distribution assumptions of the data were checked using the Shapiro-Wilk test. Evaluations of concurrent and discriminative validity were based on the mean scores from the first measurement of rater A. Pearson's correlation coefficient (r) was used to assess the concurrent validity of FSST. Pearson's correlation coefficient was interpreted as follows: 0.00-0.19=very weak; 0.20-0.39=weak; 0.40-0.59=moderate; 0.60-0.79=strong; and 0.80-1.00=very strong (14). Dynamic balance is influenced by multiple variables, including biological fitness and age (15,16). Based on the literature, we hypothesized that FSST classifies children into age groups based on the dynamic balance assessments. The children were divided into three age groups: (1) 5- to 6-year-olds (2) 7- to 8-year-olds, and (3) 9- to 10-year-olds. Between-groups discriminative validity of the FSST was evaluated using the one-way analysis of variance (ANOVA) with Bonferroni post-hoc test. Both intra- and inter-rater reliability were assessed using intraclass correlation coefficients (ICC). ICC2.1 and ICC2.2 estimates and their 95% confidence intervals were calculated based on absolute-

Table 1: Characteristics of the Participants (n=150).

Variables	Participants (n=150)	
	Mean±SD	Min-Max
Age (years)	7.79±1.68	5-10
Gender (females/males)	77/ 73 ^s	51.3/48.70 ^a
Weight (kg)	29.24±9.30	16-67
Height (cm)	128.83±10.42	104-157

^sfrequency, ^apercentages.

Table 2: Our Square Step Test Scores of Children by Age Groups Based on the First Assessment of Rater A.

Age Groups	n	Mean±SD	Min-Max	95% CI [†] for mean	p [§]
5-6 years	31	5.43±0.95	4.03-7.47	5.08-5.78	0.001*
7-8 years	53	4.41±0.80	3.05-6.53	4.19-4.63	
9-10 years	66	3.95±0.66	3.00-5.61	3.79-4.11	

*p<0.05. [§]One-way ANOVA, Bonferroni Alpha=0.016, [†]CI: Confidence Interval.

agreement (17,18). ICC estimates were interpreted as follows: <0.50=poor; 0.50-0.75=moderate; 0.75-0.90=good; and >0.90=excellent reliability (17).

Bland and Altman plots were constructed to demonstrate agreement between TUG and FSST, and between assessments of rater A and B. The mean difference between the assessments was reported as bias. The mean difference ± (standard deviation of the mean difference multiplied by 1.96) was reported as 95% limits of agreement (LOA). Bland and Altman recommended that 95% of data points should lie within two standard deviations of the mean difference for a good agreement (19). The statistical significance level was set at p<0.05.

RESULTS

The characteristics of the participants are shown in Table 1. A total of 150 subjects aged between 5 and 10 were included in the study.

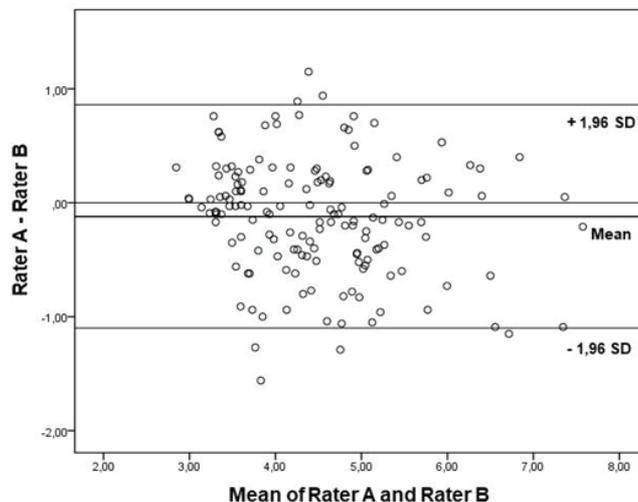
The FSST demonstrates statistically significant strong positive correlation ($r=0.685$, 95% CI; 0.589-0.761 $p=0.001$) with the TUG. As shown in Table 2, results from the one-way ANOVA analysis

showed that dynamic balance assessed using the FSST were able to distinguish between the groups ($F_{(2,147)}=38.293$, $p=0.001$). Bonferroni post-hoc analysis indicated significant differences between the mean FSST score of each age category (all p 's <0.05). The 95% confidence intervals of the FSST scores in each age category did not overlap.

ICC2.1 estimate for inter-rater reliability was 0.86 (95% CI; 0.81-0.90). ICC2.2 estimate for intra-rater reliability was calculated as 0.81 (95% CI 0.74-0.86).

On the Bland and Altman plot, for the first day assessments of the FSST, the 95% of the data points were within the $\pm 1.96SD$ of the mean difference (Figure 1). Upper LOA and lower LOA were -1.11 and 0.87, respectively. Bias was calculated as -0.12 ± 0.50 (95% CI -0.3-0.04).

The Bland and Altman plot for data based on the Rater B assessments of TUG and FSST indicated that 96% of the data points were within $\pm 1.96SD$ of the mean difference (Figure 2). Upper LOA and lower LOA were 2.04 and -1.1, respectively. Bias

**Figure 1:** Bland and Altman plot for the data based on first the four square step test assessments of Rater A and Rater B.

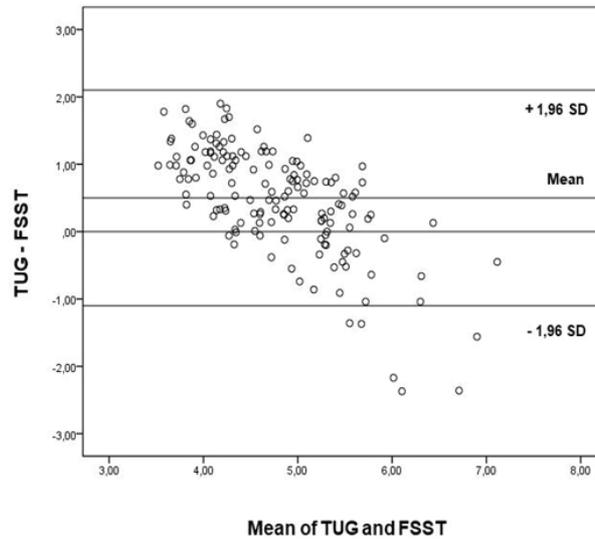


Figure 2. Bland and Altman plot for data based on the Rater B assessments of the timed up and go test and the four square step test.

was calculated as 0.47 ± 0.8 (95% CI 0.3-0.6).

DISCUSSION

The results of this study indicated that the FSST was significantly related to the TUG. It also demonstrated good discriminative validity by differentiating children according to their age groups. The FSST scores were stable and consistent over repeated administrations. There was a good inter-rater agreement. However, the level of agreement between the FSST and TUG was not sufficient.

Leizerowitz and Katz-Leurer have found a moderately significant correlation between the original form FSST and TUG scores ($r_s=0.53$; 95% CI 0.21-0.75, $p<0.05$) (11). Contrary to their finding, we found a strong correlation between FSST and TUG scores ($r=0.69$, 95% CI 0.59-0.77, $p=0.001$). If we consider 95% CI's of the size of correlation coefficients, it is clear that there is an overlapping between the interval estimates. Therefore, this result could not be interpreted as dissimilarity of the findings of two studies.

Maturation and maintenance of balance is a complex and active sensorimotor function (20,21). Proprioceptive, visual, vestibular, and cognitive systems play a significant role in balance control. The maturation of these systems does not take

place at the same time. Previous studies reported that maturation of the proprioceptive system takes place at 3 to 4 years of age, subsequently visual and vestibular systems (22,23). Lonescu et al. reported that the maturation process occurs throughout childhood until the age of 12 years (24) and does not reach adult level, even at the age of 14-15 years (25). Based on this finding, it was expected that any test used to assess the dynamic balance of typically developing children should differentiate the age groups. In this study, those children with lower age groups were rated as exhibiting low levels of dynamic balance control. This is an expected finding indicating that the FSST has also a good discriminative validity. However, it should be noted that no other studies have been conducted to investigate the discriminative validity of the FSST on these age groups, therefore it is difficult to support this finding in relation to the literature.

ICC based results of this study showed that the inter-rater reliability of the FSST scores was good. Intra-rater estimates ranged from moderate to good. Both findings were in agreement with the previous studies (10,11).

In this study, we found that 95% of the data points of Rater A and B were within ± 1.96 SD of the mean difference of their assessments. The bias was

not significant because the line of equality that is "0" line was within the confidence interval of the mean difference of two assessments. This finding indicated that the inter-rater agreement of the FSST scores was good. On the contrary, there was a significant bias on the mean difference of the FSST and TUG scores. Because the line of equality was not within the confidence interval of the mean difference of the FSST and TUG scores, based on this result, we concluded that the FSST could not be used in place of the TUG. It was not surprising that the TUG, and FSST was used to detect different types of balance and mobility. The FSST is used to assess dynamic stability and the ability of the subject to step over low objects forward, sideways, and backward. On the contrary, the TUG test is used to assess a person's mobility that requires both static and dynamic balance.

This study was conducted on children aged 5-10 years. Therefore, results cannot be generalized to children of all ages. It is a limitation of the study. The strength of this study was two folds. Firstly, the study was conducted on a sufficiently powered sample size. Secondly, in this study, both statistical and graphical methods were used to establish the agreement of two tests. In conclusion, the FSST test is a valid and reliable method to measure the dynamic balance of typically developed children. Further study is needed to establish strong evidence on the issue.

Sources of Support: No funding was received for this study.

Conflict of Interest: The authors declare no conflict of interest for the study.

Ethical Approval: The Eastern Mediterranean University BAYEK Health Ethics Subcommittee, ETK00-2018-0252 (Approval Date: 15.10.2018, and Approval Number: ETK00-2018-0252).

Peer-Review: Externally peer-reviewed.

Author Contributions: Concept – EHT; Design – EHT, SU, BS, HT, SM, LE, WD; Supervision – EHT, LE; Resources and Financial Support– SU, BS, SM, HT; Materials- SU, BS, SM, HT; Data Collection and/or Processing – EHT, Analysis and/or Interpretation – LE; Literature Research – SU, BS, SM, HT; Writing

Manuscript – EHT, SU, BS, HT, SM, LE, WD; Critical Review – EHT, LE, WD.

Acknowledgements: The material has not been previously published, in whole or in part.

REFERENCES

- Shumway-Cook A, McCollum G. Assessment and treatment of balance deficits. In: Montgomery PC, Connolly BH eds. Motor control and physical therapy: theoretical framework and practical application. Hixson: Chattanooga Group; 1991.
- Huxham FE, Goldie PA, Patla AE. Theoretical considerations in balance assessment. Aust J Physiother. 2001;47(2):89-100.
- Medical Dictionary for the Health Professions and Nursing. 2012. <https://medical-dictionary.thefreedictionary.com/dynamic+balance>. Accessed: October 20, 2018.
- Logan SW, Feldner HA, Galloway JC, Huang HH. Modified ride-on car use children with complex medical needs. Pediatr Phys Ther. 2016;28(1):100-7.
- Emery A, Cassidy JD, Klassen T, Rosychuk R, Rowe B. Development of a clinical static and dynamic standing balance measurement tool appropriate for use in adolescents. Phys Ther. 2005;85(6):502-14.
- Verbecque E, Lobo Da Costa PH, Vereeck L, Hallemans A. Psychometric properties of functional balance tests in children: a literature review. Dev Med Child Neurol. 2015;57(6):521-9.
- Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults. Arch Phys Med Rehabil. 2002;83(11):1566-71.
- Verma A, Samuel AJ, Aranha VP. The four square step test in children with Down syndrome: reliability and concurrent validity. J Pediatr Neurosci. 2014;9(3):221-6.
- Bandong ANJ, Madriaga GO, Gorgon EJR. Reliability and validity of the Four square step test in children with cerebral palsy and Down syndrome. Res Dev Disabil. 2015;47:39-47.
- Salis JS, Samuel AJ. Four square step test in children with typical development: test-retest and inter rater reliability: a pilot study. Physioaccord 2014-National Physiotherapy Conference, January 17-18, 2014, Bangalore, India.
- Leizerowitz G, Katz-Leurer M. Feasibility, stability, and validity of four square step test in typically developed children and children with brain damage. Brain Inj. 2017;31(10):1356-61.
- Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17(1):101-10.
- Williams EN, Carroll SG, Reddihough DS, Phillips BA, Galea MP. Investigation of the timed 'Up & Go' test in children. Dev Med Child Neurol. 2005;47(8):518-24.
- Evans JD. Straightforward statistics for the behavioral sciences. California: Pacific Grove; 1996.
- Largo RH, Caffisch JA, Hug F, Muggli K, Molnar AA, Molinari L, et al. Neuromotor development from 5 to 18 years. Part 1: timed performance. Dev Med Child Neurol. 2001;43(7):436-43.
- Butz SM, Sweeney JK, Roberts PL, Rauh MJ. Relationships among age, gender, anthropometric characteristics, and dynamic balance in children 5 to 12 years old. Pediatr Phys Ther. 2015;27(2):126-33.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-63.
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30-46.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.

20. Hsu YS, Kuan CC, Young YH. Assessing the development of balance function in children using stabilometry. *Int J Pediatr Otorhinolaryngol.* 2009;73(5):737-40.
21. D'Hondt E, Deforche B, Bourdeaudhuij I, Gentier I, Tanghe A, Shultz S, et al. Postural balance under normal and altered sensory conditions in normal-weight and overweight children. *Clin Biomech (Bristol, Avon).* 2011;26(1):84-9.
22. Cumberworth VL, Patel NN, Rogers W, Kenyon GS. The maturation of balance in children. *J Laryngol Otol.* 2007;121(5):449-54.
23. Steindl R, Kunz K, Schrott-Fischer A, Scholtz AW. Effect of age and sex on maturation of sensory systems and balance control. *Dev Med Child Neurol.* 2006;48(6):477-82.
24. Lonescu E, Morlet T, Froehlich P, Ferber-Viart C. Vestibular assessment with Balance Quest: normative data for children and young adults. *Int J Pediatr Otorhinolaryngol.* 2006;70(8):1457-65.
25. Micarelli A, Viziano A, Augimeri I, Micarelli B, Alessandrini M. Age-related assessment of postural control development: a cross-sectional study in children and adolescents. *J Mot Behav.* 2019;1-9.