Video Article

Computerized Adaptive Testing System of Functional Assessment of Stroke

Gong-Hong Lin[1], Yi-Jing Huang[1], Yeh-Tai Chou[2], Hsin-Yu Chiang[3], Ching-Lin Hsieh[1,4,5]

[1] School of Occupational Therapy, College of Medicine, National Taiwan University

[2] Research Center for Psychological and Educational Testing, National Taiwan Normal University

[3] Department of Occupational Therapy, College of Medicine, Fu Jen Catholic University

[4] Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital

[5] Department of Occupational Therapy, College of Medical and Health Science, Asia University

Correspondence to: Ching-Lin Hsieh at clhsieh@ntu.edu.tw

Abstract

The computerized adaptive testing system of the functional assessment of stroke (CAT-FAS) can simultaneously assess four functions (motor functions of the upper and lower extremities, postural control, and basic activities of daily living) with sufficient reliability and administrative efficiency. CAT, a modern measurement method, aims to provide a reliable estimate of the examinee's level of function rapidly. CAT administers only a few items whose item difficulties match an examinee's level of function and, thus, the administered items of CAT can provide sufficient information to reliably estimate the examinee's level of function in a short time. The CAT-FAS was developed through four steps: (1) determining the item bank, (2) determining the stopping rules, (3) validating the CAT-FAS, and (4) establishing a platform of online administration. The results of this study indicate that the CAT-FAS has sufficient administrative efficiency (average number of items = 8.5) and reliability (group-level Rasch reliability: 0.88 - 0.93; individual-level Rasch reliability: ≥70% of patients had Rasch reliability score ≥0.90) to simultaneously assess four functions in patients with stroke. In addition, because the CAT-FAS is a computer-based test, the CAT-FAS has three additional advantages: the automatic calculation of scores, the immediate storage of data, and the easy exporting of data. These advantages of the CAT-FAS will be beneficial to data management for clinicians and researchers.

Video Link

The video component of this article can be found at https://www.jove.com/video/58137/

Introduction

Dysfunctions of the upper and lower extremities (UE and LE), postural control, and basic activities of daily living (BADL) are major sequelae of stroke[1,2,3]. The assessment of these four functions in patients with stroke is fundamental for clinicians to evaluate patients' levels of dysfunctions, set treatment goals and plans, and monitor the longitudinal trajectories of these functions.

The Fugl-Meyer Assessment (FM),[4] the Postural Assessment Scale for Stroke patients (PASS),[5] and the Barthel Index (BI)[6] have good psychometric properties to assess the UE/LE motor functions, postural control, and BADL, respectively, in patients with stroke[7,8,9]. However, the total of 72 items from these three measures impedes the feasibility of assessing all three measures within a time-limited therapeutic session. A more efficient testing method is warranted. Computerized adaptive testing (CAT) is a modern measurement method. Compared with conventional measurement methods, CAT provides a more reliable estimate of the examinee's level of function in much less time[10,11,12]. In conventional measurement methods, each examinee receives the same test form (or item sets), in which many items are too difficult or too easy for the examinee. These items provide limited information for estimating the examinee's level of function and are time-intensive for examinees. In contrast, in CAT, each examinee gets a tailored item set, in which the difficulty level of the selected items meets the function level of the examinee. Because these items are tailored for that particular examinee, CAT can provide a more reliable estimate of the examinee's level of function with fewer items and, thus, in much less time. The steps of CAT development are shown in **Supplementary File 1: Appendix 1**.

Because CAT promises reliable and efficient assessments, the CAT-FAS was developed to improve the administrative efficiency of the three measures previously used (FM, PASS, and BI)[13]. This paper describes the development and administration of the CAT-FAS. This protocol provides information for researchers to develop their CATs and for prospective users of the CAT-FAS to administer it. We also address the strengths and weaknesses of the CAT-FAS.

Protocol

This study protocol was approved by a local institutional review board, and all patients gave informed consent.

1. Development of the CAT-FAS

1. Retrieve the secondary and encrypted data from the FAS study[14] to conduct simulations (**Supplementary File 1: Appendix 2**).
   NOTE: In the study, a total of 301 patients were recruited froma rehabilitation ward of a medical center and assessed at 14 d after stroke onset. Among the 301 patients, 262 patients were repeatedly assessed at 30 d after stroke onset. The study recruited patients who had (1) a diagnosis of stroke, (2) first onset of stroke, (3) onset of stroke within 14 d before hospitalization, (4) the ability to follow commands, and (5) the ability to give informed consent personally or by proxy. Patients who had other major diseases were excluded. In each assessment session, patients were assessed with the FM, PASS, and BI by a well-trained occupational therapist (**Supplementary File 1: Appendices 3-5**).
   1. Establish the **item bank** of the CAT-FAS by adopting the item bank of the FAS (**Supplementary File 1: Appendix 2A**).
      NOTE: The item bank has sufficient items fit the Rasch partial credit model[15,16] and covers a wide range of item difficulties. The item bank contains 58 items (**Supplementary File 1: Appendix 3**) selected from the FM-UE (26 items), FM-LE (11 items), PASS (12 items), and BI (nine items).
   2. Retrieve the item difficulties of all items in the item bank from the FAS study (**Supplementary File 1: Appendix 2A** - Item difficulty).
      NOTE: Each item in the item bank has a set of parameters to depict the difficulty of the item (*i.e.*, item difficulties), which are estimated by the Rasch partial credit model. The CAT-FAS uses the item difficulties to (1) select items with difficulties tailored to the examinee's level of function (step 1.3.3) and (2) estimate the examinee's level of function (step 1.3.5).
   3. Retrieve each patient's responses (*e.g.*, 0, 1, or 2 points) to the items of the item bank of the FAS (**Supplementary File 1: Appendix 2B**).
      NOTE: In previous studies[14], all items of the item bank of the FAS were administered to the patients. In this simulation study, these responses of the patients were retrieved and used as the simulated responses (patients were not administered by the CAT-FAS) to the items of the CAT-FAS (step 1.3.4).
   4. Retrieve the ability distribution (*i.e.,* the standard deviation [SD] of the scores) of the patients in the four functions (BADL, postural control, and UE/LE motor functions; **Supplementary File 1: Appendix 2C**).
      NOTE: The abilities of the patients in the four functions are the final scores of the assessment of the item bank (**Supplementary File 1: Appendix 2C**). The scores (and SD of the scores) of the four functions are estimated in a previous study[14] by the Rasch partial credit model, based on the patients' responses to each item (step 1.1.3). In this study, the SD of the scores is retrieved and used as prior information to calculate the reliability of the CAT-FAS (step 1.3.6).

2. Determine the operational algorithms of the CAT-FAS (**Supplementary File 1: Appendix 7**).
   1. Adopt the maximum *a posteriori* (MAP) method for estimating each patient's scores of the four functions with Newton-Raphson iteration[17].
   2. Use the criterion of D-optimality for the item selection[18]. An item with the maximum determinant of the Fisher information matrix is selected from the item bank for administration.
   3. Adopt 10 candidate sets of stopping rules for exploring the properties of the CAT-FAS *via* simulation (**Supplementary File 1: Appendix 8**).
      NOTE: The first five candidate sets are "reaching limited reliability increase (LRI) criterion" (*i.e.*, an LRI < 0.001, < 0.005, < 0.010, < 0.015, or < 0.020). The other five candidate sets are "reaching either LRI criterion or threshold of reliability" (*i.e.*, a Rasch reliability ≥ 0.90, paired with the aforementioned five LRI criteria). The LRI and threshold of reliability are frequently adopted stopping rules in CATs[13,17].

3. Explore the measurement reliability and efficiency (number of items needed for administration) of the CAT-FAS *via* steps 1.3.1 to 1.3.11 of simulation (**Figure 1**).
   NOTE: **Supplementary File 1: Appendix 9** shows the screenshot of the software.
   1. Use a specified set of stopping rules (*i.e.*, from the first to the last candidate sets of stopping rules which are in step 1.2.3, successively) to explore the properties of the CAT-FAS (**Figure 1A**).
   2. Set the initial CAT-FAS scores of the four functions (BADL, postural control, UE motor function, and LE motor function) to 0 for specified patients (*i.e.*, from the first to the last patient in the data, successively; **Figure 1B**,**C**).
   3. Adaptively select an item with the maximum determinant of the Fisher information matrix (*i.e.*, the criterion of D-optimality) from the item bank for administration (**Figure 1D**).
      NOTE: The information matrix of each item is calculated based on a patient's scores of the four functions and the item's difficulty (from step 1.1.2). To ensure that the CAT-FAS administers at least one item in each function/domain, the first four items of the CAT-FAS are selected from the four functions.
   4. Obtain the patient's response to the selected item from step 1.1.3 (**Figure 1E**).
   5. Simultaneously estimate the CAT-FAS scores (and standard errors [SEs] of the scores) of the four functions using the MAP method with an iterative Newton-Raphson process (**Figure 1F**).[19] During the iterative Newton-Raphson process, renew the scores and SEs of the four functions in each iteration until the criterion of convergence is met. Convergence occurs when the differences in scores between two consecutive iterations are <0.001.
   6. Count the number of items which are administered, save the latest renewed CAT-FAS scores (and SEs), and calculate the individual-level Rasch reliability of each function using the following formula:
      $1 - ([SE^2$ of step 1.3.5] / $[SD^2$ of the scores of step 1.1.4]).
   7. Calculate the LRI using the last renewed individual-level Rasch reliability (step 1.3.6) minus that of the previous estimation (**Figure 1G**).

8. Check whether the specified set (*e.g.*, the first candidate set) of stopping rules is met (**Figure 1H**). If not, repeat steps 1.3.3 - 1.3.8 until the specified set of stopping rules is met. If so, save the latest renewed CAT-FAS scores (and SEs) as the final CAT-FAS scores (and SEs).
9. Repeat steps 1.3.2 to 1.3.8 until all patients' administrations are completed (**Figure 1I**).
10. Finish the simulation of the CAT-FAS with the specific set of stopping rules and save the results of the simulation (**Figure 1J**).
   NOTE: The results should include (1) the final CAT-FAS scores (and SEs) of the four functions, (2) the number of items needed to complete the CAT-FAS, (3) the Rasch reliability of each patient (*i.e.*, individual-level Rasch reliability), and (4) the average Rasch reliability of all patients.
11. Repeat steps 1.3.1 to 1.3.11 to explore the properties of the CAT-FAS with other candidate sets of stopping rules until all candidate sets of stopping rules are explored (**Figure 1K**).

4. Select the final set of stopping rules for the CAT-FAS according to the average Rasch reliability of ≥0.90 in at least three functions and the average items of administration of ≤10.0.
5. Develop an online administration platform for the CAT-FAS by writing a computer program to establish a website (**Supplementary File 1: Appendix 10**).

2. Administration of the CAT-FAS

1. Connect the examiner's electronic device (*e.g.*, personal computer, tablet, or smartphone) to the online administration platform of the CAT-FAS using an internet browser.
2. Log in to the administration system (**Supplementary File 1: Appendix 11**).
3. Click **Data management** to access data from previous examinees (**Supplementary File 1: Appendix 12**).
4. Click **New examinee** to create an account for a new examinee by entering the examinee's name and ID number.
5. Select an examinee and click **Start** (**Supplementary File 1: Appendix 13**).
6. Click **New assessment** to create a new assessment or click Results to review the results of the examinee's previous assessments.
7. Administer the items shown on the screen to the examinee (**Supplementary File 1: Appendix 14**).
8. Rate the examinee's performance or responses by clicking the rating scale shown at the bottom of the screen(**Supplementary File 1: Appendix 14**).
9. Explain the results of the CAT-FAS to the examinee, including the T-scores with a 95% interval, the percentile ranks of the T-scores, and the reliabilities of the four functions of the CAT-FAS. These results are calculated and shown automatically by the CAT-FAS (**Supplementary File 1: Appendix 15**).
10. Click **OK** and return to the **Data management** page.

Representative Results

The results of the simulation showed that the 10 candidate sets of stopping rules had sufficient average Rasch reliability (0.86 - 0.95) and varied administrative efficiency (the average number of items = 6.4 - 17.5). Considering the trade-off between reliability and administrative efficiency, the set of **LRI < 0.010** was selected as the optimal set of stopping rules for the CAT-FAS because of its sufficient average Rasch reliability (0.88 - 0.93, see **Table 1**), individual-level Rasch reliability (≥70% of the patients had a Rasch reliability of ≥0.90), and administrative efficiency (the average number of items = 8.5, see **Table 2**).
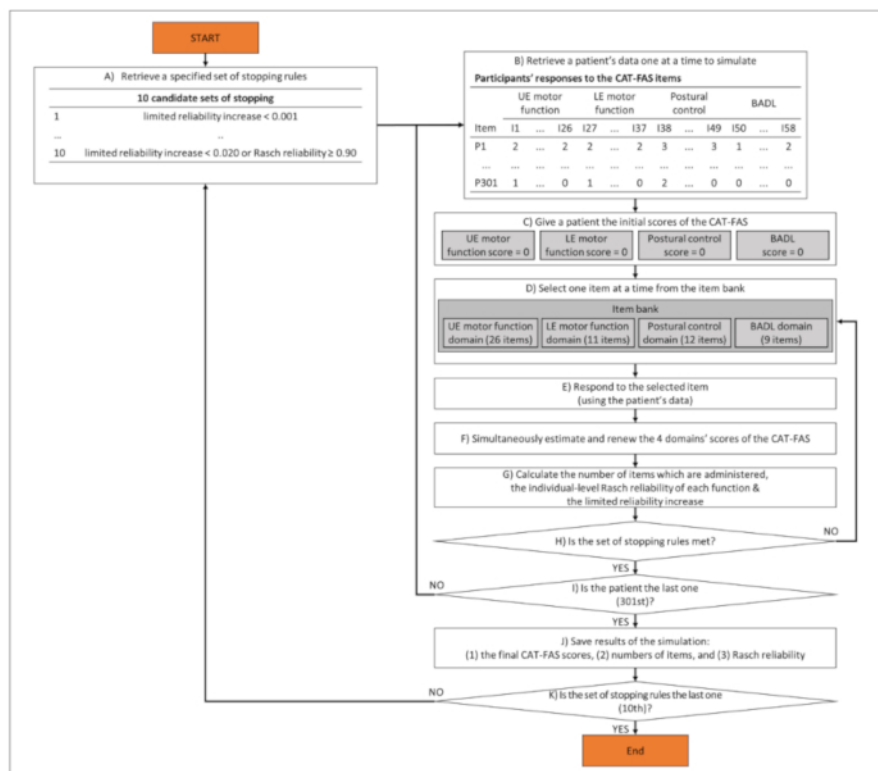
Figure 1: Process of exploring the performance of the CAT-FAS *via* simulation analysis. This figure shows the process of exploring the measurement reliability and efficiency (number of items needed for administration) of the CAT-FAS with 10 candidate sets of stopping rules. Please click here to view a larger version of this figure.

	Average	% of the patients with reliability ≥ 0.90
CAT-FAS		
UE motor function	0.88	69.8
LE motor function	0.9	76.2
Postural control	0.93	88.6
BADL	0.9	78.9
Item bank (58 items)		
UE motor function	0.9	69.4
LE motor function	0.92	77.4
Postural control	0.96	96
BADL	0.94	93.4
UE: upper extremity; LE: lower extremity; BADL: basic activities of daily living		

Table 1: Rasch reliability of the CAT-FAS. For the CAT-FAS, the average Rasch reliability of the four functions ranged from 0.88 to 0.93, and the individual-level Rasch reliability shows ≥70% of the participants with a Rasch reliability of ≥0.90.

	Average	Range	% of the patients using 5–10 items	% of the patients using > 10 items
CAT-FAS	8.5	~4-13	66.4	19.5

Table 2: Efficiency (number of items) of the CAT-FAS. The average number of items needed for administration is 8.5. Most participants (66.4%) were assessed using 5 - 10 items.

Supplementary File 1. Please click here to download this file.

Discussion

The results presented here showed that the CAT-FAS administered about 10% of the items in the original tests (the average number of items used in the CAT-FAS: 8.5 items *vs.* the original tests: 72 items). These findings indicate that the CAT-FAS has good administrative efficiency. The results were in line with previous studies, which reported that a CAT administered only about 10 items or less to assess social function, balance, or activities of daily living in patients with stroke[10,11,20]. The CAT-FAS, having good administrative efficiency, has great potential to reduce the time and burden for both patients and clinicians.

The average Rasch reliability of the CAT-FAS was 0.88 - 0.93, and more than 70% of the patients had a Rasch reliability of ≥0.90. These results reveal a good Rasch reliability of the CAT-FAS in patients with stroke. The good Rasch reliability of the CAT-FAS can be ascribed to two factors: a sound item bank and the feature of multidimensionality. First, the item bank of the CAT-FAS contains 58 items that cover a wide range of functional level for each domain[14]. The sufficient item coverage of the item bank can provide sufficient information to reliably estimate the examinee's level of function. Second, the CAT-FAS is a multidimensional CAT (*i.e.*, four domains of the CAT-FAS), in which a patient's item response on any domain can be used to simultaneously estimate the patient's abilities (scores) of all four domains by considering the correlations among all domains. This feature of a multidimensional CAT has been proven to improve the Rasch reliability in previous studies on developing multidimensional CATs[21,22]. The CAT-FAS with good Rasch reliability can be used to precisely calibrate the patients' levels of the four functions (UE/LE motor function, postural control, and BADL) with limited random measurement error.

In addition, because the CAT-FAS is a computer-based test, the CAT-FAS has three additional advantages: an automatic calculation of scores, an immediate storage of data, and the easy exporting of data. The automatic calculation of scores saves examiners' time and reduces mistakes in scoring. The immediate storage of data improves the efficiency of monitoring an examinee's longitudinal changes in the four functions. The easy exporting of data enhances the efficiency of processing electronic medical records, sharing administration results between/within clinicians and patients, and analyzing data for research. These advantages of the CAT-FAS improve the overall efficiency of data management for clinicians and researchers.

The results presented here revealed that the CAT-FAS, with different sets of stopping rules, showed different performances on administrative efficiency and reliability. In general, a trade-off relationship was found between administrative efficiency and reliability. For example, the set of **LRI < 0.001** had a higher reliability and lower administrative efficiency compared to the set of **Rasch reliability ≥ 0.90** or **LRI < 0.020**. The set of **LRI < 0.010** had both sufficient administrative efficiency and sufficient reliability, so it was selected as the final set of stopping rules for the CAT-FAS. If prospective users need the CAT-FAS to have a higher administrative efficiency or reliability, they can select another set of stopping rules for administering the CAT-FAS.

The first four items of the CAT-FAS were selected within each of the four domains. This design can prevent an unexpected situation that may occur in a multidimensional CAT. The unexpected situation is that a domain's score of a multidimensional CAT might be estimated without administering any items from that domain. The unexpected situation occurs because a multidimensional CAT can use (1) the scores of the other domains and (2) the correlations among domains to estimate the scores of the domain without any items being administered[15]. In contrast, the CAT-FAS's item selection rule of the first four items promises that at least one item from each domain is administered. Thus, the CAT-FAS can provide more representative information to estimate patients' four functions.

Three limitations of the CAT-FAS are noticed. First, the training time for administration may be long because prospective users have to become familiar with the 58 items in the item bank, as well as with the instructions and rating criteria. Second, the four domains of the CAT-FAS cannot be administered separately. Third, the results presented here were from a simulated study instead of actual administrations of the CAT-FAS in patients with stroke. Therefore, the results may be somewhat different from those of an actual administration. Field tests of the CAT-FAS are warranted in the future.

Disclosures

The authors have nothing to disclose.

Acknowledgements

References

1. Kim, S. S., Lee, H. J., You, Y. Y. Effects of ankle strengthening exercises combined with motor imagery training on the timed up and go test score and weight bearing ratio in stroke patients. *Journal of Physical Therapy Science.* **27** (7), 2303-2305 (2015).
2. Langhorne, P., Coupar, F., Pollock, A. Motor recovery after stroke: A systematic review. *Lancet Neurology.* **8** (8), 741-754 (2009).
3. Lum, P. S., Burgar, C. G., Shor, P. C., Majmundar, M., Van der Loos, M. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Archives of Physical Medicine and Rehabilitation.* **83** (7), 952-959 (2002).
4. Fugl-Meyer, A. R., Jaasko, L., Leyman, I., Olsson, S., Steglind, S. The post-stroke hemiplegic patient 1: A method for evaluation of physical performance. *Scandinavian Journal of Rehabilitation Medicine.* **7** (1), 13-31 (1975).
5. Benaim, C., Perennou, D. A., Villy, J., Rousseaux, M., Pelissier, J. Y. Validation of a standardized assessment of postural control in stroke patients: The Postural Assessment Scale for Stroke Patients (PASS). *Stroke.* **30** (9), 1862-1868 (1999).
6. Mahoney, F. I., Barthel, D. W. Functional Evaluation: The Barthel Index. *Maryland State Medical Journal.* **14**, 61-65 (1965).

7. Duffy, L., Gajree, S., Langhorne, P., Stott, D. J., Quinn, T. J. Reliability (inter-rater agreement) of the Barthel Index for assessment of stroke survivors: Systematic review and meta-analysis. *Stroke.* **44** (2), 462-468 (2013).

8. Lin, J. H., Hsueh, I. P., Sheu, C. F., Hsieh, C. L. Psychometric properties of the sensory scale of the Fugl-Meyer Assessment in stroke patients. *Clinical Rehabilitation.* **18** (4), 391-397 (2004).

9. Mao, H. F., Hsueh, I. P., Tang, P. F., Sheu, C. F., Hsieh, C. L. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke.* **33** (4), 1022-1027 (2002).

10. Hsueh, I.-P. *et al.* Development of a computerized adaptive test for assessing balance function in patients with stroke. *Physical Therapy.* **90** (9), 1336-1344 (2010).

11. Hsueh, I.-P., Chen, J.-H., Wang, C.-H., Hou, W.-H., Hsieh, C.-L. Development of a computerized adaptive test for assessing activities of daily living in outpatients with stroke. *Physical Therapy.* **93** (5), 681-693 (2013).

12. Wong, A. W., Heinemann, A. W., Miskovic, A., Semik, P., Snyder, T. M. Feasibility of computerized adaptive testing for collection of patient-reported outcomes after inpatient rehabilitation. *Archives of Physical Medicine and Rehabilitation.* **95** (5), 882-891 (2014).

13. Lin, G.-H., Huang, Y.-J., Lee, S.-C., Huang, S.-L., Hsieh, C.-L. Development of a computerized adaptive testing system of the Functional Assessment of Stroke. *Archives of Physical Medicine and Rehabilitation.* **99** (4), 676-683 (2017).

14. Wang, Y.-L., Lin, G.-H., Yi-Jing, H., Chen, M.-H., Hsieh, C.-L. Refining three measures to construct an efficient Functional Assessment of Stroke. *Stroke.* **48** (6), 1630-1635 (2017).

15. Adams, R. J., Wilson, M., Wang, W.-c. The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement.* **21** (1), 1-23 (1997).

16. Masters, G. N. A Rasch model for partial credit scoring. *Psychometrika.* **47** (2), 149-174 (1982).

17. Wang, W.-C., Chen, P.-H. Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement.* **28** (5), 295-316 (2004).

18. Mulder, J., Van der Linden, W. J. Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika.* **74** (2), 273-296 (2009).

19. Segall, D. O. General ability measurement: An application of multidimensional item response theory. *Psychometrika.* **66** (1), 79-97 (2001).

20. Lee, S.-C. *et al.* Development of a social functioning assessment using computerized adaptive testing for patients with stroke. *Archives of Physical Medicine and Rehabilitation.* **99** (2), 306-313 (2018).

21. Paap, M. C. S. *et al.* Measuring patient-reported outcomes adaptively: Multidimensionality matters! *Applied Psychological Measurement.* **42** (5), 327-342 (2018).

22. Paap, M. C. S., Kroeze, K. A., Terwee, C. B., van der Palen, J., Veldkamp, B. P. Item usage in a multidimensional computerized adaptive test (MCAT) measuring health-related quality of life. *Quality of Life Research.* **26** (11), 2909-2918 (2017).