COMPUTERIZED COGNITIVE ADAPTIVE TESTING

计算机化的认知适应性测试

何莲珍 著

Lianzhen He

浙江大學出版社

Zhejiang University Press

Computerized Cognitive Adaptive Testing

计算机化的认知适应性测试

何莲珍　著

浙江大學出版社

图书在版编目（CIP）数据

计算机化的认知适应性测试＝Computerized Cognitive Adaptive Testing／何莲珍著．—杭州：浙江大学出版社，2004.10

ISBN 7-308-04060-7

I. 计… II. 何… III. 计算机—应用考试—适应性—研究—英文 IV. ①H08②TP391.4

中国版本图书馆 CIP 数据核字(2004)第 131875 号

责任编辑	张颖琪
封面设计	刘依群
出版发行	浙江大学出版社
	(杭州浙大路 38 号　邮政编码 310027)
	(E-mail: zupress@mail.hz.zj.cn)
	(网址: http://www.zjupress.com)
排　　版	浙江大学出版社电脑排版中心
印　　刷	浙江大学印刷厂
开　　本	787mm×960mm　1/16
印　　张	13.5
字　　数	272 千
版 印 次	2004 年 10 月第 1 版　2004 年 10 月第 1 次印刷
书　　号	ISBN 7-308-04060-7/H・324
定　　价	25.00 元

序　言

中国是考试的故乡，有其悠久的历史和丰富的经验。但是作为一种测量工具，考试必须经受现代科学技术的磨勘，与时俱进。现代教育测量学正是在现代教育理论、现代统计理论和现代教育技术的共同作用下，不断地展示着新貌。

何莲珍博士的博士论文《计算机化的认知适应性测试》(Computerized Cognitive Adaptive Testing)结合我国实际，提出了一种在计算机上实现的英语适应性测试。论文完成于1998年，现在看来，它仍然具有很大的现实意义。它的出版问世，当会对我国的测试发展提出一种新的思路。

何博士所提出的模型有几个特点:

1. 它是适应性的，也就是说，它针对每个考生的原有水平和现场的答题情况，不断地从题库里抽出适应学生能力的题目。只要题库足够大，每个考生所做的题目都不是一样的，这不但对测量学生的能力提出一种更为可靠的方法，而且可以防止作弊。适应性考试看来已是新一代考试的必经之路，美国的GRE和TOEFL就是走这样的路子。

2. 既然是适应性，它必然是一种计算机辅助考试。从题库建设、题目挑选和能力的判断都必须由计算机实现。

3. 从测试理论上，它必须以项目反应理论(Item Response Theory)为基础。项目反应理论有好几种模型，何博士所采用的是目前较多考试所使用的二参数模型。它的技术含量比较高。

4. 作为一种尝试，论文还提出考试的认知模型，其出发点是语言能力往往体现为反应时间。对英语的考生来说，答题的准确性以外，反应的流畅应该也是一个测量的参数。这个问题比较复杂，因为增加一个参数，与项目反应理论的单维度原则有悖，而且认知能力涵盖面很广，不单是一个反应能力的问题。但是何博士到底是作了有益的探索。

论文的另一个特点是，它不但介绍了所设计的模型，编制了一套可以实施的软件，而且从不同的角度(它和传统测试的比较、使用了不同的题目组的考生的能力比较、同一考生在能力没有发生变化时使用不同题目组的分数比较)去检验模型的信度和效度，说明这确实是一种值得推广的测试方法。当然这种测试要正式实施，还有许多问题要进一步解决，例如软件(题库)和硬件(计算机考场)的建设就不能一蹴而就。但这正好说明，何博士的探索确实是起到了筚路蓝缕的作用。

桂诗春

2004 年国庆节

“The foreign languages must make computer adaptive testing a priority in the 1990’s.”

——Stansfield 1990

Acknowledgements

Acknowledgements should be made first and foremost to my supervisor, Professor Gui Shichun, who has provoked the ideas for my dissertation and who, throughout the process of my writing the dissertation, has always been very helpful in giving me sincere advice, unfailing support and timely encouragement. Never will I forget the talks I had with him along the narrow roads on the campus, in the small computer lab and in his cozy house. Never will I forget the occasional nudge from him which has proved so valuable and so helpful. And never will I forget the move exchanges between us over the phone or via email. Without his help and guidance my dissertation would never have been finished. All my life I will benefit from his earnest and tireless instruction and, above all, his cultivation over the past years in the field of linguistics and applied linguistics.

Sincere thanks also go to Professor Ning Chunyan, Professor He Ziran, Professor Wang Chuming, Professor Wu Xudong and Professor Xu Luoman for their instruction, inspiration and encouragement. They have always been a source of strength and their dedication to their research in various fields has had enduring influence on me.

I would also like to express my heart-felt thanks to Professor Shao Yongzhen and Professor Ying Huilan at Zhejiang University who have been extremely helpful in various ways over the past years. Without their continuous help and encouragement, I would never have had the courage to enter into the PhD program and the process of my dissertation writing would have been much longer and much more painful.

At this moment, I feel deeply indebted to members of my family, both the extended and the nuclear family, for sharing my responsibility for the family and taking up most of the housework over the past three years. Particular mention should be made of my husband who has been so supportive and patient during the whole process. As a mother, I have always been under the guilty conscience of not being able to render my son the maternal love and care other children are enjoying profoundly. He, instead, offers me help in his unique way, by taking care of himself.

I also feel grateful to my colleagues and students at Zhejiang University and the BFT training and testing centers in Shanghai and elsewhere and many other people who have helped me, one way or another, with the CCAT system and with the formidable amount of pretesting and experiments. I particularly want to thank Mr. Shao Xuhui, Miss Zhou Qing, Mr. Zeng Yongqiang, Ms. Xiong Haihong, Mr. Zhang Quan, Mr. Liu Hanbing and Mr. Tang Hong for their great help and support.

Thanks to all who have helped me.

Abstract

The present research is an attempt to improve the precision and efficiency of educational measurement with reference to the various testing models developed over the past years and with a view to developing new item types and tapping the great potential of computer technology in language testing.

The dissertation begins by examining the educational measurement literature in four areas: (a) the use of computer technology in language testing, (b) computerized adaptive testing, (c) item response theory as a mathematical model for computerized adaptive testing, and (d) cognitive psychology and its implications for language testing.

The dissertation then reports on a pilot study in computerized cognitive adaptive testing (CCAT). CCAT has five components, none of which is dispensable. The five components are: (a) an item pool with calibrated items, (b) the two-parameter item response model, (c) the adaptive procedure for both vocabulary and structure items and reading comprehension items, (d) cognitive testing with the speed parameter incorporated in the final ability estimate, and (e) computer-generated reports for both the examinee and the examiner.

Four hypotheses are formulated and tested. The four hypotheses are: (1) CCAT is as reliable and valid as the conventional test in the measurement of the examinee's language ability and that in computerized cognitive adaptive testing fewer items need to be administered for equal or greater precision; (2) Comparisons between examinees can be made even though they take different sets of items; (3) If the examinee does not change in his ability, the same result can be obtained even though he takes the adaptive test at different times and different sets of items are administered; (4) The result of CCAT is more close to the examinee's true ability level than either CT or CAT.

Experiments on two item types, i.e., bank cloze format and short paragraph reading are done with promising results and a proper weight is found to be assigned to the speed parameter in the final ability estimate. Experiment on CCAT yields fairly satisfactory results.

The findings of the research prove the four hypotheses to be true and suggest that computerized cognitive adaptive test has advantages over the conventional paper-and-pencil test in various ways and that the inclusion of speed parameter in the final ability estimate improves the precision of educational measurement.

KEY WORDS: *item response theory* *computerized adaptive testing*
cognitive psychology *computerized cognitive adaptive testing*

摘　　要

语言测试的精度及效率问题一直是语言测试专家关注的问题。随着该领域研究的深入，认知心理学的发展与计算机技术尤其是多媒体技术的日趋成熟，对计算机适应性考试的研究也有了长足的进步。认知信息处理模型试图深入测试行为的过程，而不是仅仅看测试的结果。本研究旨在联系我国实际建立一个认知计算机适应性测试模型，以期提高目前我国一些大规模考试的精度及效率，同时对易于在计算机适应性考试中采用的题型进行了研究。

本研究从理论上对计算机在语言测试中的应用，计算机适应性考试的数学模型、起点问题、项目选择、能力估计、终止原则以及认知科学及其在语言测试中的应用问题进行了深入探讨。在理论研究的基础上结合我国的实际设计出一个阅读理解和词汇结构的计算机化的认知适应性测试模型(CCAT)。该模型由五个部分组成：1) 包括集库式完形填空、篇章阅读、短语境阅读、词汇结构四种题型的带项目参数的题库; 2) 认知计算机测试的数学模型，即项目反应理论的双参数模式，把项目难度及项目区分度作为能力估计及项目选择的依据，比单参数模式更为精确；3) 双阶计算机适应性测试模型，以根据考生集库式完形填空完成情况所作出的能力估计作为篇章、短语境阅读和词汇结构测试的起点；4) 引进速度参数的认知测试，将实验得出的最合适的速度权重包括在最后的能力估计中，提高测试的精度；5) 计算机自动生成的测试报告，报告中包含的信息对项目分析、题库维护及题库发展都具有重要的意义。CCAT 的这五个组成部分缺一不可。

本研究对该模型的四个假设进行了验证：1) CCAT 在测量考生的语言能力方面与常规测试一样有效，而在 CCAT 中题量大大减少，从而提高测试效率；2) 考了不同题目的考生之间仍然存在可比性；3) 只要考生能力没有发生变化，考生在不同时间考不同的题目，其能力保持不变；4) CCAT 比常规测试及其一般的计算机适应性测试更接近考生的实际语言能力。CCAT 的实验结果证明了以上四个假设都成立，而且 CCAT 在测量考生的语言能力方面比常规测试有更高的信度与效度。

对两种新题型，即集库式完形填空和短语境阅读的研究证明前者是测试考生综合语言能力的理想题型，后者在测试考生的阅读能力方面与篇章阅读有着较高的相关，但单句阅读更易于在计算机适应性测试中采用。对阅读测试中速度研究

的结果表明，阅读速度的差异反映出考生语言能力的差异，在阅读测试中引进速度参数可以提高语言测试的精度。

在对整个 CCAT 系统的设计中充分运用了计算机多媒体技术及软件开发工具，努力改善用户界面及测试环境，以减轻考生测试时的焦虑度，对题库中题目曝光频率的控制提高了语言测试的安全性。

关键词：项目反应理论　　计算机适应性测试
认知心理学　　计算机化的认知适应性测试

List of Abbreviations

BC	bank cloze
BFT	Business Foreign-languages Test
BIF	base information function
CAE	computer-aided education
CALT	computer-adaptive language test
CAT	Cambridge Advanced Test
CAT	computerized adaptive test(ing)
CBELT	computer-based English language testing
CCAT	computerized cognitive adaptive testing
CTT	classical test theory
CET	College English Test
CIP	cognitive information-processing
CM	continuous measurement
CPE	Cambridge Proficiency Certificate in English
CRT	criterion-referenced test
CT	computerized testing
EPM	educational psychometric measurement
EPT	English Proficiency Test
ETS	Educational Testing Service
GITEST	Guangzhou Institute of Foreign Languages Testing Packages
GUFS	Guangdong University of Foreign Studies
ICC(s)	item characteristic curve(s)
IELTS	International English Language Testing System
IIF	item information function
IM	intelligent measurement
IRT	item response theory
MCC	multiple-choice cloze
MET	Matriculation English Test
ML	maximum likelihood
MLAT	Michigan Language Aptitude Test
MMAP	marginal maximum a posteriori
MML	marginal maximum likelihood
PCF	person characteristic function
RST	reading speed test
PETS	Public English Test System
SAT	Scholastic Aptitude Test
SPR(T)	short paragraph reading (test)
TIF	test information function
TRT	traditional reading test
UCLES	University of Cambridge Local Examinations Syndicate

Contents