COMPUTER ADAPTIVE MEMORY TEST MAINTAINS DECISION ACCURACY AND REDUCES TEST LENGTH

Edward S. Ericksen, ATS, Jae-Hyeon Park* Ph.D., Minsoo Kang†, Ph.D., Brian G. Ragan, Ph.D., ATC

Athletic Training Outcomes Research Laboratory, University of Northern Iowa

***Korea National Sport University, Seoul, Korea; Middle Tennessee State University, Murfreesboro, TN**

INTRODUCTION, OBJECTIVE, & DESIGN

Researchers in Athletic Training have often used patient reported outcome measures to assess health and functional status. These measures require the patient to respond to all of the items to determine ability levels, which is considered a fixed length test. Computer Adaptive Testing (CAT) is an approach that reduces test length, while maintaining test sensitivity and reliability.[1] Sequential Probability Ratio Testing (SPRT), a type of CAT, has been reported effective in criterion-referenced testing (master/non-master). In the classification of cognitive dysfunction from traumatic brain injuries, SPRT may be a quicker, more efficient device used to identify injuries.

First developed by Wald[2] in 1945, SPRT was a quality control measure in testing manufactured goods for the military during World War II. The original idea behind this method was to reduce the number of samples needed to determine if the goods were of acceptable or unacceptable quality before shipping them to the troops.[3] The accuracy and increased speed from SPRT provided a valuable manufacturing breakthrough. It was only a short time before the applications of SPRT were developed to speed up the classification of human behavior and function. SPRT has been successfully applied as a method to shorten criterion-referenced classifications of human functions.[4,5,6]

Sequential Probability Ratio Testing uses the pattern of sucesses on sequential trials to determine group membership (i.e., master, non-master) by examining the posterior probabilities (probabilities of success and failure) of the trials for both groups. Unlike more complicated CAT procedures that require Item Response Theory and a large sample size (n>200), this type of CAT (e.g., SPRT) only requires a small sample (n≤100) to calculate the posterior probabilities.[7]

Relationship to Healthcare: With an increased emphasis of establishing evidence-based practice attention and resources are being used to collect patient-reported data, which then can be used to identify or classify the patient. The use of CAT, specifically SPRT, could potentially have a tremendous contribution to healthcare by maintaining classification accuracy and decreased test length. This would allow for researchers to collect more data without increasing the burden to the patient and allow clinicians the speed to identify conditions more quickly and initiate care sooner, thus improving the quality of care athletic trainers provide.

Objectives: The objectives of this study were to directly observe and measure the accuracy and efficacy of SPRT when utilized for measuring simulated cognitive dysfunction.

Research Design: A prospective design was used.

METHODS

Sampling/Subjects: A total of one hundred participants' response vectors were randomly selected from a larger data set of previously completed study on short-term memory. The data were from a Short-term Memory Assessment Recall Tool (SMART)[8], developed at the University of Illinois at Urbana-Champaign to test short-term memory. The participants' responses were used for all three thresholds.

Methods: The participants' responses were entered one at a time into an excel macro **(Figure 1)**, which calculated the probability ratio (Probability of Master/ Probability of Non-master) from the posterior probabilities. With SPRT there are four parameters, α, β, θ_0, and θ_1 which determine the classification of master and non-master. The α parameter refers to the probability of a false negative, and β is the probability of a false positive classification.[9] The θ_0 parameter is the minimum probability of success for a master and θ_1 is the maximum probability of success for a non-master from the posterior probabilities.

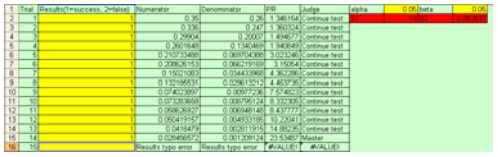


Figure 1.

The SPRT follows a simple algorithm to make a classification, which is depicted in **Figure 2**. In this figure a threshold (cut-score) is predetermined and posterior probabilities calculated. The participant's response to the first item is entered. The response is scored correct or incorrect, then a probability ratio is calculated. If the probability ratio is greater than α or less than β the test is terminated and a classification of master or non-master is made. If this condition is not satisfied then another items response is entered and the probability ratio is recalculated. This process is repeated until either the α or β criteria are met. **Figure 3** represents the probability ratio calculations for an individual over many items and until the ratio has passed the β criteria. **Validity:** There is substantial validity evidence for the methods used in this study.

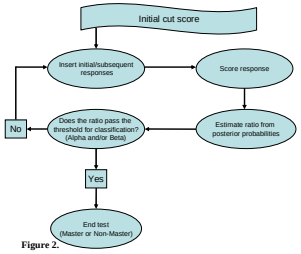


Figure 2.

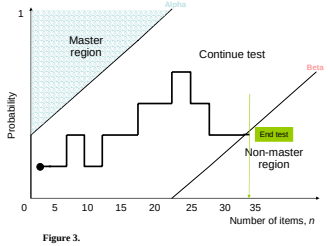


Figure 3.

DATA ANALYSIS and RESULTS

Statistical Analyses: Data were analyzed with C coefficients, which represents the accuracy of the decisions of master and non-master for each of the three cut-scores along with an overall average. The total number of items needed to make a decision were also recorded for each decision.

Independent Variables: The three cut-scores (59%, 70%, and 80%) were used to identify master (non-injured) and non-master (injured). Three cut-scores were selected that represented a low, middle and high scores. Multiple cut-scores were used to reflect the current recommended practice in concussion assessment of using an individual centered method of detection where an individual post-injured score is compared to their pre-injury (baseline) score. This practice require SPRT to be able to perform accurately over a wide range of possible cut-scores.

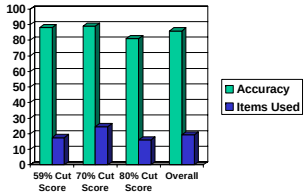
Dependent Variables: The accuracy of the classification of participants' response vectors as master or non-master; the number of items needed for classification.

Table 1. Statistical Results

Results:

Cut-score	# Masters	# Non-masters	# Misclassifications	Accuracy	Number of Items Used (% of items)
59%	81	19	12	88%	33 (17.5%)
70%	57	43	11	89%	46 (24.5%)
80%	29	71	19	81%	30 (16.0%)
Overall				86%	36 (19.2%)

N=100



Note. This table depicts the breakdown of the sample into master and non-master classifications based on each cut-score. The number of misclassifications and the accuracy (C-coefficient) for each cut-score is also presented. In addition, the average number of items needed for classification and the percentage of total items used are reported.

INTERPRETATION of CONCLUSIONS

Interpretation: Athletic trainers make decisions and classify athletes multiple times a day (e.g., play, no-play). The use of CAT holds great potential in assisting the athletic trainer in making those decisions. The results indicate that the application of a criterion-referenced standard (pass fail) with the CAT using the SPRT could decrease test time, maintain high level of accuracy, and reduce item exposure, which may limit any potential practice effects.

Related Literature: The application of SPRT has been successfully applied to the physical performance field with success. Safrit et al.[10] applied SPRT to classify master/non-masters in a motor task where the test accurately classified 71% of students using the instructor as the criterion. The number of trials was reduced from 62 to 31 for males and 23 for females. In the present study the results were even better with classification accuracy of 86% and a reduction in test length of 80.8%.

The uniqueness of the present study was it used multiple cut-scores to replicate the individual-centered standard used in the identification of concussions. This means that the posterior probabilities have to calculated based on the cut-score determined from the baseline test for that individual and these probabilities would be different across the range of scores. That is why three separate cut-scores (59, 70, 80) were selected for this study that represented a wide range in test scores. Previous applications have only dealt with a stationary cut-score.[11] The results are very encouraging for future classification applications in concussion testing.

The accuracy of classification using SPRT in identifying memory performance decreases is comparable to other valid tests, which are longer and require more time and possible suffer from practice effects. In the present study participants' were correctly classified 86% of the time. This rate is similar to the data reported by Schatz et al.[12] where a accuracy (sensitivity) rate of almost 82%.

Conclusions/Impact: Our results have produced evidence to show that SPRT is an effective clinical tool when used for Mastery/Non-mastery classifications. The results show that SPRT on average, decreased the test length by approximately 80%. Application of the SPRT would be easy to incorporate into practice because with computers. Palm pilots, laptops, and personal computers would be able to handle SPRT applied to a clinical test.

REFERENCES

1. Ware JE, Bjorner JB, Kosinski M. Practical implications of item response theory and computerized adaptive testing. *Medical Care*. 2000; 38(9); 73-82.
2. Wald A. Sequential method of sampling for deciding between two courses of action. *Journal of American Statistical Association*. 1945; 40; 277-306.
3. Wald A. Sequential analysis. 1947. New York: Wiley.
4. T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, New York: Academic Press, 1967.
5. Reckase, M. D. (1983). A procedure for decision making using tailored testing. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 237-255). New York: Academic Press.
6. Eggen, T. J. H. M. (1999). Item selection in adaptive testing with the sequential probability ratio test. Applied Psychological Measurement, 23, 249-261.
7. Safrit MJ, Stamm CL, Douglass JA. The consistency of mastery classifications for a criterion-referenced test of motor behaviour: The effect of varying sample size. *Journal of Human Movement Studies*. 1981; 7; 131-143.
8. Ragan, B. G., & Gandhi, M. (2004). Short-term Memory Assessment Recall Tool (SMART) (Version 1.0) [Computer software]. Urbana, IL.
9. Welch ER, Frick TW. Computerized Adaptive Testing in Instructional Settings. *Educational Technology Research and Development*. 1993; 41(3); 47-62.
10. Safrit MJ, Wood TM, Ehlert SA, Hooper LM, Patterson P. The application of sequential probability ratio testing to a test of motor skill. *Research Quarterly For Exercise And Sport*. 1985; 56(1); 58-65.
11. Frick TW. Computerized adaptive mastery tests as expert systems. *Journal of Educational Computing Research*. 1992; 8(2); 187-213.
12. Schatz P, Pardini JE, Lovell MR, Collins MW, Podell K. Sensitivity and specificity of the impact test battery for concussion in athletes. *Archive of Clinical Neuropsychology*. 2006; 21(1); 91-99.