



University of Nebraska at Omaha
DigitalCommons@UNO

Journal Articles

Department of Biomechanics

Spring 2007

Methods to Improve the Reliability of the Functional Reach Test in Children and Adolescents With Typical Development

Kathleen G. Volkman

University of Nebraska Medical Center

Nicholas Stergiou

University of Nebraska at Omaha, nstergiou@unomaha.edu

Wayne A. Stuber

University of Nebraska Medical Center

Daniel Blanke

University of Nebraska at Omaha, dblank@unomaha.edu

Julie Stoner

University of Oklahoma Health Sciences Center

Follow this and additional works at: <https://digitalcommons.unomaha.edu/biomechanicsarticles>

 Part of the [Biomechanics Commons](#)

Recommended Citation

Volkman, Kathleen G.; Stergiou, Nicholas; Stuber, Wayne A.; Blanke, Daniel; and Stoner, Julie, "Methods to Improve the Reliability of the Functional Reach Test in Children and Adolescents With Typical Development" (2007). *Journal Articles*. 153.

<https://digitalcommons.unomaha.edu/biomechanicsarticles/153>

This Article is brought to you for free and open access by the Department of Biomechanics at DigitalCommons@UNO. It has been accepted for inclusion in Journal Articles by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.



Methods to Improve the Reliability of the Functional Reach Test in Children and Adolescents With Typical Development

Kathleen Gerri Volkman, MS, PT, Nicholas Stergiou, PhD, Wayne Stuberg, PhD, PT, PCS, Daniel Blanke, PhD, and Julie Stoner, PhD

Munroe–Meyer Institute, University of Nebraska Medical Center (K.G.V., W.S.), School of Health, Physical Education and Recreation, University of Nebraska at Omaha (N.S., D.B.), and Department of Preventive and Societal Medicine, University of Nebraska Medical Center (J.S.), Omaha, Nebraska

Address correspondence to: Kathleen Volkman, Department of Physical Therapy, Munroe–Meyer Institute, University of Nebraska Medical Center, PO Box 985450, Omaha, NE 68198-5450. E-mail: kvolkman@unmc.edu This work was completed in partial fulfillment of a Master of Science degree program in Exercise Science for the first author.

Key words: adolescent, child, measure, musculoskeletal equilibrium, posture/physiology, reproducibility of results

Purpose: Test-retest reliability of the Functional Reach Test was examined in children with typical development by comparing standard and alternate methods.

Methods: Eighty subjects ages seven to 16 years were tested and 69 retested for four methods of Functional Reach Test (ie, one-arm finger-to-finger, two-arm finger-to-finger, one-arm toe-to-finger, and two-arm toe-to-finger). Intraclass correlation coefficients and limits of agreement were calculated.

Results: Intraclass correlation coefficients were high in toe-to-finger measurement methods (0.97–0.98) for the entire group and specific age groups (0.83–0.93). Toe-to-finger methods were more reliable than finger-to-finger methods. The two-arm toe-to-finger method had the best limits of agreement with approximately ± 5 cm indicated by the 95% confidence interval.

Conclusions: Test-retest reliability using a toe-to-finger method of measuring is stronger than previously reported when using traditional methods. Limits of agreement analyses imply a change of 5 cm or more is likely to represent a true clinical difference when using the two-arm toe-to-finger method.

INTRODUCTION

The Functional Reach Test (FRT) has been studied extensively in adults and has been found to relate to movement of the center of pressure.¹ The FRT is accepted as a clinical test for balance in the elderly population and has demonstrated high test-retest reliability in various adult populations ($r = 0.89 - 0.92$).^{1,2} In children, the FRT also has been considered a tool to measure balance. However, because children show more variability in their movements than adults and undergo changes in body proportions and size during development, the reliability of various postural control tests in children has been difficult to establish.^{3,4} Recent studies showed fair test-retest reliability (Intraclass correlation coefficient [ICC] = 0.64 – 0.75) of the FRT with typically developing children⁵ and poor reliability ($r = 0.31$) with balance-impaired children.⁶ Westcott and colleagues⁷ stated that the FRT might be useful as a discriminative test to document feed-forward mechanisms of postural control. However, because of test-retest reliability inconsistencies, they did not recommend it as an evaluative measure for change over time.

This study addresses reported reliability inconsistencies of the FRT. The effect of a biomechanical change in the performance of the FRT was examined as well as the effect of a change in the method of measurement on test-retest reliability coefficients and FRT scores. The FRT consists of subjects reaching forward with one arm as far as they can without taking a step or falling. Sources of variability in the FRT, such as trunk rotation, shoulder retraction/protraction, and base of support position, were considered. It was hypothesized that a two-arm reach would limit trunk rotation and, thereby, could improve reliability of the test. It was also hypothesized that measuring from a stationary point, such as the toes, would improve test-retest reliability compared with measuring from a nonstationary point as done in the traditional method.

The purpose of this study, therefore, was to compare the test-retest reliability of the traditional FRT in typically developing children using two alternate protocols. The first protocol was a symmetrical two-arm style of reach vs the original asymmetrical one-arm reach, and the second protocol involved measuring reach from the end of the toes

(stationary point) versus the original method of measuring reach from the end of the hand (non-stationary point).

METHODS

Subjects

Eighty children with typical development (40 boys and 40 girls) were recruited by personal contact or a letter to the parents of prospective subjects. This study was approved by the institutional review board committee at the University of Nebraska Medical Center, and informed consent was obtained from parents and subjects before participation. A parent questionnaire requesting their child's demographic information, health conditions, and current medical treatment was used to screen the subjects. Exclusion criteria included recent history of orthopedic or neurological injury or disease, current school physical therapy treatment, and lack of active ankle range of motion in standing. Subjects were divided into three age groups: seven to eight years, 11 to 12 years, and 15 to 16 years, as shown in Table 1. Sixty-nine subjects were retested on the same day or within two weeks (with four subjects retested within eight weeks due to scheduling difficulties).

Research Design

Efforts were made to protect internal validity by holding constant many variables in the environment and the task through the design of the testing apparatus, the choice of measurement tools and the consistency of instructions provided to subjects. A learning effect in subjects was minimized by the provision of only one practice trial and three measurement trials for each style of reaching. An order effect was minimized through reversing test order in half of the subjects.

Procedures

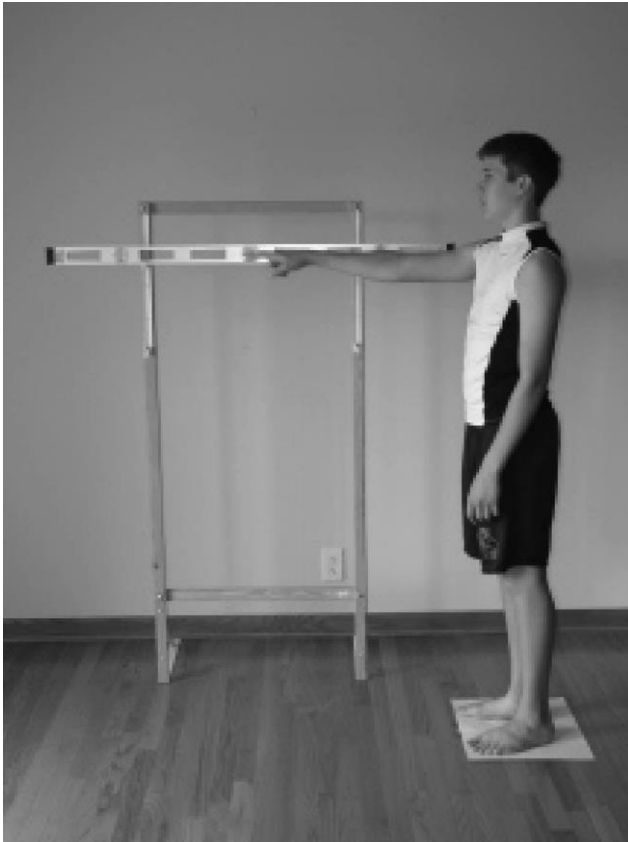
The FRT protocol is clearly defined in the literature.¹ The FRT consists of a subject reaching along a measuring stick with one arm extended with the hand in a fist, as far forward as possible without losing balance or taking a step. The score is the difference between the starting position (standing upright with 90 degrees of shoulder flexion) and the reaching position of the hand (leaning forward with the arm outstretched). In this study, the positioning of the hand for measurement was changed from a fist to a pointed index finger. This permitted more consistent measurement of the end point of the reach, as it was difficult to lay the ruler at the ends of the metacarpals in a perpendicular fashion.

For this study, a wooden frame was devised to hold a metric measuring stick (120 cm long) with a level, which was held by spring-loaded grips (Fig. 1). The measuring stick permitted reach measurements to the nearest 0.20 cm. The stick was positioned by moving it up or down, sliding the grips along their vertical metal runner, to align it with the subject's shoulder. A level was used to insure the stick was horizontally aligned. During measurement of the FRT, a ruler was used as a straight edge for marking the end of reach along the measuring stick so a numerical score could be read. Before the FRT measurements, the subject's height was measured to the nearest 0.10 cm using a metric measuring tape fixed to the wall. Then each subject stood barefoot on a 14" by 17" sheet of heavy duty rough newsprint paper, which was taped to the floor and aligned with the end of the measuring stick before the tests. A carpenter's square was used to measure the distance horizontally from the frame to the end of the stick. This measurement was repeated on the floor from the frame to the edge of the paper to align the edge of the paper with the end of the measuring stick above. The carpenter's square was also used to align the vertical components of the frame at a 90-degree angle to the floor. The subject stood on the paper with the tips of the great toes aligned with the paper's edge, and his or her feet were traced with a pen. To further insure consistency in the testing, only hard floor surfaces were used under the paper during testing. The tracings were used to insure the same foot placement for the initial tests and retests. The dominant writing hand was chosen for the reaching arm, and this arm was placed closest to the measurement device during the test. In female subjects with long fingernails, the ruler was positioned beneath the fingernail to align it with the end of the finger. Furthermore, an alternate protocol for reaching with both arms also was developed, and it was compared with the traditional one-arm FRT.

TABLE 1

Subject Characteristics (total n = 80/ retest n = 69)

Age	7–8 Years		11–12 Years		15–16 Years	
	Test	Retest	Test	Retest	Test	Retest
N	29	21	26	23	25	25
Male	13	8	13	11	14	14
Female	16	13	13	12	11	11
Mean height \pm SD (cm)	127.75 \pm 6.06	127.70 \pm 6.16	149.22 \pm 7.38	149.07 \pm 7.61	172.31 \pm 7.74	172.31 \pm 7.74

**Fig. 1.** One-arm functional reach test starting position.

To obtain the one-arm FRT score, the subject's starting position was measured, and the score was recorded. The subject leaned forward to reach as far as possible without losing balance or taking a step (Figs. 1 and 2). The subject held the reaching position for approximately three seconds while the reach position was measured to the nearest 0.20 cm. The subject was not allowed to put pressure on the ruler during the testing. To obtain the two-arm FRT score, the starting and the reaching positions were measured in a similar manner to the one-arm reach. The two arms were extended forward at 90 degrees of shoulder flexion with the hands clasped and the index fingers extended together (Figs. 3 and 4). The reach measurement was obtained at the tip of the longer index finger. In addition, an alternate method of calculating the FRT score from toes to fingers was explored. In this method, the FRT was measured as the distance between the starting point of the meter stick (vertically aligned above the tips of the great toes) and the fingertip(s) at the end of reach. For the calculation of the toe-to-finger score, there was no measuring of initial hand position as for the finger-to-finger score. The toes were aligned with the edge of the paper and the end of the measuring stick as previously described. This method was applied for both the one-arm and the two-arm reach tests.

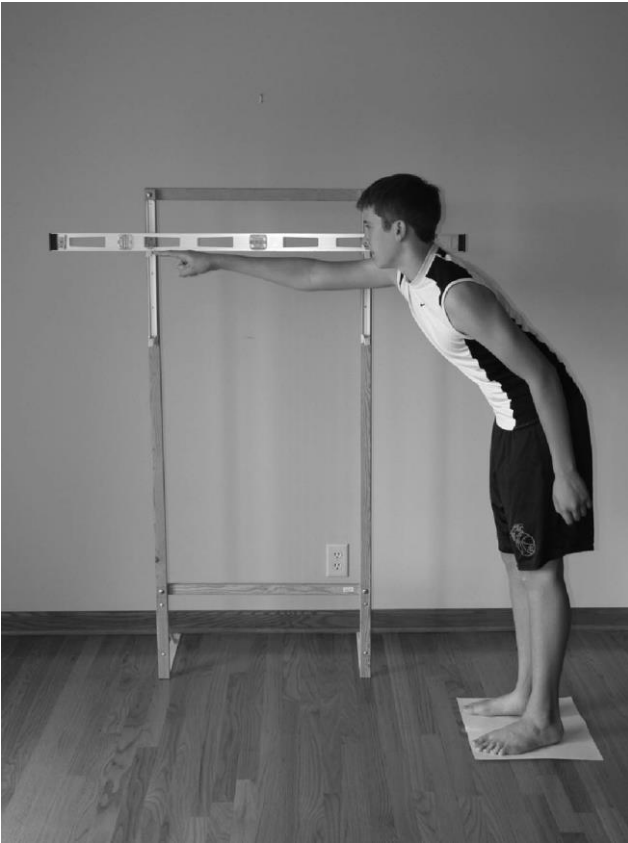


Fig. 2. One-arm functional reach test final position (example of heels down position).



Fig. 3. Two-arm functional reach test starting position.



Fig. 4. Two-arm functional reach final position (example of heels up position).

Verbal instruction was given along with a demonstration of positioning. Because it was hypothesized that the base of support could affect reliability, subjects were instructed to choose one of two strategies when performing the reach tests: either bending at the hips with feet flat or bending at the hips with heels lifted. The subjects had to select a specific strategy and maintain it for all trials. The strategy used was recorded. One practice trial of each style of reach was allowed. During the testing, the subjects were told to stand with feet a comfortable width apart, to stand straight, and to raise the arm/arms so the investigator could obtain the starting position measurement as described previously. Then, they were told to reach as far as possible without taking a step or falling and the reach measurement was taken. The subjects were given verbal encouragement for optimal performance and were reassured that the investigator would be standing close to them for safety. During the test, three trials of the one-arm reach were performed and recorded. This was repeated for the two-arm reach with similar verbal instruction and demonstration of arm and hand positioning. For each method, the trial with the greatest score of the three was used as each subject's FRT score.

The order of the presentation of the two types of the FRT was reversed for alternate subjects. In addition, for the retesting of each subject, the order of the presentation of the FRT was again reversed. These procedures were used to minimize any kind of ordering effect on the results.

Statistical Analysis

Statistical analyses were performed using SPSS 11.5 (SPSS Inc., College Station, TX) and SAS 8.02 (SAS Institute, Cary, NC) software. Means and standard deviations were reported for demographics and FRT scores. To achieve adequate power for the data analysis, we determined that a total of 26 subjects in each age group would need to be measured twice for the reliability study, assuming a one-sided 0.05 alpha level, to have 80% power to detect a true difference between the historical reliability coefficient of 0.75 and a coefficient of at least 0.90.8 Test-retest reliability of the FRT was analyzed under two styles of reach (one-arm or two-arm) and two measuring techniques (measuring from finger or from toes). Test-retest scores from the subsample of 69 subjects were analyzed under the four conditions and across the three age groups. The four conditions were: one-arm style using finger-to-finger measurement (1AFF), two-arm style using finger-to-finger measurement (2AFF), one-arm style using toe-to-finger measurement (1ATF), and two-

arm style using toe-to-finger measurement (2ATF).

The ICC frequently is used in the literature to report test-retest reliability. However, the literature on reliability has advocated the calculation of repeatability coefficients to show more clearly the level of agreement between each subject's repeated measurements expressed in actual units (cm). Therefore, Bland-Altman plots^{9,10} also were graphed to show the 95% limits of agreement, a prediction interval for the differences between repeated measurements for each of the four conditions.

The ICC [1,1] was estimated separately for each of the four measurement methods (ie, 1AFF, 2AFF, 1ATF, and 2ATF). These were based on a one-way random effects analysis of variance because the design involved only a single rater measuring each subject on two occasions. The measurement was a single value, rather than a mean of several values, so the form ICC [1,1] was used.¹¹ The 95% confidence interval (CI) for the ICC also was estimated. Informal comparisons were made among the four methods to see whether one method had higher correlation coefficients.

The agreement between repeated measures summarized through Bland-Altman plots displayed the difference (test minus retest) of each score as well as the overall average difference and the 95% limits of agreement (± 2 standard deviations) on the vertical axis by the average of the repeated measures on the horizontal axis. The effect of lag time between repeated measures, dichotomized as within the same day or not, on the difference values was investigated using a Wilcoxon rank sum test. A paired *t*-test was used to compare the mean measurements of the one-arm and the two-arm methods. The nonparametric Wilcoxon rank sum test was used to compare the distribution of differences of repeated FRT measures between groups, defined by the lag time between repeated measurements, because of the small number of subjects with a lag of at least one day.

TABLE 2

Reliability Coefficients of Four Methods of the FRT (n = 69)

Method	ICC [1,1]	95% CI
One-arm, finger to finger	0.83	0.73–0.89
One-arm, toe to finger	0.97	0.95–0.98
Two-arm, finger to finger	0.83	0.74–0.89
Two-arm, toe to finger	0.98	0.96–0.99

TABLE 3

Reliability Coefficients of Four Methods of FRT by Age Group (n = 69)

Method	ICC [1,1]	95% CI
One-arm, finger to finger		
Age 7–8 years	0.39	0–0.70
Age 11–12 years	0.59	0.25–0.80
Age 15–16 years	0.87	0.73–0.94
One-arm, toe to finger		
Age 7–8 years	0.83	0.63–0.93
Age 11–12 years	0.89	0.76–0.95
Age 15–16 years	0.90	0.79–0.95
Two-arm, finger to finger		
Age 7–8 years	0.60	0.25–0.82
Age 11–12 years	0.50	0.13–0.75
Age 15–16 years	0.81	0.62–0.91
Two-arm, toe to finger		
Age 7–8 years	0.86	0.68–0.94
Age 11–12 years	0.88	0.74–0.95
Age 15–16 years	0.93	0.85–0.97

RESULTS

Results of the correlation analysis for the entire sample are presented in Table 2. All methods demonstrated

greater reliability than the values reported in the literature for children, which historically were no greater than 0.75. Both toe-to-finger methods revealed higher ICC values when compared with the finger-to-finger, regardless of the style of reach used, ie, one arm or two arms. The finger-to-finger methods were only marginally better than the historical values as indicated by the lower end of the 95% CI.

The ICC and 95% CI for each of the four methods, subdivided by age group, are listed in Table 3. The reliability was lower in the seven- to eight-year-old group, most notably under the 1AFF method. However, toe-to-finger methods had greater reliability coefficients in the ages seven- to eight-year-old group and the 11- to 12-year-old groups compared with finger-to-finger methods. The reliability coefficients had higher values in the 15- to 16-year-old group in all methods, particularly in the 2ATF method.

The Bland–Altman limits of agreement plots are included in Figures 5 through 8. In these plots, the average difference between the repeated measurements (solid line) and the 95% CI for a randomly chosen difference between repeated measures (dotted lines) is presented. The average difference between repeated measures for the 1AFF method was 1.04 cm with limits of agreement ranging from -7.34 to 9.42 cm. For the 1ATF method, the average difference between repeated measures was 0.53 cm with limits of agreement ranging from -5.47 to 6.53 cm. It is, therefore, reasonable to expect the premeasure to be anywhere from 7.34 cm less than the postmeasure to 9.42 cm greater than the postmeasure. The average difference between repeated measures for the 2AFF method was 0.08 cm, with limits of agreement ranging from -8.21 to 8.36 cm. Finally, in the 2ATF method, the average difference of 0.09 cm was found with limits of agreement ranging from -4.88 to 5.06 cm. Under all methods, agreement was fairly constant across the age groups.

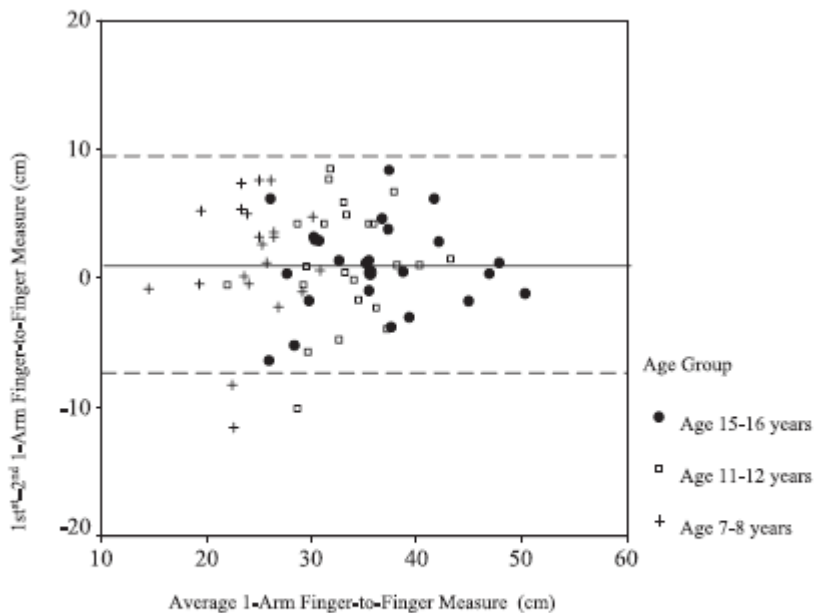


Fig. 5. Bland–Altman plot of limits of agreement for 1AFF. Solid line represents mean difference between tests one and two. Dotted lines represent mean difference ± 2 SD.

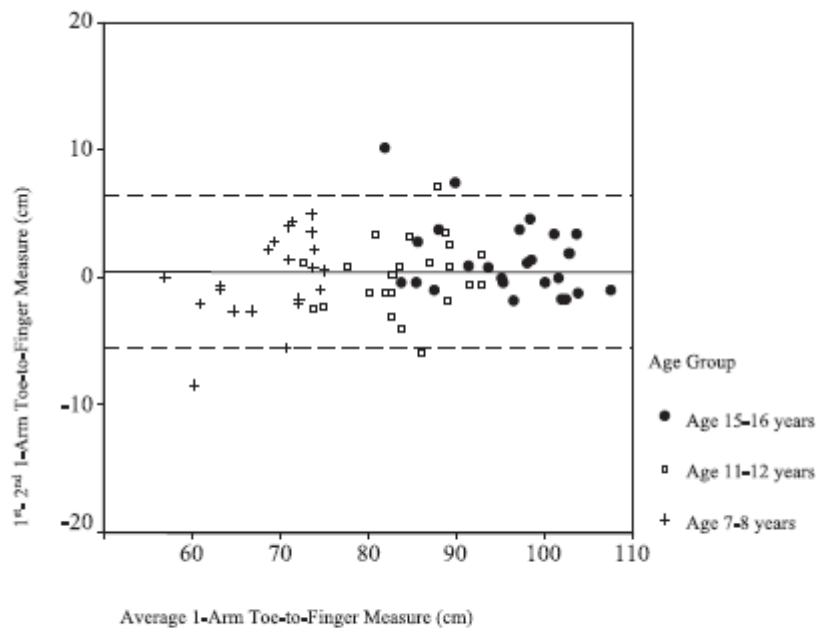


Fig. 6. Bland-Altman plot of limits of agreement for 1ATF. Solid line represents mean difference between tests one and two. Dotted lines represent mean difference ± 2 SD.

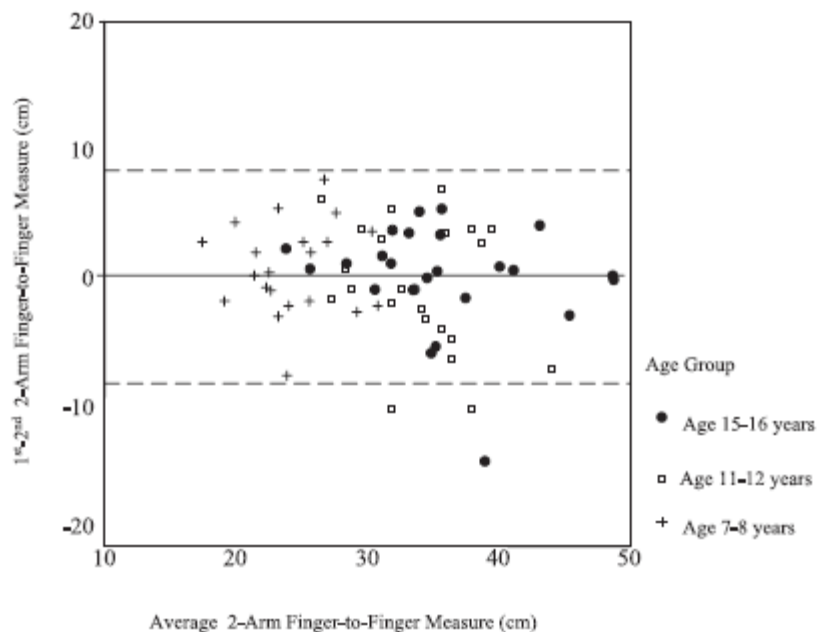


Fig. 7. Bland-Altman plot of limits of agreement for 2AFF. Solid line represents mean difference between tests one and two. Dotted lines represent mean difference ± 2 SD.

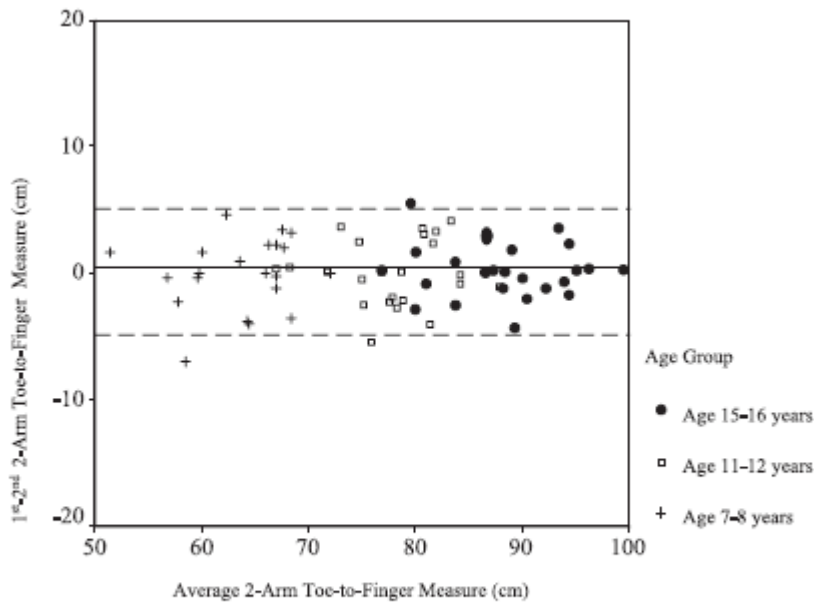


Fig. 8. Bland–Altman plot of limits of agreement for 2ATF. Solid line represents mean difference between tests one and two. Dotted lines represent mean difference ± 2 SD.

TABLE 4

Descriptive Statistics of Reach Scores at Initial Test by Method for Each Age Group

	7–8 years (n = 29)	11–12 years (n = 26)	15–16 years (n = 25)
1AFF (cm)			
Mean, (SD)	24.69 (4.52)	33.68 (5.08)	36.52 (7.09)
1ATF (cm)			
Mean, (SD)	69.39 (6.08)	84.29 (6.20)	95.95 (6.57)
2AFF (cm)			
Mean, (SD)	24.35 (4.13)	33.98 (4.28)	35.83 (6.17)
2ATF (cm)			
Mean, (SD)	63.73 (5.12)	77.88 (5.47)	88.42 (5.82)

Means and standard deviations for each method under each age group are listed in Table 4. It was noted that the highest scores actually were distributed among the three trials performed by the subjects, rather than consistently being the first or the last trial. The time between repeated measures ranged from zero to 52 days. Fifty-four (78%) had repeated measurements taken on the same day. Fifteen (22%) subjects had a lag time between visits that ranged from one day to 52 days, with a median lag of eight days. Only four (27%) of these 15 subjects had lag times longer than two weeks. No significant differences were found between the subjects with a lag of zero days and subjects with a lag of at least one day, in terms of the repeatability of the FRT scores ($p > 0.6$ for each method; using the Wilcoxon rank sum test).

The hypothesis that the two-arm reach score should be less than the one-arm reach score was not supported by the data acquired from the finger-to-finger measures. Table 4 shows that in the toe-to-finger measures 1ATF scores were approximately 6 cm greater than 2ATF scores. However, there were no significant differences between 1AFF and 2AFF scores ($p = 0.9$, paired t test). A possible explanation for this result is that the starting position might move backward during the two-arm reach. As the two arms were raised and the center of mass subsequently moved forward,

the center of pressure would shift further backward as compared with the one-arm reach. The traditional finger-to-finger method of measurement would not reflect this sway. To investigate this explanation, further analysis was performed to detect differences between the starting position in the 1AFF and the 2AFF in each subject. The distance between the end of the measuring stick vertically aligned over the tips of the great toes and the starting location of the finger was calculated as the difference between the toe-to-finger and finger-to-finger reach measurements for the one-arm and two-arm styles. A paired t-test was used to compare the mean measurements between the one-arm and the two-arm methods. Results of this analysis showed that the mean distance was greater for the one-arm method as hypothesized (Table 5; $p < 0.001$ for first and second measures). For the first-time measurements, the mean starting position distance was 6.25 cm less for the two-arm reach method as compared to the one-arm reach and 6.64 cm less for second-time measurements. Therefore, the results supported the hypothesis that the starting position of the hand moved consistently further backward with the two-arm style of reach even though this was not reflected in the finger-to-finger scores.

TABLE 5
Results of Paired Samples Test of Difference Between One-Arm and Two-Arm Starting Position

Pairs	Mean (cm)	SD	SEM	95% CI	t	d	p Value
One-arm starting position –two-arm starting position (first test)	6.25	4.25	.47	5.30–7.19	13.15	79	<0.001
One-arm starting position –two-arm starting position (second test)	6.64	3.82	.46	5.72–7.56	14.45	68	<0.001

DISCUSSION

The purpose of this work was to analyze test-retest reliability coefficients of FRT scores under four different conditions. It was hypothesized that the reliability of the traditional one-arm reach measured finger-to-finger would be different from the alternate protocols of measuring from toes to finger and the two-arm protocol. The results showed that all methods had greater ICC coefficients than the results reported in the literature⁵ for children and were closer to those reported for adults.^{1,2} The toe-to-finger measurement methods were more reliable than the finger-to-finger, regardless of the style of reach used, ie, one arm or two arms. Furthermore, the 2AFF method was only marginally better than the 1AFF method, indicated by the lower end of the confidence interval (Table 2). Considering the sources of variability in the FRT, it had been assumed that a two-arm reach would limit trunk rotation and, thereby, improve reliability of performance. This hypothesis was not supported by the results obtained using the ICC statistical approach. The limits of agreement evaluation, however, showed that the 2ATF had the smallest average differences between the repeated measures (0.09 cm). On the basis of this evaluation, it can be suggested that the toe-to-finger methods are more reliable than the finger-to-finger methods when measuring FRT in children.

The FRT methods were also analyzed by age group. Since similar protocols were used to measure the FRT (ie, tracing the feet, stance chosen by subject), reliability for the 1AFF method was expected to be similar to the results by Donohoe et al.⁵ In that study, reliability coefficients ranged from 0.64 to 0.75 in a sample with fewer subjects and a younger age range. Coefficient results in the current study ranged from 0.39 to 0.87 with a larger sample. It was found that the 2ATF method had higher coefficients (Table 3) for all age groups and that the 1ATF and the 2ATF were approximately the same. The seven- to eight-year-old group had lower reliability coefficients as compared with the other two age groups in three of the four methods. The 11- to 12-year-old group also had lower reliability coefficients as compared with the 15- to 16-year-old group, particularly in the finger-to-finger method. This result is consistent with data on postural control reported by Hay and Redon,¹² where young children six to eight years of age exhibited more variability in the amplitude of the center of pressure displacement during self-initiated movements as compared with older children and adults. Therefore and particularly in younger children, using a reliable method of FRT is important for measuring change over time. Using a protocol with a stationary starting point for the FRT, rather than the traditional method of starting at the end of the arm, could result in improved reliability.

It should be noted that the ICC is influenced by the amount of variability between the subjects. When there is more variability across subjects, the ICC will generally be larger. When the height statistics were examined, the SD of the oldest age group was slightly higher than the other groups (Table 1). Because height has been found to correlate with reach distance,¹³ this variable may have been an influential factor.

The 2AFF reliability coefficients were lower than the 1AFF coefficients in the two older age groups, indicating

that using a two-arm strategy may not result in improved reliability of the FRT as hypothesized. This implies that a large source of variability may not arise from the trunk position. It can be concluded from the analysis on the starting hand position that the starting center of pressure (where the weight is distributed on the feet at the beginning of the test) may affect reliability each time the test is performed. No literature was found defining how a subject should stand for the FRT other than to stand comfortably. When the 2AFF method is used, the shift of the center of pressure backward is more noticeable. If the traditional 1AFF method is used, a less noticeable shift occurs, but nonetheless, the shift may be different each time the test is done. When using the toe-to-finger method of measurement, this variable would not affect the reliability of the scores.

If the FRT is to be used in children to measure change in postural control, it is important to know with reasonable certainty that an increase in score is the result of improved balance and not to error. The toe-to-finger methods had the best agreement between measures. The 2ATF method had the best limits of agreement of the four methods with approximately ± 5 cm indicated by the 95% confidence interval. This means that a second measurement, using this method, could reflect a real change in reaching ability if the measurement exceeded 5 cm in this population. The 2ATF and the 1ATF methods both provide higher reliability coefficients and limits of agreement compared to the finger-to-finger methods.

Mean results for 1AFF (traditional test) scores were 24.69 cm in the seven- to eight-year-old group and 36.52 cm in the 15 to 16 year old group. This compares similarly to the published mean of 24.21 cm in seven to eight year olds and is slightly greater than the mean of 32.30 cm for 13 to 15-year-old youths.⁵ Duncan et al¹ published mean FRT (1AFF) scores of 36.60 cm for young adult females and 41.83 cm for males. Hageman et al¹⁴ published scores of 43.33 cm in young adults of both sexes. In this study, the 15- to 16-year-old group had a mean 1AFF reach of 36.52 cm. The median 1AFF values by gender for the oldest group (15 to 16 years) were 32.20 cm for females and 38.80 cm for males.

The subjects in this study were children with typical development. However, balance testing is more likely to be performed on children with impairments that affect their balance. Further study of the toe-to-finger method of the FRT in these populations would yield correlation coefficients or limits of agreement more appropriate to the patients seen in pediatric practice. Also, additional research could determine if directing children to place their weight on their heels and thus moving the center of pressure further backward prior to measuring the FRT would decrease variability in scores.

CONCLUSION

In the present study using toe-to-finger FRT measurement methods and without changing the biomechanics of the test, reliability coefficients improved over previously reported values for children who are developing typically. Correlation coefficients were similar between the 1ATF and 2ATF methods. However, of the four methods, the 2ATF method had the best limits of agreement of ± 5 cm around the mean.

Toe-to-finger measurement methods have the advantage of using a stationary starting point rather than an inconsistent one. This method removes variability resulting from sway and initial shoulder position. It would be no more difficult to perform in the field than the traditional method. Tracing the base of support is recommended for increasing reliability of repeated measures and this can be accomplished simply enough with paper and tape. For the school-based therapist, a plumb line could be used to mark the toe position on a chalkboard, as well as the reach position, to measure functional reach in school children. It may also be important to measure the extended arm with a pointed index finger rather than a fist and to instruct children to use the same lower extremity strategy each time the test is performed. Because the FRT is also used in children with balance impairments, further research into the reliability of FRT is needed in such populations.

REFERENCES

1. Duncan PW, Weiner DK, Chandler J, et al. Functional reach: a new clinical measure of balance. *J Gerontol.* 1990;45:M192–M197.
2. Frzovic D, Morris ME, Vowels L. Clinical tests of standing balance: performance of persons with multiple sclerosis. *Arch Phys Med Rehabil.* 2000;81:215–221.
3. Shumway-Cook A, Woollacott MH. The growth of stability: postural control from a developmental perspective. *J Mot Behav.* 1985;17:131–147.

4. Wolff DR, Rose J, Jones VK, et al. Postural balance measurements for children and adolescents. *J Orthop Res.* 1998;16:271–275.
5. Donahoe B, Turner D, Worrell T. The use of functional reach as a measurement of balance in boys and girls without disabilities ages 5 to 15 years. *Pediatr Phys Ther.* 1994;6:189 –193.
6. Pellegrino TT, Buelow B, Krause M, et al. Abstract. Test-retest reliability of the Pediatric Clinical Test of Sensory Interactions for Balance and the Functional Reach Test in children with standing balance dysfunction. *Pediatr Phys Ther.* 1995;7:197.
7. Westcott SL, Lowes LP, Richardson PK. Evaluation of postural stability in children: current theories and assessment tools. *Phys Ther.* 1997;77:629 – 645.
8. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17:101–110.
9. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician.* 1983;32:307–317.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307–310.
11. Shroud PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 86:420 – 428.
12. Hay L, Redon C. Feedforward versus feedback control in children and adults subjected to a postural disturbance. *Brain Res.* 1999; 125:153–162.
13. Habib Z, Westcott S. Assessment of anthropometric factors on balance tests in children. *Pediatr Phys Ther.* 1998;10:101–109.
14. Hageman PA, Liebowitz M, Blanke D. Age and gender effects on postural control measures. *Arch Phys Med Rehabil.* 1995;76:961– 965.