The University of Iowa
Iowa Testing Programs

The preassembled, parallel computerized fixed-test (CFT) forms are among the most popular computer-based testing models. In item response theory, the test information function plays the dominant role for designing and comparing the measurement precision of the CFT forms. In addition to Base SAS®, the current paper develops an automated procedure by utilizing several SAS® software and procedures (i.e. PROC IML, PROC SQL, SAS/GRAPH®, GTL, & ODS) for the construction of the CFT forms. The purpose is to demonstrate an automated and efficient way to obtain test and item information functions for the CFT forms and also, to plot the corresponding test and item characteristic curves along with informative summary statistics. Meanwhile, the paper investigates how measurement precision is interwoven with conventional item statistics. For test developers and practitioners, the handy automated procedure through SAS® and informative results are both provided for practical uses and for future references.

COMPUTERIZED FIXED-TEST FORMS

Incremental growth in computer-based testing (CBT) initiates the development of the automated procedure. The preassembled parallel, computerized fixed tests (CFT) model and its variants provide all of the general advantages of CBT (Leucht & Sireci, 2011). The automated CFT form assembly procedures usually need only deal with a single target test information function which dominates the development of the preconstructed, intact test forms that are administered by computer.

IRT MODEL, INFORMATION FUNCTION, & MEASUREMENT PRECISION

- A 3PL unidimensional IRT model (Birnbaum, 1968; Lord, 1980) is assumed
- Item characteristic function/curve (ICF/ICC) shows the probability that a person with a given ability level will answer the item correctly
  Item information function (IIF) tells individual contribution of an item
- Test characteristic function/curve (TCF/TCC) shows the relationship between total score and person ability estimate
  Test information function (TIF) is simply the sum of all IIFs on the test
- *The higher the test/item information, the smaller the measurement error <=> the more measurement precision achieved*

IMPLEMENT the AUTOMATED PROCEDURE

- Constructing four 10-item tests from a math item pool to show "how item difficulty (i.e., *p*-value) and item discrimination (i.e., biserial correlation) relate to the measurement precision"
- Constructing four tests of different lengths to show the straightforward relationship between test length and test information

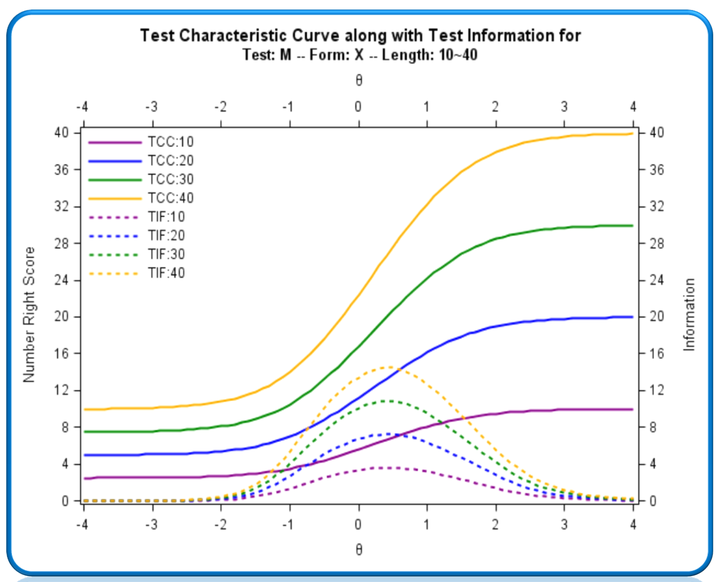

Figure 1 .
TCCs and test information of 10-, 20-, 30-, and 40-item tests

Automated Procedure in SAS®

Base SAS®
--Building item pool from test blue print & results from IRT parameter estimation program (e.g., BILOG-MG)

PROC IML
--Fast computing ICF, IIF, TCF, & TIF
--Fast look up for max info. & ability estimates

PROC SQL
--Creating macro variables for displaying info. in graphs

SAS/GRAPH® plus GTL
--Creating TCC & ICC graphs

Repetitive until the desired maximum test information obtained

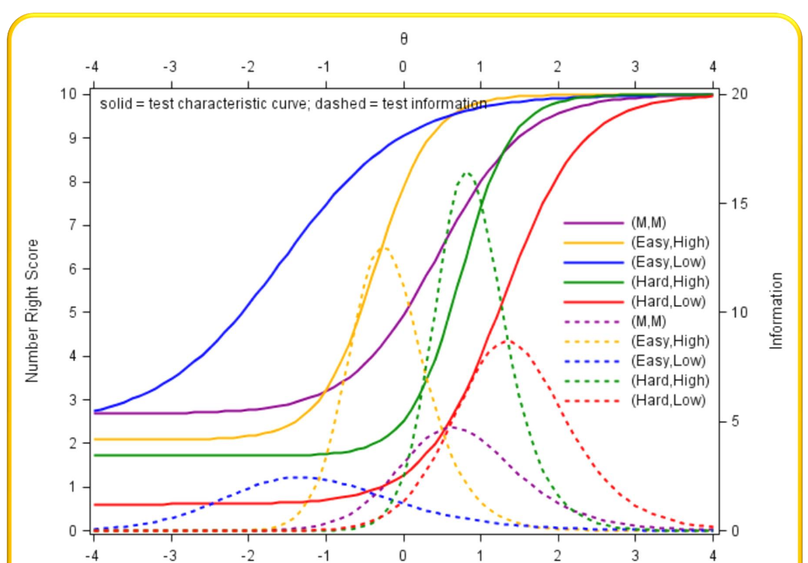

Figure 2 .
TCCs and test information of five 10-item tests of varying item difficulty and item discrimination

Figure 1—Given the definition of test information function, longer tests have more information than shorter tests.

Figure 2—Test of moderate-difficulty-and-discrimination items (M,M) has the max test information near $\theta = .5$. Tests of more high discrimination items have higher test information (orange & green). TCCs of tests of harder items shift to the right (green & red). Alternatively, one can use relative efficiency to compare tests (Lord, 1980).

WHAT'S in the ITEM POOL?

Item-level information, including test, form, and content domain to which items belong, unique item ID, conventional item statistics such as difficulty (*p*-value) and discrimination (biserial), and also, IRT parameter estimates (*a*, *b*, & *c*), etc.

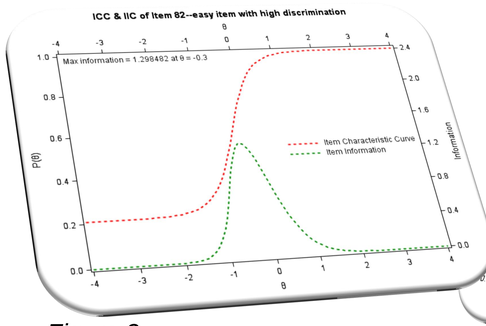

Figure 3.
Easy item with high discrimination

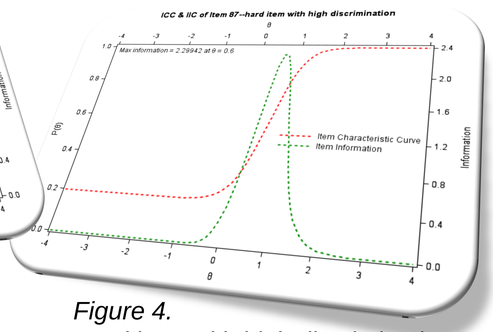

Figure 4.
Hard item with high discrimination

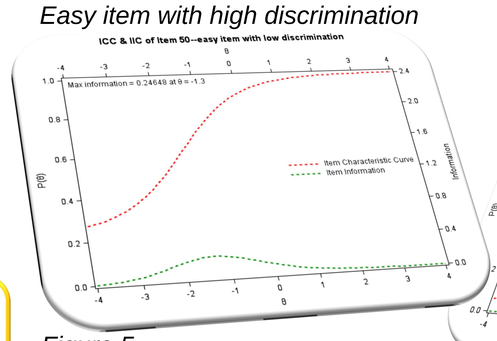

Figure 5.
Easy item with low discrimination

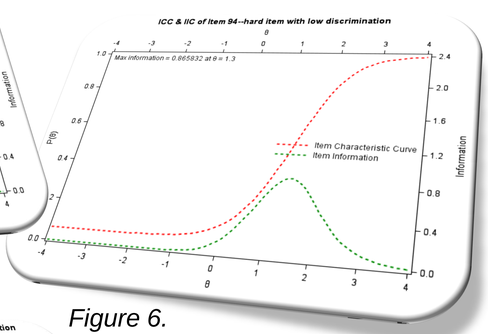

Figure 6.
Hard item with low discrimination

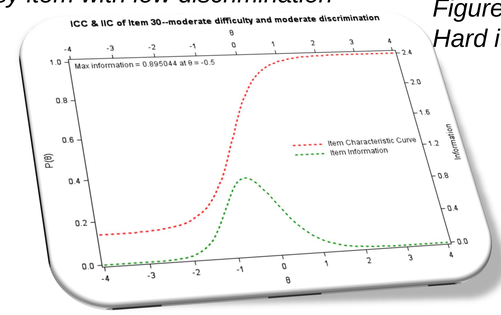

Figure 7.
Item with moderate difficulty and discrimination

Figures 3-7 show ICC s and item information for items of different difficulties and discriminations.

- Max information occurs at the inflection point of ICC/TCC where the ability can be measured at the greatest precision.
- Easy, low discrimination item has little information.

CONCLUSION

- To achieve maximum measurement precision for a certain ability, choose the item(s) that can offer the most information at the particular ability. By inspecting the item-level information and graphs, one can decide which item(s) is required to maximize desired test information and measurement precision.
- Motivated by the need to construct computerized fixed-test forms and activated by the power and flexibility of SAS, the automated procedure developed can meet the need!

References

- Birnbaum, A. (1968). Estimation of an ability. In F. M Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 423–479). Reading, MA: Addison-Wesley.
- Cartier, J. (2006). A programmer's introduction to the Graphics Template Language. Proceedings of *the Thirty-first Annual SAS® Users Group International Conference, Paper 262*. Cary, NC: SAS Institute Inc
- Heath, D. (2009). Secrets of the SG procedures. Proceedings *of the SAS® Global Forum 2009 Conference, Paper 324*. Cary, NC: SAS Institute
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum Associates.
- Luecht, R. M. & Sireci, S. G. (2011). *A review of models for computer-based testing* (Research Report NO. 2011-12). Retrieved from College Board website: http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-12-review-models-for-computer-based-testing.pdf
- Matange, S. (2008). Introduction to the Graph Template Language. Proceedings of *the SAS® Global Forum 2008 Conference, Paper 313*. Cary, NC: SAS Institute Inc.
- Murphy, W. C. (2007). Changing data set variables into macro variables. Proceedings of *the SAS® Global Forum 2007 Conference, Paper 50*. Cary, NC: SAS Institute Inc.
- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- SAS Institute Inc. (2009). *Base SAS® 9.2 Procedures Guide*. Cary, NC: SAS Institute Inc.
- SAS Institute Inc. (2008). *SAS/IML user's guide, Version 9.1*. Available at http://support.sas.com/91doc/docMainpage.jsp
- Terry, R, Lee, S. H., & Milburn, N. (2005). *IRT-FIT: Fitting IRT models in SAS®* . Technical Manual for Users.
- Wicklin, R. (2010a). *Statistical programming with SAS/IML software*. Cary, NC: SAS Institute Inc.
- Wicklin, R. (2010b). *LOC: The most useful function you've never heard of*. Retrieved from http://blogs.sas.com/content/iml/2010/09/10/the-most-useful-function-youve-never-heard-of/
- Wicklin, R. (2011). *Random access: How to read specific observation in SAS/IML software*. Retrieved from http://blogs.sas.com/content/iml/2011/05/18/random-access-how-to-read-specific-observation-in-sasiml-software/
- Zimowski, M., Muraki, E., Mislevy, R., & Bock, D. (2003). *BILOG-MG (Version 3.0)* [computer software]. Lincolnwood, IL: Scientific Software International.

Acknowledgements

The author would like to thank Dr. Wicklin Rick's contribution to the field of SAS/IML. To accomplish the paper, the author has been benefitted enormously from his book, his online paper, and his blog (http://blogs.sas.com/content/iml/author/rickwicklin/).

Contact the Author

Your comments and questions are valued and encouraged. Please contact the author at:

Yi-Fang Wu
Educational Measurement and Statistics, University of Iowa
E-mail: wuyifang91@gmail.com or yi-fang-wu@uiowa.edu

Sample code for the published examples is available from the author upon request.

SAS GLOBAL FORUM

THE UNIVERSITY OF IOWA

ITP Iowa Testing Programs