開発した改訂版の単純検査（改訂版STEFT）の構造、信頼性、妥当性、および応答性

著者：入江 啓輔

学位授与番号：学位甲第342号
学位名：博士（保健学）
Development of the modified Simple Test for Evaluating Hand Function

(modified STEF):

Construct, reliability, validity, and responsiveness

Keisuke Irie OTR, MSc\textsuperscript{a,b}, Hirokatsu Iseki OTR\textsuperscript{a,b}, Satomi Okamoto OTR, MSc\textsuperscript{a},

Kazuhiro Okamoto RST, MSc\textsuperscript{a}, Seiji Nishimura OTR, PhD\textsuperscript{c}, and Kenji Kagechika MD\textsuperscript{d}

\textsuperscript{a}Department of Rehabilitation Medicine, Kanazawa Medical University Hospital

\textsuperscript{b}Division of Health Sciences, Graduate School of Medical Sciences, Kanazawa University

\textsuperscript{c}School of Health Sciences; College of Medical, Pharmaceutical and Health Sciences;

Kanazawa University

\textsuperscript{d}Department of Rehabilitation Medicine, Kanazawa Medical University

Corresponding author:

Keisuke Irie, MSc

Department of Rehabilitation Medicine, Kanazawa Medical University Hospital

1-1 Daigaku, Kahoku-gun uchinadamachi, Ishikawa, 920-0293 Japan

Tel & Fax: +81-076-286-3511

E-mail: keisuke@kanazawa-med.ac.jp
Abstract

Study Design: Clinimetric evaluation study.

Introduction: Despite the availability of numerous performance tests to measure finger dexterity, there is no international consensus on hand function evaluation.

Purpose of the Study: To evaluate the reliability, validity and responsiveness of the modified version of the Simple Test for Evaluating Hand Function (STEF), which is widely used in Japan.

Methods: The intra-rater (n=40) and inter-rater (n=32) reliability of the modified STEF was evaluated by calculating the intraclass correlation coefficient, models (1,1) and (2,1), respectively, in healthy individuals. The criterion validity of the modified STEF (n=50) was evaluated by calculating the Pearson’s correlation coefficient relative to the STEF, the Purdue Pegboard Test (PPT) and the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH). The standardized response mean of the scores was calculated to determine responsiveness (n =35). The modified STEF was used prospectively to measure the change in hand function in a cohort of patients with hand trauma injuries and inflammatory diseases (n = 30), as well as in a cohort of patients with cervical spondylosis (n = 20), from preoperative baseline to 1 and 3 months postoperatively.

Results: ICC_{1.1} and ICC_{2.1} values were ≥0.80, indicative of high intra- and inter-rater
reliability. All correlation coefficients were significant (p<0.05): STEF (r = 0.89); PPT (r = 0.69); and DASH (r = -0.34). The standardized response mean indicated greater responsiveness of the modified STEF (0.89) than the STEF (0.71) and PPT (0.68), but a lower responsiveness than the DASH (1.11).

Conclusions: The modified STEF is a reliable measurement tool, with a moderate positive correlation with the PPT and a greater responsiveness than the STEF and PPT.

Key words: modified Simple Test for Evaluating Hand Function, Reliability, Validity, Responsiveness

Level of Evidence: Not applicable.
1. Introduction

The assessment of general hand function in patients having sustained hand trauma or in those with an inflammatory condition is complex. Many instruments have been developed to quantify general hand function, which is defined as the nature and extent of daily activities that an individual can perform using his or her hands.\textsuperscript{1,2} Traditionally, outcome assessments in hand therapy have focused on the structure and function of the hand, including range of motion, grip strength and sensation. However, since the introduction of the International Classification of Functioning, Disability and Health (ICF) by the World Health Organization (WHO) in 2001, there has been a shift in clinical assessments, from a focus on disability to functioning and health, with assessment of activities and participation providing the primary clinical outcomes. Within this framework, activity is defined as ‘the execution of a task or action by an individual’, whereas participation is defined as ‘a person’s involvement in a life situation’. Fine motor function of the hand is essential to the performance of meaningful activities, including the coordinated actions of handling objects, such as grasping, manipulation and release.\textsuperscript{3,4} With the focus on activities and participation,\textsuperscript{5,6} hand therapists are increasingly using performance tests (such as the Nine-Hole Peg test,\textsuperscript{7} Purdue Pegboard Test (PPT),\textsuperscript{8} Jebsen-Taylor Hand Function test (JTT)\textsuperscript{9}) to evaluate activity limitations, and patient-reported outcomes (such as the Disabilities of the Arm, Shoulder and
Hand (DASH) and the Michigan Hand Outcomes Questionnaire (MHQ), to evaluate participation restrictions. However, Van de ven-Stevens et al. reported that none of the existing measurement tools for the assessment of hand-related activities and participation satisfy all clinimetric properties (reliability, validity and responsiveness) required for clinical application. Furthermore, previous studies have reported that the JTT should not be used to evaluate the therapeutic effect of a surgical intervention due to concerns with its responsiveness and validity. As such, there is no international consensus regarding the appropriate tools to use in clinical practice for the assessment of hand activities.

In Japan, the Simple Test for Evaluating Hand Function (STEF) is widely used to evaluate hand-related limitations in activities. The STEF was developed by an occupational therapist in 1969, with its rating system adapted for clinical practice in 1986. The STEF specifically assesses the accuracy, smoothness, speed, and dexterity of voluntary movements of the upper extremity, and the hand more specifically. It is composed of 10 tasks, involving 58 objects, and includes the assessment of various grip patterns, including the precision grip and power grip. The STEF scores the time taken to move ten objects, of different shapes and sizes, to predetermined locations, with scores from 1-10 used to quantify the time requirement, the time scores summed to provide a total test score.

The reliability of the STEF was established by Kaneko during its initial development.
terms of validity, the STEF has a high correlation with the Action Research Arm Test, which is widely used for the assessment of upper limb function in patients post-stroke.\textsuperscript{17} With respect to hand therapy, the STEF correlates with the range of wrist motion and the DASH in patients with a distal radius fracture or other wrist joint diseases.\textsuperscript{18,19} In our previous study, we determined that the responsiveness of the STEF was sufficient to support its use as a measure of clinical change after a therapeutic intervention.\textsuperscript{20} However, the difficulty level and the evaluation criteria of the examination tasks may be limiting factors in patients with mild impairment in hand function. To address this issue, we developed the modified STEF to include assessment items requiring a higher degree of difficulty of hand function. Therefore, our aim in this study was to evaluate the reliability, criterion validity and responsiveness of the modified STEF. We hypothesized that there would be a positive correlation between the modified STEF and the PPT, and that the responsiveness of the modified STEF would be greater than that of the STEF and PPT.

2. Material and Method

2.1 Equipment Construction Standards

The STEF consists of a rectangular assessment board (40 x 80 cm, height, 3.5 cm) with 10 different tasks, using objects of different size and shape: item 1 (large ball); item 2 (medium
sphere); item 3 (large rectangle); item 4 (neutral direction); item 5 (wooden disk); item 6 (small cube); item 7 (cloth); item 8 (gold disc); item 9 (small ball); and item 10 (pin). The structure of the assessment board and the contents of the test are shown in Figure 1 and described in Table 1, with the various shapes and sizes of objects used shown in Figure 2.

2.1 Modified STEF

The modified STEF is a performance-based test for patients with hand trauma and inflammatory diseases. The differences in the assessment items between the STEF and the modified STEF are summarized in Table 2. The specific differences between the modified STEF and the STEF were as follows (Figure 3): item 3, increase in weight of object from 200 g to 600 g; item 7, change in surface from vinyl to polypropylene; item 8, decrease in thickness of the object from 2.0 mm to 1.5 mm; item 9, decrease in the object diameter from 6.0 mm to 4.0 mm; and item 10, a decrease in diameter from 3.0 mm to 1.0 mm.

2.2 User’s Guide

The following materials are needed to perform the assessment: the modified STEF board and contents; a desk to place the board on; the examiner’s handbook; a stopwatch; and a pen or pencil. The distance between the patient and the desk are adjusted such that the patient’s hands are reasonably placed at the start position. The following instructions are given to the patient ‘Please perform this test as quickly as you can, doing so in a
comfortable manner’. At the start of the test, the patient places his or her hand at the center of the board. Each subtest is started using the following instructions: ‘ready, go’. The time is stopped when the object is placed in the intended position. All subtests have a time limit, after which point the assessment is stopped. If the patient drops an object or places it in the wrong location on the board, that particular subtest is restarted. However, a maximum of three attempts for each subtests is allowed. If the patient fails on the third attempt, the item is recorded as ‘incomplete’.

2.3 Scoring

The time required to complete each subtest is measured using a stopwatch, from the ‘go’ signal to the correct placement of the object on the board, and the time is recorded (Table 3). The time-to-completion for each item is then scored from ‘1’ to ‘10’, using the cutoff times established for 100 healthy individuals, 50 males and 50 females, 18 to 24 years of age. The time score for all items are summed to yield the total test score, with a maximum possible score of ‘100’.

2.4 Reliability estimates

Reliability is defined as the ability to evaluate similar patient characteristics equally at different time points, and is usually assessed using a test-retest analysis. The intra-class correlation coefficient (ICC), and its 95% confidence interval (CI), were used to evaluate
inter- and intra-rater reliability. Inter-rater reliability (ICC1,1) was evaluated in 40 healthy adults (male: 23, female: 17, age: 20-30 years), using the total score of the modified STEF. Intra-rater reliability (ICC2,1) was evaluated in 32 healthy adults (male: 12, female: 20, age: 18-27 years), again using the total score of the modified STEF.

For absolute reliability, the standard error of mean (SEM) and minimal difference change (MDC) were calculated. The SEM is calculated as the square root of the within participant variance: SEM = SD\sqrt{1-ICC}. The MDC refers to the minimal change measured by an assessment that falls outside of the measurement error of the instrument used, and was calculated as follows: MDC = SEM \times \sqrt{2} \times 1.96.

**Validity and Responsiveness of the test**

Validity is the ability to measure an outcome precisely by making a comparison to a reference, or gold, standard, usually an instrument that has been previously validated. Responsiveness is the ability of a test to detect clinical change, which is typically measured by calculating the effect size or standardized response mean (SRM). To measure the criterion validity and responsiveness of the modified STEF, we collected data from a prospective cohort of patients. The modified STEF was administered by three therapists, each having 4 to 10 years of experience in hand therapy. In addition, the study was supervised by a certified hand therapist from the Japan Hand Therapy Society.
Patients with cervical spondylosis myelopathy were evaluated from the preoperative stage up to discharge. Patients with hand trauma, such as fractures of the distal radius and fingers, were evaluated immediately after surgery and up to three months postoperatively. Patients older than 20 years of age and with no history of other surgical procedures were included. Patients were excluded from the analysis if they had a medical history of rheumatoid arthritis or osteoarthritis of the hands or other serious complications. After screening for the exclusion criteria, 50 patients were included in the evaluation of validity estimates (Table 4), with 35 included in the evaluation of responsiveness estimates (Table 5). Pearson’ correlation coefficient was used to evaluate the relationship between the total score of the modified STEF and that of the STEF, PPT and DASH. Analyses were performed using the Statistical Package for the Social Sciences (SPSS), version 18 (SPSS, Inc., Chicago, Illinois, USA), and the significance level was set to 5%.

The SRM of the clinical change in hand function as a result of an intervention was calculated as the mean change in score divided by the standard deviation of the score change, and was used as the measure of responsiveness. We calculated the SRM for the modified STEF, as well as for the STEF, PPT, and DASH, using the change score from baseline to the final follow up.

Ethical consideration
Verbal and written information were provided to all participants before enrollment, with all participants providing their consent. This study was approved by the Kanazawa Medical University Hospital Expert Committee on University Research Ethical Evaluation.

3. Results

Reliability

The inter-rater reliability was high, with an ICC$_{1,1}$ value of 0.89 (95% CI: 0.81-0.95), and a SEM and MDC of 0.26 and 0.72, respectively, for the right hand, 0.94 (95%CI: 0.89-0.97) and the SEM and MDC were 0.12 and 0.33, respectively, for the left hand. The intra-rater reliability was also high, with an ICC$_{2,1}$ value of 0.95 (95% CI: 0.91-0.98), and a SEM and MDC of 0.28 and 0.78, respectively, for the right hand, 0.86 (95% CI: 0.74-0.93) and the SEM and MDC were 0.39 and 1.08, respectively, for the left hand.

Validity

The average score for the modified STEF, STEF, PPT, and DASH were as follows: 85.3±8.5, 92.1±5.8, 8.7±3.5, and 34.6±17.9, respectively. The correlation ($r$) values between the modified STEF and the other assessments were as follows: STEF, $r = 0.89$, p<0.05 (Fig 4); PPT, $r = 0.62$, p <0.05 (Fig 5); and DASH, $r = -0.34$, p<0.05 (Fig 6).

Responsiveness
The mean total scores for the modified STEF, STEF, PPT, and DASH, with the associated SRM, for the affected hand are listed in Table 6. The SRM (to detect a meaningful clinical change) was higher for the modified STEF (0.89) than for the STEF (0.71) and PPT (0.68), but was lower than the SRM for the DASH (1.11).

4. Discussion

Dexterity, defined as "the skillful and controlled manipulation of a tool or an object by the fingers,"\textsuperscript{21, 22} is considered essential to the successful performance of tasks of daily living, work, school, play, and leisure,\textsuperscript{23, 24} and is an important component of a comprehensive assessment of upper extremity function. Despite the fact that many performance tests have been developed internationally, there is no consensus on the ideal test to assess hand function in clinical practice. The STEF is robust to effects of cultural background owing to the variety of assessment items and grasping styles that it incorporates, with few elements being linked to specific activities of daily living which, to a great extent, are culturally determined. However, previous studies have indicated the need to pay attention to the ceiling effect of the STEF in patients with mild impairment in hand function.\textsuperscript{19} In our previous study, the STEF score tended to be higher than the score in these previous studies, particularly among individuals with mild impairment in hand function.\textsuperscript{20} On the basis of these results, we considered the possibility that the difficulty level and evaluation criteria of the test
items might be limiting factors of the clinical utility of the STEF. To address this limitation of the STEF, we developed the modified STEF by increasing the difficulty of selected test items (3, 7, 8, 9, 10) to improve the discrimination of the evaluation criteria. In this study, we evaluated the clinimetric properties of the modified STEF by evaluating test scores across a larger group of patients who had undergone surgical treatment.

We determined that the reliability of the modified STEF is almost perfect when the test is performed by different evaluators within a short time-frame and when repeated by the same evaluator within a 7-day time frame. However, the 95% CI of the ICC$_{2,1}$ ranged between 0.74-0.93, indicative of variation in the scoring of the modified STEF by different evaluators. This inter-rater variation in scoring can be explained by the absence of a clear indication for repeating a trial in the modified STEF. In the STEF, an assessment item is repeated when the test object is dropped. However, in the modified STEF, as individuals had better hand function, errors in performance were not due to “dropping” the test object but rather due to “pushing” the object to the correct location on the board. The extent to which “pushing” rather than “placing” the test object on the correct location was tolerated might have varied between examiners. Therefore, clarification of this criterion of the performance is needed to provide examiners direction as to when a test item should be repeated.

To evaluate validity, we chose to use the PPT and DASH as opposed to other hand
specific measures of disability for several reasons. First, the inter-rater reliability and construct validity of the PPT have been established.\textsuperscript{26, 27} In addition, at present, the PPT is recommended in clinical practice because it includes both bilateral and unilateral hand use, and has a broad age range of normative data available. The DASH also has demonstrated reliability, validity and responsiveness.\textsuperscript{28} It is the most extensively studied assessment tool of the upper limb, with large supporting evidence of its good clinimetric quality. We identified a moderate correlation between the modified STEF and PPT. This lower than expected correlation might be explained by differences in test items between these two assessment tools. Similar to the PPT, the modified STEF includes the task of manipulating a pin with the thumb, index finger and middle finger. However, the modified STEF also includes subtests with objects of different sizes and shapes, with these tasks requiring the use of different unilateral grasps that use all the digits of the hand, as well as incorporating tasks that require anterior and posterior rotation of the forearm. The PPT may be used for a patient with an occupation that requires integration of fine motor skills and bimanual activity. However, the modified STEF should be used for a patient who uses a variety of unilateral grips.

We also identified a low correlation between the modified STEF and the DASH. These two assessment tools are very different, with the modified STEF being a criterion-based,
standardized performance tool, while the DASH is a self-report questionnaire which evaluates an individual’s perception of his or her capacity to perform an activity in a real life environment. Because of these differences, the two types of tools should complement each other when measuring someone’s activity and participation level, but may not correlate as they are measuring different underlying constructs.

We used the SRM as an indicator of responsiveness. According to the evaluation criteria proposed by Cohen, a SRM >0.80 is indicative of a larger responsiveness. As such, the SRM of 0.89 indicates good responsiveness of the modified STEF, which was comparable to the SRM of 1.1 that we calculated for the DASH in our study, the latter value being comparable to previously reported values of 0.7 to 1.37 for the DASH.

The STEF is limited in its ability to detect subtle changes in performance among patients with sufficiently good hand function as the test items used can be performed with different grips. As such, impairments in one finger can be effectively compensated by the use of an alternate type of grip using uninvolved fingers. Therefore, the test items of the STEF may not be sufficiently sensitive to evaluate more subtle limitations in ROM, sensation and muscle strength. In contrast, the test items of the modified STEF cannot easily be compensated for by using an alternative grip and, therefore, subtle impairments will affect the test score. This differentiation is improved by the reference scores of over 100 men and
women, which are provided with the modified STEF.

5. Limitation

In this study, we evaluated the correlation and responsiveness of the modified STEF with established clinical tools, namely the STEF, PPT, and DASH, for patients in whom hand function was impaired by trauma, inflammatory disease and cervical spondylosis myelopathy. However, the sample size was small and included a variety of etiologies. This may have influenced our evaluation of the correlation of the modified STEF to established assessment and of its responsiveness. Future studies are warranted to evaluate the responsiveness of the modified STEF for specific conditions, such as distal radius fracture, carpal tunnel syndrome.

6. Conclusion

The modified STEF is a reliable measurement tool, with a moderate positive correlation with the PPT and a greater responsiveness than the STEF and PPT. Based on our results, we propose that the modified STEF has sufficient the reliability, criterion validity and responsiveness for use in clinical practice.
Sources of Funding:

No funding agency played any role in study design, collection of data, analysis, or interpretation of results, drafting or finalizing the manuscript, or journal submission for manuscript publication.

Conflict of Interest:

The authors have no conflicts of interest to declare.

Acknowledgements:

We would like to thank both Katsumi Inoue and Norio Kawahara for his assistance and guidance in this research.
References


8) Costa LD, Vaughan HG, Levita E, Farber N. Purdue Pegboard as a predictor of the


33) Gay RE, Amadio PC, Johnson JC. Comparative responsiveness of the disabilities of the arm, shoulder, and hand, the carpal tunnel questionnaire, and SF-36 to clinical change after carpal tunnel release. *J Hand Surg Am*. 2003;28:250-254.

34) MacDermid JC, Tottenham V. Responsiveness of the disability of the arm, shoulder, and hand (DASH) and patient-rated wrist/hand evaluation (PRWHE) in evaluating change after hand therapy. *J Hand Ther*. 2004;17:18-23.