

2015 · 第6辑

日语教育与 日本学

■ 日语教育研究 ■ 日本语学研究 ■ 日本文学研究
■ 日本文化研究 ■ 中日语言对比研究

执行主编◎徐 曙



华东理工大学出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS

2015 · 第6辑

日语教育与 日本学



华东理工大学出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS

· 上海 ·

图书在版编目(CIP)数据

日语教育与日本学(第6辑)/徐曙执行主编. —上海:
华东理工大学出版社, 2015.6
ISBN 978-7-5628-4350-4

I. ①日… II. ①徐… III. ①日语-语言教学-文集
②日本-研究-文集 IV. ①H369-53②K313.07-53

中国版本图书馆 CIP 数据核字(2015)第 180154 号

日语教育与日本学(第6辑)

执行主编 / 徐 曙

责任编辑 / 嵇 蕾

责任校对 / 金慧娟

封面设计 / 戚亮轩

出版发行 / 华东理工大学出版社有限公司

地 址: 上海市梅陇路 130 号, 200237

电 话: (021)64250306(营销部)

(021)64252875(编辑室)

传 真: (021)64252707

网 址: press.ecust.edu.cn

印 刷 / 常熟市华顺印刷有限公司

开 本 / 787mm×1092mm 1/16

印 张 / 13

字 数 / 312 千字

版 次 / 2015 年 6 月第 1 版

印 次 / 2015 年 6 月第 1 次

书 号 / ISBN 978-7-5628-4350-4

定 价 / 58.00 元

联系我们: 电子邮箱 press_wy@ecust.edu.cn

官方微博 e.weibo.com/ecustpress

天猫旗舰店 <http://hdlgdxcb.tmall.com>



日语教育与日本学

2015 · 第6辑

中国日语教学研究会 上海分会
日语教育分会 主办
民族院校分会

编委会主任: 徐一平

编委会副主任: 刘晓芳 曹大峰 丁毅

本辑执行主编: 徐曙

编委会顾问(按姓氏汉语拼音为序)

陈俊森(华中科技大学)	冈崎眸(茶之水女子大学)
刘金才(北京大学)	皮细庚(上海外国语大学)
宿久高(吉林大学)	谭晶华(上海外国语大学)
文秋芳(北京外国语大学)	吴寄南(上海国际问题研究所)
修刚(天津外国语大学)	张辉(华东理工大学出版社)

编委会成员(按姓氏汉语拼音为序)

庵功雄(一桥大学)	蔡凤林(中央民族大学)
曹大峰(北京日本研究中心)	杜勤(上海理工大学)
高宁(华东师范大学)	侯仁锋(县立广岛大学)
胡令远(复旦大学)	冷丽敏(北京师范大学)
李晓博(深圳大学)	林洪(北京师范大学)
刘晓芳(同济大学)	刘雨珍(南开大学)
毛文伟(上海外国语大学)	潘钧(北京大学)
潘世圣(华东师范大学)	钱晓波(东华大学)
邱根成(上海对外经贸大学)	杉村泰(名古屋大学)
盛文忠(上海外国语大学)	田野村忠温(大阪大学)
王宝平(浙江工商大学)	王铁桥(河南大学)
王婉莹(清华大学)	王勇(浙江工商大学)
望月圭子(东京外国语大学)	吴川(日本大学)
毋育新(西安外国语大学)	徐静波(复旦大学)
徐敏民(华东师范大学)	徐曙(上海对外经贸大学)
徐一平(北京日本研究中心)	许慈惠(上海外国语大学)
尹松(华东师范大学)	张文丽(西安交通大学)
赵刚(西安交通大学)	赵华敏(北京大学)
周异夫(吉林大学)	朱桂荣(北京日本研究中心)

卷首语

本刊《日语教育与日本学》自 2011 年 5 月创刊至今已出 5 辑,去年更是初创 1 年出版 2 辑的佳绩。5 辑文集,彰显了日语学界资深学者厚积薄发的结晶,展示了年青新锐思想火花的碰撞,创刊伊始就被“中国知网”收入,为全国日语学界添了一份砖瓦。

为了进一步突出本刊的学术性和跨专业跨地区研究的特色,经共同协商决定自今年起由中国日语教学研究会上海分会、日语教育分会和民族院校分会共同主办《日语教育与日本学》研究集刊,每年出版两期专辑。三分会将紧密合作,加大力度,联袂全国日语教育及日本学研究方面的专家学者进一步办好集刊,并积极争取外语教育学和外国学领域同行专家的指导和支持,不断展宽学术视域,扩大集刊的影响力,提升集刊的学术地位。这一决定和办刊方针也得到了华东理工大学出版社的鼎力支持。

为了保证《日语教育与日本学》研究集刊的良好发展,主办方对编委会进行了充实调整,聘请了中日两国 10 位资深学者任编委会顾问为文集把关,还聘请了 38 位日语教育学、日本语学、日本文学及翻译、日本文化方面的专家学者为编委会成员,负责集刊的组稿和审稿事宜,为继续办好《日语教育与日本学》集刊夯实了组织基础。

《日语教育与日本学》集刊从本辑起,将启动匿名审稿程序,提高刊用文章的质量。期待学界同仁的鼎力扶持和倾心相助。

《日语教育与日本学》编委会

2015 年 5 月

版权声明:作者投稿或接受集刊编委会约稿,即视为遵守集刊主办方与本社关于集刊出版合同之约定,同意授予本社该作品的专有许可使用权,包括但不限于该作品的专有出版权、复制权、发行权以及电子与网络传播权。

目 录

• 日语教育研究 •

- 日语计算机自适应性考试综述与考察 侯仁锋(1)
- 日本人学生と留学生の文化の捉え方の相違
- コミュニケーション促進の観点から— 田崎敦子(14)
- 中日同声传译过程中的“问题词”——基于训练实例的分析 徐旻(23)

• 日本语学研究 •

- 「V1—慣れる」における母語転移の可能性 杉村泰(35)
- 近世的“素读”文化与汉文训读 潘钧(46)
- 多義動詞「カケル」の学習についての一考察
- 話し言葉における日本語母語話者の使用実態を踏まえて— 鷺見幸美(54)
- 日语使役句在会话中的意义和功能 凌蓉(64)
- 从认知语言学角度考察「ていく」和「てくる」的意义构造 张晓(73)
- 汉日同形词的概念、概念化和操作化 王志军(84)
- 论日语汉字词拟声拟态词的语义变化特征 赵耀(96)
- 关于外来语变迁的考察——以《朝日新闻》社论为研究对象 鲁峥(105)
- 《生成词库语义论》评介 邓超群(117)

• 日本文学研究 •

- “中日文学关系研究”的本体论意识——关于我的“厨川白村研究”的思考 李强(125)
- 论日本女性主义文学文本中女性形象的嬗变 李先瑞(131)
- 青年荷风汉文素养小考——以《上海纪行》为中心 钱晓波(139)

• 日本文化研究 •

- 疾病の文化史的研究の可能性 福田真人(147)
- 日本人非语言表达的文化学释义 刘军(156)
- 从多视角透视日本的自杀现象 金灵(163)
- 粤剧《胡不归》与新剧《不如归》 杨文瑜(173)
- 流行语中现代日本女性形象的动态考察——基于女性主义语境下 张丽梅(183)

• 中日语言对比研究 •

- 试析汉日方位词“里”与「の中」的异同 黄爱民(192)

日语计算机自适应性考试综述与考察^①

县立广岛大学 侯仁锋

[摘要] 计算机自适应性考试(CAT:Computerized Adaptive Test)是依靠项目反应理论(IRT:Item Response Theory)和计算机技术的支撑而发展起来的新测试形式。这一新测试形式于网络时代愈发显现出其优点,因而被应用于越来越多的考试中。本文介绍了日语计算机自适应性考试的测试理论依据、测试系统的基本原理及其考试项目构成,探讨了其优点和开发的困难所在,认为该考试基于因特网施测,使测试更加有效、更加高效,也更加客观和准确,代表了测试发展的方向。

[关键词] 自适应性 项目反应理论 测试形式

1 引言

计算机自适应性考试(CAT:Computerized Adaptive Test)是依靠项目反应理论(IRT:Item Response Theory)和计算机技术的支撑而发展起来的新测试形式。这一新测试形式于网络时代愈发显现出其优点,因而被应用于越来越多的考试中。

自适应性考试的基本原则,可以追溯到最早的自适应性测试——罗德测验^②,在罗德测验中,呈现给被试的下一个测验项目(试题、题目,下同)是根据他对前一个项目的反应决定的,即对各种不同能力水平的被试给予难度适宜的测试项目。随着计算机技术的发展,后来出现了计算机辅助测试,但这种机辅测试仍以经典测量理论(CTT:Classical Testing Theory)为依据,多为纸笔测试形式的电子版,除了测试媒介的改变外,并无实质性进步。直到70年代末80年代初,项目反应理论与计算机技术相结合,出现了真正意义上的计算机自适应性测试,现在又基于因特网施测,使测试更加有效、更加高效,也更加客观和准确。

① 此文得到“日语计算机自适应性考试”开发者日本筑波大学今井新悟教授、山口大学赤木弥生副教授的支持和认可,在此深表谢意。

② 在20世纪60—80年代,罗德在美国教育测试服务中心(ETS)进行了大量研究,在罗德的实验中,有一种名为“灵活测验”的考试。这种考试包含了难度从高到低的一系列试题,被试一开始都回答难度中等的题目,如果答对则会被要求回答更高难度的试题,答错则会接受较低难度的题目。这个实验通过灵活可变的分支来选择试题。一般认为由此确立了自适应性测试的基本原理。

2 日语考试的种类与形式

我们这里探讨的种类,是指大规模标准化考试的种类。这种大规模标准化考试也到了“百花齐放、百家争鸣”的时代。如日本推出的以下四种考试。

①日语能力考试(日本語能力試験 JLPT)^①,由「独立行政法人 国際交流基金」和「公益財団法人 日本国際教育支援協会」开发,面向世界施测。

②实用日语鉴定考试(J.TEST 実用日本語検定),由「日本語検定協会・J.TEST 事務局」开发,主要面向我国等亚洲地区施测。

③GNK 日语检定考试(GNK 生活・職能日本語検定試験),由「公益財団法人 国際人材開発機構」开发,已在我国开始施测。

④日语计算机自适应性考试(コンピューターによる自動採点日本語テストJapanese Computerized Adaptive Test, J-CAT),由筑波大学的今井新悟教授、山口大学的赤木弥生副教授等课题组利用日本国家科研基金开发,面向世界施测。

我国国内的则有:

⑤大学日语四、六级考试,由教育部委托“大学日语四、六级考试设计组”开发,已经施测了20个年头。

⑥日语专业四、八级考试,由教育部委托“日语专业四、八级命题组”开发,也早已在国内施测。

在基于网络的信息化时代,如果将上面的各种考试做个分类的话,除了“日语计算机自适应性考试”使用电脑终端外,其他均属于传统的纸笔媒介测试。测试水平的提高与发展,除了对语言本质、测试理论的认识加深之外,还有赖于相关技术的进步。计算机技术在测试中的应用,堪称使测试发生了革命性的飞跃。

3 测试的理论依据

任何一种语言测试都是一定的语言学理论、教育测量学理论(心理学理论)和教学内容及其教学法的体现。一般而言,大规模标准化考试所依据的测试理论有两个,一个是“经典测量理论”,另一个是“项目反应理论”。对于这两个理论,我们下面只做提纲挈领性的提示:

3.1 经典测量理论

经典测量理论,亦称真分数理论,以考试实得分数为前提条件,以真分数和平行复本为基本概念。在这一理论的基本框架内,建立起测验项目的计量学指标,如测验的信度、效度、题目的难度、区分度等,并以此来筛选试题,编制试卷或建立题库,及其考后项目分析等。它于十九世纪末开始兴起,二十世纪30年代形成比较完整的体系而渐趋成熟。50年代格里克森的著作使其具有完备的数学理论形式,而1968年洛德和诺维克的《心理测验分数的统计理论》一书,将经典测量理论发展至巅峰状态,并实现了向现代测量理论的转换。经典测量理论经过大量的实践,为测试迈向现代化和科学化做出了重大贡献。

尽管如此,在心理和测量领域,人们一般认为经典测量理论有四个弱点。第一是样本依

① 此处指(旧日语能力考试)(下同)。

赖。这是经典测量理论的最大弱点,被试特性和测验特性相互依赖;测验或题目的难度依赖被试,而被试的能力水平又依赖于测验或题目的难度。第二是信度和误差问题。第三是无法知道被试在单个题目上的表现。第四是不能“因人施测”。因此被认为测试欠精确度。

3.2 项目反应理论

项目反应理论也称潜在特质理论或潜在特质模型,是针对经典测量理论的局限性(弱点)而提出来的,是一系列心理统计学模型的总称。这些模型的目标用以确定潜在心理特征(latent trait)是否可以通过测试项目被测试出来,以及测试项目和被试之间的互动关系。即,项目反应理论认为被试在测试项目上的反应和成绩与他们的潜在特质有必然关系。

与经典测量理论相比较,项目反应理论就被试对项目的反应行为做了更强的假设,由此显示出以下优点:(1)有可能不依赖被试样本而对项目难度做出描述。(2)有可能不依赖项目样本对被试能力做出描述。(3)有可能估计出被试在新项目(下一试题)上的正确反应概率。项目反应理论的最大特点是:找到了一条项目特征曲线(ICC),并且以数学函数表达式(数学模型)来描述它,逼近它。不同的数学模型对曲线有不同程度的逼近,也会有不同的参数。项目特征曲线包含两个方面的参数:(1)对测试项目的特征进行刻画的项目参数。(2)对被试特征进行刻画的潜在特征或特质参数。

因此,项目反应理论中测验的项目参数具有跨群体不变性,具有恒久性的特点。也就是说,在项目反应理论中,即使抽取不同被试样组,估计出的项目参数仍然是相对不变的,所以任何能力的一组被试的估计参数,都能运用于其他被试组,包括全部被试。项目反应理论的这一优良性质为建设大型题库,编制各种测验提供了方便,使计算机自适应性考试的推出成为可能。

但是,虽然项目反应理论具有很大的优势,它也有自身的不足:(1)由于其理论假设建立在较深奥的数学基础之上,所以在普及和运用上有一定的难度。(2)由于项目反应理论从测量模型的理论框架来讲,多采用1\0计分资料的单维模型,故造成其应用上的严重局限。(3)由于受到苛刻的假设限制,必须有大样本进行配合,否则精确性不高。(4)在应用过程中需以先进的电脑设备及技术等作为辅助。

上述的①(旧日语能力考试)②③⑤⑥种考试,应该是在经典测量理论下设计、开发的考试,而只有④是在项目反应理论下设计、开发的考试。这两种考试类型在测试技术上有本质的不同。

4 日语计算机自适应性考试(J-CAT)简介

4.1 何为 J-CAT

日语计算机自适应性考试(J-CAT;Japanese Computerized Adaptive Test)是利用因特网和计算机技术,测试日语熟练程度(水平)的一种考试。考试不受时间和地点的限制,只要具备网络环境,就可随时随地进行考试,而且免费。考试由听解、词汇、语法和读解四部分组成,题型为大家所熟悉的四选一的答题方式。

J-CAT 测试的是一般日语能力,即通用日语,而非专用日语,如校园日语、商务日语、科技日语、医护日语等。

测试对象为日本国内外日语学习者,学习者可以以个人登录方式参加考试,由此了解自

己的日语水平,过一段时间后(推荐半年以上),可再登录考试,了解自己的日语水平的进步情况。当然,学校、企事业单位也可利用这个考试,如大学的摸底、分班考试、成绩考试、企事业单位录用新人时的日语水平考试等。前者称个人考试,后者称团体考试。

考试各部分各100分,满分400分。答完试题考试结束后,显示器上即可显示出考试结果(所得成绩),提供PDF形式成绩单下载、保存及打印。考试不分级别,任何水平的学习者均可参加考试。因此无须提前决定考哪个级,考试系统根据被试的日语水平(答题情况)自动提供不同试题。所以,这种测试形式被称之为自适性考试(adaptive test)。

欲更具体了解,建议去该考试网站(<http://www.j-cat.org/>)亲自参加一次考试。

4.2 J-CAT 的构成

如上所述,所谓自适应性考试,是指计算机根据被试答题情况(对否),推断其能力,提供最适合该被试能力的试题。也就是说,试题既不难也不容易,其难度正好适合该被试。显而易见,自适性是指测试系统对不同的被试,根据其能力水平,自动选出适应其水平的试题测试。即考试“量体裁衣”“因人施测”。这个原理,打个比喻说,如同视力检查。

我们知道,视力检查时,通常使用一张视力检查表(表因国家不同而有异)。我们这里使用一张环形符号表(参见图1),环由大到小、自上而下排列,环上某部分开有一个开口。检查时,首先指一个大环,如果被检查者能看清开口所在,即答对,那么再指一个比这个小几个号的环,如果这个环的开口还能看清,那就再指一个更小的,直至看不清,即没答对,这时再回指一个大一号的,检查是否能看清开口所在,再据此结果,或上或下反复检测若干次,确定出被检查者的最大视力。最后,所能看清的那个最小的环边上写的那个数字,就是其视力了(参见图1)。

○	○	い	0.2
○	○	ㄣ	0.3
○	○	<	0.4
○	○	ㄣ	0.5
○	○	し	0.6
○	○	に	0.7
○	○	と	0.8
○	○	っ	0.9
○	○	τ	1.0
○	○	っ	1.2
○	○	と	1.5

图1 视力检查表

这张视力检查表每行边上写的那个数字,就相当于考试时的难度(难易度)。自适应性考试如同检查视力一样,通过交替出难度大(难题)的试题和难度小的试题(易题),来检测出被试能做对的最难试题。同视力检查一样,通过被试答题的正误情况,决定下一道出什么难度的试题。这一系列工作,都是由计算机代替人工进行的,所以称作计算机自适应性考试(Computerized Adaptive Test)。

图2是自适应性考试的试题推移与推定能力值变化的示意图。×表示答错,○表示答对。答错后会给出一道较简单的题目,如果答对,就会给出一道较难的题目,这样误差就会逐渐变小,难易度变化也逐渐越来越小,到接近无波动处便是最终能力值(参见图2)。

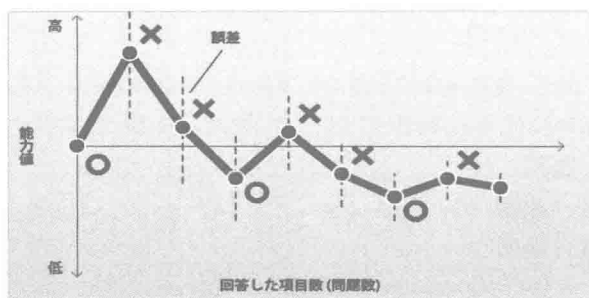


图2 自适应性考试的试题选择推移与能力值确定的示意图

试题由专家命题,然后进行预测。通过预测取得答题数据,针对每道试题,使用该数据,根据项目反应理论计算出难度(难易程度)和区分度(区别能力的程度),建立这样两个参数,标注到试题上。项目反应理论根据使用参数的数量不同,分单参数模型、双参数模型和三参数模型。J-CAT 采用双参数模型。把这种带两个参数的试题放到题库中,构建成考试系统,然后挂到网上形成网考。

考试开始后,系统分两次估计被试的能力值,首先,计算机选出若干题目,通过网络逐一发送至被试的电脑终端呈现在显示器上,被试开始做题,选择自己认为正确的选项,然后单击该选项,答题结果便会通过网络传回到考试系统,根据答题结果,计算机依据项目反应理论的算法大致估算出被试的水平,这是第一次估计被试能力值,也称粗略估计。在这个基础上,计算机从题库中抽出最符合该能力水平的试题逐一再发送到被试的电脑终端,通过这种反复,能力值就会逐渐收敛在一定的误差内趋于稳定,这时考试便结束,也就确定了能力值,即第二次估计被试能力值。如果出了相当数量的试题,不知何故,误差仍大于基准而不能呈逐步收敛变小,此时如果超过了一定的出题数量,那么也结束考试(难度高的试题能答对,而难度低的试题则又答不对,如果频繁地出现这种情况,误差就很难收敛变小)。

整个考试过程(以上的流程)均是自动进行的,所以考试的组织者、监考无须印制大量的试卷,不用回收答案,也不必阅卷评分,同时也就避免了阅卷误差等。试题的呈现方式,自适应性考试和传统考试如图3所示,完全不同。

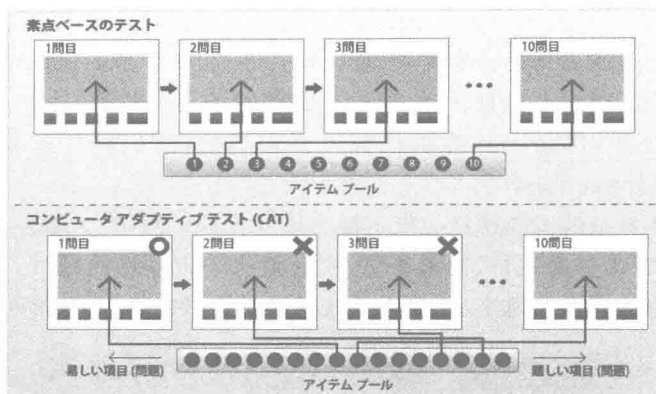


图3 传统考试与自适应性考试的试题呈现方式示意图

(上图为计分方式的传统考试,下图为计算机自适应性考试)

アイテムプール: 题库

易しい項目(問題): 易题

難しい項目(問題): 难题

5 试题形式(题型)

如前所述,考试分四个部分,不同部分、不同试题,答题时间各异。即每道试题都规定了答题时间,如词汇、语法试题,从短的30秒,到读解试题中比较长的文章5分钟,根据试题的难易度或篇幅的长短,答题时间各不相同。每道试题出现时,显示器右上方出现一个时间框,显示答题开始后时间逐渐递减。如果在答题时间内没有选择答案,便视为答错,并自动显示出下一道试题。

这一点就和纸笔考试大不相同了,能测出来被试在每一个项目上的语言熟练程度。我们知道,语言测试不仅仅只测试答对率,还要测试熟练(流畅)程度,计算机自适应性考试能很好地测试到被试的语言熟练程度。

5.1 题型举例

前文提到,题型是大家所熟悉的四选一形式。我们现在看一道“词汇部分”的试题,如下:

_____に入るものを選びなさい。

その場の雰囲気は_____言動が要求される。

a. ふさわしい

b. はなはだしい

c. ひさしい

d. まぎらわしい

下面再看一道“语法部分”的试题：

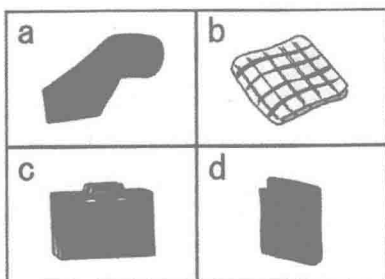
_____に入るものを選びなさい。

わたしは まいにち かいもの _____ 行きます。

a.が
b.で
c.に
d.を

我们再看一道“听解部分”的例子：

选项为影像形式，听与该影像有关的对话或一段讲话，看图选择正确答案。



试题首先播放提问，告诉被试要听什么。这是出于我们日常中的听解并不是漫无边际地什么都听，而是根据需要要选择地听。为此，听解开头首先播放提问，以使考生有目的地听。播放完提问后，马上播放试题的对话或一段讲话，然后再播放一遍提问，从选项中选择正确答案。

选项的给出，有的是图片形式，如上图（这里虽是黑白图片，但实际上是彩色的），有的是声音形式，采用声音形式时，显示器上显示“絵はありません。No picture in this question”，只给出“a.b.c.d.”，听完点击自己认为正确的选项。

最后，我们看一道“读解部分”的例子。

筆者の意見はどれですか。a・b・c・dから最も適当なものを一つ選びなさい。

ある中国人留学生(男性)が、日本社会の「男女平等」について下記のような意見を述べている。

「教育現場や制度上は男女平等社会が実現されたが、日本人が風習・しきたり等による古い考え方を引きずっているため、政界^{せいけい}や企業及び高等教育機関において、女性が要職^{ようしやく}を占める割合が少ないのではない^か」また、「特に日本の女性は中国に比べても、伝統的に生計を維持するために職業につくという意識が足りないのではない^か。女性が生計を維持する責任を自覚し、自らを高める努力の必要性を自覚することによって要職につく機会を拡大することができるのではない^か」と述べている。

注1) 政界: 政治または政治家の社会

注2) 要職: 職務上の重要な地位。

- a. 日本では、教育現場や制度上も男女平等社会が実現されていない。
- b. 政界や企業でも日本女性が要職を占める割合は少ない。
- c. 日本も中国も女性は生計を維持する責任を持っていない。
- d. 日本人は風習やしきたり等による、古い考え方を引きずっていない。

如上所示,读解的形式是,阅读一篇(一段)文章,然后回答一个问题,即读解均为一篇一问一答的形式。这与日语能力考试或其他考试的读一篇长文章,然后回答若干个问题的形式不一样。至于原因,我们将在后文提及。

5.2 题型亮点

(1) 彩照和动画的使用

计算机的应用,实现了一些纸笔形式的考试难以做到的一些形式,选项不仅可以使用彩色图片,还可以使用照片甚至动画,如:



何をしていますか。

- a. あくしゅをしています。
- b. はくしゅをしています。
- c. きょしゅをしています。
- d. はしゅをしています。

这原本是个使用动画的例子。这里印成了黑白颜色,但实际考试时是彩色的动画画面,显示两人握手过程。测试这种动作性动词或拟声拟态词最适合用动画。现在印在了纸上成了静止画面,握手这个动作就不太明显。

我们再看一个使用照片的例子,这是一位男性公司职员在出差前和上司说话的场面,使用照片能够使场面显示得一目了然(这里变为黑白的了,实际上是彩色的)。不言而喻,照片更能准确地显示试题的场面。



彩色照片和动画形式的试题,传统的纸笔形式的考试是难以实现的。一般而言,试题形式和真实语言生活的接轨程度越高,测试信度就越高。

(2) 一题一问

计算机自适应性考试的试题,一般只能一题一问,即便是较长的读解试题,也只能如此。是

因为项目反应理论的四个强假设之一所要求的,即“局部独立性假设。”局部独立性指考生在各道题目上的答对概率相互独立,即考生的潜在能力是影响作答的唯一因素,当排除这个因素的影响后,考生在不同题目上的作答行为之间不存在任何关系。但是在大规模外语测试中,局部独立性这一假设往往会被违反,因为常见的题型是几道选择题基于同一篇章。在局部独立性假设违反的情况下,项目分析不仅会导致模型与数据的不拟合,而且会导致对项目区分度的估值过高,从而影响测量的精确度。一题一问,在具体试题上也就避免了一题多问造成的相互纠结或难易度各异的弊端,使题目自身更独立,质量更高,更能准确测量出考生的实际水平。

6 J-CAT 的分数导出与水平参照

6.1 分数算出法

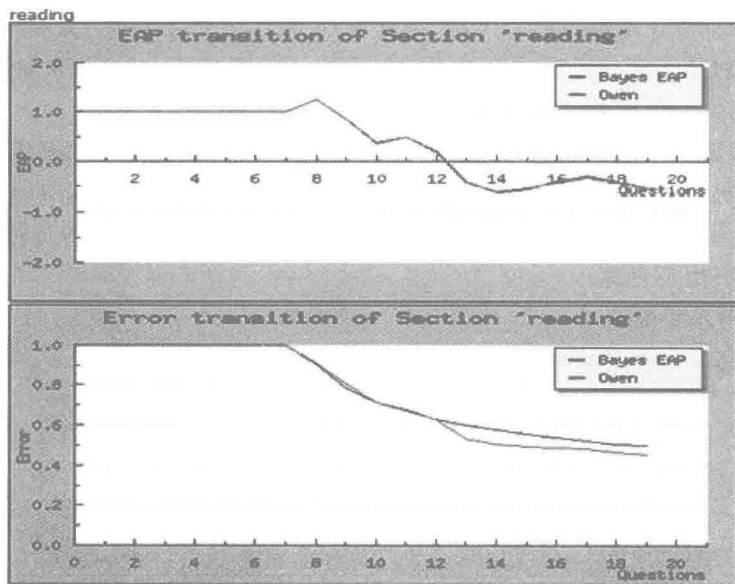


图4 能力推定与误差的收敛情形示意图

J-CAT 基于项目反应理论推定被试能力值,其过程如图4上下两张图所示,上面那张是能力值推定的推移,下面这张是误差值的推移。首先,决定能力值的假定值,让考生回答几道难度不同的试题,根据其答题结果而大致估算出,即前文提到的“第一次估计能力值(粗略估计)”。我们这里假定一个被试参加了这场考试,其答题结果如图4的上图所示,可见直至第7题,一直无变化以1.0推移,这部分就相当于假定值,即根据这部分的答题正误情形,而推定出假定能力值为1.0(第一次粗略估计)。在这部分答题正误情形的基础上,再看这之后的答题正误情形,即随着答题连续推算其后的能力值,在图上可以看到,数值在上下波动中,逐渐在-0.5附近呈收敛,于是在第19题结束了考试。这时的值就是最终的能力推定值(第二次估计)。同时,如下表所示,也可知误差为0.5左右。这种误差当然越小越好。如果在第19题不结束考试而继续出题,误差就会随之变小。但是,误差图表显示,误差虽然最初下降幅度很大,但其后愈趋平稳。这就意味着,即使再增加题量,误差也很难再缩小。所以考试不可能无限制地(也没有必要)持续下去,需要适可而止。这样就有必要决定一个基准作为终止规则。实际上误差和出题数量呈现一种鱼和熊掌不能兼得的关系。一方面,为了尽量提高测量精度,需

要大量出题,而另一方面,考虑到考生的负担,又想尽量减少题量。所以这就需要制定一个最合适的终止规则。那么如何决定这个规则呢?J-CAT通过模拟实验,分析考生的反馈及其酌情兼顾教学实际情况等,从而依据出题数量和误差值两方面而定,即其中的某一方达到了规定值时,便终止考试。

J-CAT如此获得的能力值,至今为止的数据显示,基本在 $-3.5 \sim +3.5$ 的范围。然而,这种基于项目反应理论所言的能力值,一般考生和使用考试者不易理解其意,所以有必要将其转换为大家所熟悉的满分为100分的数值。项目反应理论所求的能力值是以0为原点作比例尺度的,所以将其做线性变换,间隔也不会发生变化,因此可以如同偏差值一样换算成分数。J-CAT使用如下换算公式算出得分。

$$\text{得分} = \text{最终能力值} \times 15 + 50$$

用这种方法换算出听解、词汇、语法、读解四部分的得分,基本在0分至100分。理论上也应该有不满0分和100分以上的分数,但实际考试中几乎不出现。即便有这种极端结果出现,则将不满0分和100分以上的分数按照0分和100分给出分数。

关于得分换算公式中的系数“15”这个数值,是通过模拟实验的结果而导出的。这个系数如果过小,换算出的分数间幅度狭窄,得分上反映不出差别,也就意味着对所出试题即使反复答对,也考不到高分。另外,如果系数过大,就会出现超过想设定的得分幅度(这里为0~100分)的情形,产生不合理。

6.2 分数解释与水平参照

上文提到,考试结束时,考试系统会即刻提供考试成绩单(参看图5)。

成績証 Score report J-CAT® Japanese Computerized Adaptive Test <http://www.j-cat.org/>

あなたの J-CAT スコアは以下の通りです
Your J-CAT scores are as follows.

氏名 (Name)	Sora Hara	所属機関 (Institution)	St. Valentine University
ID 番号 (ID No.)	62000106	受験日 (Exam Date)	2012/02/14
受験番号 (Exam No.)	20120214	合計スコア (Total Score)	63

(watermark: dd039e6dd2082be1721f44c255397f44)

J-CAT.org

Interpretation of J-CAT score		
J-CAT score	Proficiency level	J-CAT 日本語能力試験
-100	Basic 初級	Level 4
101 - 150	Pre-Intermediate 中級前半	Level 3
151 - 200	Intermediate 中級	
201 - 250	Intermediate-high 中級後半	Level 2
251 - 300	Pre-Advanced 上級前半	Level 1
301 - 350	Advanced 上級	
350+	Near native 日本語母語話者相当	

* "Basic" learners can exchange basic ideas
 * "Intermediate" learners can manage daily communication
 * "Advanced" learners can manage academic and professional communication

J-CAT.org

图5 成绩单

成绩证书上给出了水平参照,即回答了“Interpretation of J-CAT score”的分数相当于什么水平?参照的是(旧)日语能力考试(JLPT)的相应级别。另外,为了方便考生或相关单位的理解,成绩证书下方还印有下列说明。但是,说到底这只是个参照,并不保证两者相关。这一点将在后文提及。

“初级”,学习者能表述自己的基本的想法。

“中级”,学习者能就日常生活进行会话。

“高级”,学习者能就学术及其专业进行交流。

J-CAT 的结果(分数)是基于项目反应理论换算出的能力推定值。在预测抽样时,考虑到要能反映出整体日语学习者的学习情况,所以各个部分得分50分、综合得分200分,就可认为这个分数就比较接近整体日语学习者的平均水平,即相当于所谓的“中级水平”。

J-CAT 提供的分数具有不变性,可以在不同批的被试、不同次的考试之间进行比较。但是这个成绩相当什么水平,其判断由使用者自行决定。比如说“中级”,所代表的水平也会因国家、地区或者教学单位的不同而有异。

所以,基于 J-CAT 的分数而划分水平,这完全由使用者酌情自行决定,J-CAT 不做决定。在这个意义上,可以说 J-CAT 不是水平认定考试,它致力提供可靠而可信的分数。尽管如此,作为“参照”,J-CAT 提供了与旧日语能力考试的级别挂钩的分数,是为了便于个体被试相对理解其水平,否则被试对其成绩将莫名其妙,但这仅仅只是个水平“参照”而已。

另外,就与 JLPT 的关系而言,系统开发者没有提供相关系数等统计数据,据说是因为 JLPT 的难度不稳定,每次考试起落不定,所以即便求得两者的相关性也无意义。现在已经推出新的日语能力考试,新的考试也是基于项目反应理论算出试题的难易度而进行等化,这样两个考试间就有了验证相关性的可行性。但是,毕竟两个考试形式不同、构想效度^①也不同,所以对两个考试的相关性解释,似乎也需要慎重。

教学单位利用 J-CAT 时,建议不要被这里所示的“参照”所左右,最好将 J-CAT 得分结合本单位的水平酌情予以解释和使用。J-CAT 的得分和一些具体教学单位的水平如何进行比较,感兴趣者,可阅读『J-CAT オフィシャルガイド』(详见参考文献)一书。

J-CAT 所能保证的是分数的不变性,本身不具备划分水平及其解释的功能。例如,有不少大学使用 TOEFL 和 TOEIC 考试的分数,作为学分的认定条件或作为入学条件,但设多少分以上为认定标准,均由各大学自行决定。这两个考试的举办单位美国教育考试服务中心(Educational Testing Service,ETS)提供的只是不变形的指标,即分数,其解释和使用方法则由各教学单位自行决定。J-CAT 的得分及其解释也与此完全一样。

7 计算机自适应性考试建设的特点

7.1 主要优势

计算机自适应性考试的理论基础是项目反应理论。项目反应理论与计算机技术两者紧密

① 语言测试中的构想效度,一般指对语言构成、语言能力在理论上的认识,即某个考试是依据什么语言理论开发的。