

日本研究

文集 7

百家出版社

日本研究文集

7

季林根 主编

百家出版社

图书在版编目(CIP)数据

日本研究文集(7)/季林根编著. —上海:百家出版社,2006.4

ISBN 7-80703-446-7

I. 日... II. 季... III. 日本—研究—文集
IV. K313.07-53

中国版本图书馆 CIP 数据核字(2006)第 026659 号

书 名 日本研究文集(7)

主 编 季林根

责任编辑 陈闵梁

装帧设计 王 俭

出版发行 上海文艺出版总社(www.shwenyi.com)

百家出版社(200032 上海市茶陵路175弄3号)

经 销 全国新华书店

印 刷 常熟市大宏印刷厂

开 本 850×1168 1/32

印 张 15.625

字 数 385 000

版 次 2006年4月第1版 2006年4月第1次印刷

ISBN 7-80703-446-7/K·22

定 价 38.00 元

责任编辑 陈冈梁
封面设计 王 俭

编者的话

《日本研究文集(7)》即将出版问世,此刻我感到十分高兴。回顾10年前出版的第一本论文集,那只是一本薄薄的小册子,而且论文内容的覆盖面也相当有限。近年来,随着国内日语界对日本研究的深入开展,我们的教师与国内同行们的交流日益频繁,日本文化经济学院的科研氛围也越来越浓,教师们的视野愈加开阔。与往年相比,如今我们研究的课题涉及面更广,涵盖了日本语言文学、社会、文化、经济等各个方面。而且研究的角度也从过去的概念上的剖析,提升到了深层次的比较研究、语言的社会文化背景研究等更深的层面。可以说日本文化经济学院在科研方面已经迎来了一个新的时期。

本论文集收集了我院教师在2003年度和2004年度科学报告会上发表的科研文章以及他们近期的研究成果,共45篇。其中语言类21篇、文学类6篇、文化类8篇、经济类6篇、日语教学类2篇、社会类2篇。

我相信,只要大家共同努力,我们的科研工作一定能更上一层楼,《日本研究文集》也会更具有学术水平,进一步发挥其学术交流的作用。然而,由于水平有限,汇编工作中难免存在不足或错误之处,恳请各位原谅。

2005年11月

目 录

全国日语专业四、八级测试的信度及效度研究

..... 谭晶华 杨础人(1)

我国日语教育的回顾 王 宏(27)

语言文学类

「は・が」の使い分けについての教え方 皮细庚(37)

「思う」と「考える」の意味用法について 沈宇澄(48)

連語と連語論 吴大纲(57)

论语義在日语的遣词造句中的作用 许慈惠(71)

谈谈现代日语动词的分类 许慈惠(81)

日本語の心理動詞について 周 星(89)

受身構文における中国語と日本語の表現形態に

ついて 俞瑞良(97)

「XはYが」構文 徐志强(106)

コーパスによる実証的研究の信憑性を高め

るために 毛文伟(121)

機能辞性の判定基準について 毛文伟(133)

迂言的な受身表現 凌 蓉(148)

日本語と中国語の受身文の機能について 凌 蓉(157)

「怒」表現への考察(一) 关 薇(167)

「怒」表現への考察(二) 关 薇(179)

「手紙文」から見た日本語と中国語における敬語

表現について	林 彬(186)
日中対照から見る日本語受身文の使用	陆 洁(195)
日本語教育におけるマルチメディアの導入	顾伟坤(204)
中国語の三声は二音節である	王 磊(213)
日本語肯否表現における非対称性の関与要素 ...	王华伟(219)
肯否表現における対称性と非対称性	王华伟(229)
条件を視点化する「と」の用法	徐 卫(236)
「風の歌を聴け」	高 洁(246)
大江健三郎の目に映った新興宗教	徐 旻(251)
大江健三郎文学に見られる「周縁性」	徐 旻(263)
川端文学における伝統性とモダニズムの位置 ...	何文晔(274)
三好達治と「花筐」	许 晨 陆静华(281)
『生れ出づる悩み』における「海」	张昕宇(290)

社会、文化、经济类

日本の議員選挙制度について	刘玉飞(303)
日本人の排他意識について	季林根(324)
日本人の食卓に見られる風土の影響	季林根(334)
五山文学について	顾伟坤(344)
企業文化のパワー	蒋步青(353)
企業文化の見直し	蒋步青(362)
日本のマンガ文化	焦毓芳(373)
日本のポピュラー文化が中国における影響に について	焦毓芳(381)
洛陽・竜門石窟の奉先寺と奈良の東大寺	孙宗明(390)

「プラザ合意」の啓示	陈小芬(405)
日本の新しい企業統治システムについて	王 萍(420)
環境の激変と日本的経営の変容について	徐宝妹(431)
統計データからみた日本の高齢化問題	张建华(440)
高齢者労働供給規定要因の一考察	张建华 王 娟(449)
日本経済は復興になるのか	瞿晓华(467)
试论泡沫经济的本质特征	瞿晓华(477)

A Study of Reliability and Validity on China's Graded Test for Japanese Majors 4 & 8

Tan Jinghua Yang Quren

全国日语专业四、八级测试的信度及效度研究

谭晶华 杨隹人

Abstract: The present paper makes an analysis of the answers to different types of questions in China's Graded Test for Japanese Majors 4 & 8 (2003). Based on the students' answers, the paper analyzes how well the students master Japanese, identifies the existing drawbacks of the test, studies how the reliability and validity of the test can be improved in the future, and puts forward some practical suggestions on the revision of the test syllabus, the perfection of the test's content and the improvement of the reliability and validity of the test.

提要: 本文对 2003 年全国日语专业四、八级考试的各种答题数据进行了分析,同时根据答题情况剖析了目前考生的实际情况,找出了目前尚存在的不足之处,对今后如何提高考试信度和效度的问题进行了研究,提出了修改考试大纲、完善考试内容等可以切实提高信度和效度的建议。

Key words: test, reliability, validity, level of difficulty, discrimination

关键词:测试;信度;效度;难易度;区分度

1. 研究目的

日语专业自 2002 年起,正式启动了大学日语专业四级(大学本科日语专业二年级水平)和八级(大学本科日语专业四年级水平)考试。实施这一考试的目的在于检查日语专业教学大纲的落实情况,对全国日语专业的所有教学点的教学质量进行全面、客观、公正地评估,如实地把握全国各日语专业点的教学质量,以及考察学生掌握外语能力和知识的情况,便于指导委员会乃至教学行政机关对外语教学的指导和约束,鞭策和促进各专业点提高教学质量,保证有限的可利用资源发挥更大的效益,培养出更多更好的外语人才。

但是,假如日语四、八级考试没有一定的信度和效度,也就是说,我们的考试如果不能测出考生的真实水平,那就无法达到上述目的。为此,有必要对已经实施了 2 年多的日语四、八级考试本身进行一番评估,在客观评估的数据上进行分析研究,找出不足,并在此基础上研究出一套提高日语四、八级考试的信度和效度的方案,以便能真正客观、公正地测出考生的真实水平,达到检查日语专业教学大纲落实情况的目的。

2. 本研究的依据

用于本研究的依据是对 2003 年全国日语专业四、八级考试全体考生答题情况的原始数据。日语专业四、八级考试虽然始于 2002 年,但该年四级考试因手工阅卷,未积累有关数据,无法进行统计分析;该年的八级考虽已利用光电管阅卷,但当时所用为老式阅读机,数据亦未能妥善保管;而 2004 年的赶不上本次研究所用,故此次研究只能利用 2003 年一年的四、八级考后得到的原始数据。

3. 本研究的数据

用于本研究的原始数据是 2003 年四、六级考试的答卷数据。该数据包括利用光电阅读机对 2003 年的四、六级考试客观试题答卷阅卷后得到的电子数据和对主观题进行人工阅卷所得主观成绩两种(六级考试的主观题有作文和翻译两种)。该卷面成绩数据为 BDF 文件。

我们对 2003 年的四、六级考试卷面成绩原始数据的处理情况是这样的:

首先,我们将此数据转化为文本文件,然后利用广东外语外贸大学文科基地曾永强教授专门编写的项目分析软件进行项目分析统计。分析主要为项目分析,内容包括难易度和区分度分布表、总成绩统计数据、各大题统计数据、各大题的相关分析以及所有客观试题各小题的各类分析数据。

上述项目分析统计委托广东外语外贸大学文科基地的曾永强教授完成。

4. 项目分析数据的分析

下面根据项目分析得到的四、六级考试的难易度和区分度分布表、总成绩统计数据表、各大题统计数据表、各大题的相关分析表等所列项目分析数据进行逐项分析。通过分析找出问题,然后再在第 5 节中探讨解决问题以及提高考试信度和效度的方策。

4.1 四级考试客观题项目分析数据的分析

4.1.1 总体分析

首先由表 1 可以看出两点情况:一、整套题目的难易度偏易:难度大的题目偏少,尤其是很难的题目应该有 5 题,但实际为零,偏难的题目应该有 14 题,实际为 4 题;难度低的题目偏多,尤其是

表 1:2003 年四级考试的难易度及区分度分布表

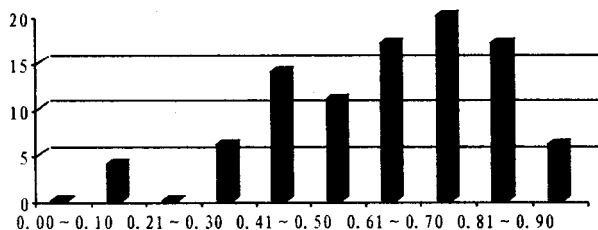
难易程度	答 对 率	小题数	合格题数	应合格题数
很难	0.00 ~ 0.10	0	0	5(5%)
偏 难	0.11 ~ 0.20	4	0	14 (15%)
	0.21 ~ 0.30	0	0	
难度 适中	0.31 ~ 0.40	6	4	57(60%)
	0.41 ~ 0.50	14	11	
	0.51 ~ 0.60	11	9	
	0.61 ~ 0.70	17	10	
偏 易	0.71 ~ 0.80	20	16	14 (15%)
	0.81 ~ 0.90	17	14	
很易	0.91 ~ 1.00	6	4	5(5%)
合 计		95	68	95

表 2:总成绩统计数据表

试题数	95
分 值	
平均分	61.63
标准差	11.43
难 度	0.65
区分度	0.12
信 度	0.87
偏态值	-0.05
峰 值	-2.93

注:合格题指正确答案的区分度为 0.30 以上的小题。

偏易的题目只能有 14 题,但实际达到了 37 题,大大超出标准。由此导致难度适中的题目偏少,应该有 57 题,而实际只有 34 题。整套试题的难易度分布情况如下图。



其次,区分度达 0.30,即能区分考分高和低的考生之合格题目偏少,约占整套共 95 小题的 71.6%。通常,区分度小于 0.30 的小题数不能超过整套题目总题数的 5%(李 1997:127)。难易度分布与各段合格题数对比情况见下图。

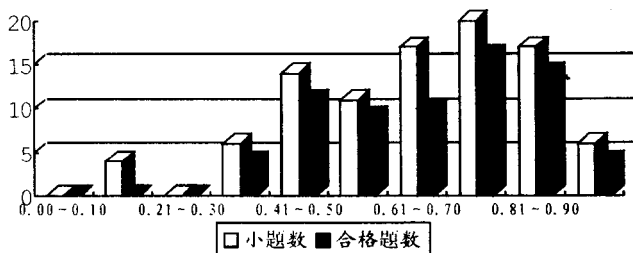


表2的数据也证实了上述判断。首先从难易度来看,通常的难易度应该为0.55~0.60,而本套试题的整体难易度为0.65,这表明了整套试题偏易了些。这一点从偏态值为-0.05可以得到证实,因为假如题目难易度适中的话,偏态值应为0(而当偏态值为正值时,则表明该套题偏难)。

再看一下成绩分布的峰值。理想的峰值应为0,但如果能控制在 ± 1.0 之内,其分数基本符合正态分布(李1997:124)。而本考题的峰值为-2.93,可见本套题的考生成绩分布曲线的高峰低于正态分布曲线的高峰,表明中等成绩考生的考分过于分散。

由于以上原因,总区分度只有0.12,低于0.30的要求。信度为0.87,低于0.90的标准。

4.1.2 听力题分析

好的听力题无疑可以测出学生获取口头信息的能力,而且可以考出学生能否在短时间内作出正确反应的能力。这次听力题情况,由表3可以看出,整体来说略偏易,根据各小题分析数据反映的答对率来看,答对率最低为0.40,其中有2道题分别达到了95%和94%。达到95%的那道题区分度仅0.27,为不合格题。整套大题的偏态值仅-0.02,所偏不大,基本上属于正态分布。但是峰值为-2.98,表明听力题的考生成绩分布曲线的高峰低于正态分布曲线的高峰,说明了中等成绩考生的考分过于分散。

在区分度方面,虽然总的区分度仅 0.28,但接近 0.30。仔细看一下各小题分析数据,发现 25 题中,小题区分度低于 0.30 的仅 4 题,其余 22 题中,有 18 题的区分度均达 0.40 以上。这表明 25 道听力题虽然偏易了些,但基本上能够区分出考分高的考生和考分低的考生。但是,所有各小题的干扰项的区分度均为负值(此表未列印,下同),表明干扰不得当,这在一定程度上影响了区分度。

表 3:听力题统计数据表

试题数	25
分 值	1
平均分	15.91
标准差	4.59
难 度	0.64
区分度	0.28
信 度	0.78
偏态值	-0.02
峰 值	-2.98

4.1.3 语言知识题分析

由于文字、词汇、文法等问题都是测试学生所习得的语言知识,而非语言能力,因此,我们将它们归并到一起进行统计分析。

由表 4 可知,知识题的题目量较大,达 50 题。从这 50 题的难易度来看,虽然有 3 题答对率低于 0.20,但超过 0.90 的达 4 题,其中 2 题答对率高达 98%,整个知识题的难易度达到了 0.68,且偏态值为 -0.05,因此总体来说还是偏易了些。

峰值为 -2.96,如同听力题一样,表明考生成绩分布曲线的高峰低于正态分布曲线的高峰,说明了中等成绩考生的考分过于

表 4:文字、词汇、文法题统计数据表

试题数	50
分 值	1
平均分	33.77
标准差	6.08
难 度	0.68
区分度	0.37
信 度	0.78
偏态值	-0.05
峰 值	-2.96

分散。

语言知识题的整体区分度达到 0.37, 已经高出 0.30 的合格线。但低于这条合格线的小题有 13 题。同样, 所有干扰项中, 仅 4 题中有个别干扰项的区分度为正值, 其余 46 题的干扰项全部都是负值。

4.1.4 读解题分析

虽然在这读解题中包含了两种类型的题目: 一是长篇读解, 共 10 小题, 还有一种由 4 篇短文组成的短篇读解, 也是 10 小题。均用于测试考生的阅读理解能力, 即获取书面信息的能力, 可见这些读解题都是测试语言能力的。

由表 5 数据可知, 四级读解题的难度适中, 偏态值仅为 -0.01, 已非常接近正态。但峰值与前面的几道大题一样, 显得太低, 为 -2.98, 这表明读解题的考生成绩分布曲线低于正态分布曲线的高峰。

表 5: “读解”统计数据表

试题数	20
分 值	1
平均分	11.95
标准差	2.89
难 度	0.60
区分度	0.12
信 度	0.53
偏态值	-0.01
峰 值	-2.98

本大题的区分度亦较低, 为 0.12。在 20 小题中, 区分度低于 0.3 的达 7 题, 20 小题的 60 个干扰项中, 只有 2 个干扰项为正值, 4 个干扰项为 0, 其余均为负值。

4.1.5 相关系数分析

下面探讨各大题之间的相关系数。

首先看一下客观题之间的情况。由表 6 可知, 客观题整体与听力、语言知识(含文字、词汇、语法)、读解(含长、短篇读解)之间的相关系数均大于 0.70。其中客观题整体与听力之间的相关系数

表 6:四级题相关分析表

		客观	听力	文字	读解	主观
客观	相 关 系 数	1.000	.822	.897	.762	.087
	意义水平(2 - tailed)	.	.000	.000	.000	.000
	样 本 数	3834	3834	3834	3834	3834
听力	相 关 系 数	.822	1.000	.549	.508	.068
	意义水平(2 - tailed)	.000	.	.000	.000	.000
	样 本 数	3834	3834	3834	3834	3834
文字	相 关 系 数	.897	.549	1.000	.573	.073
	意义水平(2 - tailed)	.000	.000	.	.000	.000
	样 本 数	3834	3834	3834	3834	3834
读解	相 关 系 数	.762	.508	.573	1.000	.081
	意义水平(2 - tailed)	.000	.000	.000	.	.000
	样 本 数	3834	3834	3834	3834	3834
主观	相 关 系 数	.087	.068	.073	.081	1.000
	意义水平(2 - tailed)	.000	.000	.000	.000	.
	样 本 数	3834	3834	3834	3834	3834

* * Correlation is significant at the 0.01 level (2 - tailed).

达到 0.822,与语言知识之间的相关系数达到 0.897,与读解之间的相关系数达到 0.762。由于客观题整体与听力、语言知识、读解之间是整体与部分之间的关系,因此它们之间的相关系数达到 0.7 以上,为相互关系密切,这是正常的,也是应该的。

各种能力或知识(听力、语言知识、读解)之间的相关系数情况是:听力与语言知识之间的相关系数达 0.549,与读解之间的相关系数为 0.508;语言知识与读解之间的相关系数达 0.573。这些数据表明各种能力或知识之间的相关系数均大于 0.50,介于 0.40~0.70 之间,表明它们之间有一定的相互关系。这种关系也是正常的,因为假如两者的相关系数太低,则表明其中某一方与实

际情况有一定的差异,没有考出水平。反之,假如相关水平太高,则表明只需测试一方,就可以测得两种水平,即其中有一方是多余的。何况,由于测试的是各不相同的3种能力,而人的各种能力也不可能都比较一致,因此它们之间的相关系数介于0.40~0.70之间是正常的。

下面再看看主观题(作文)与客观题(包含客观题整体、听力、语言知识、读解)之间的相关系数。其中作文与客观题整体之间的相关系数为0.087,与听力之间的相关系数为0.068,与语言知识之间的相关系数为0.073,与读解之间相关系数为0.081,均低于0.20,表示它们之间关系较少。正因为关系少,表明了很有必要对学生的写作能力进行测试。由于客观题(无论是语言知识题还是语言能力题)都是以理解为主的,而作文题则是以表达为主,说明了学生的理解能力与表达能力之间存在着较大的差异,至少有相当部分的考生并未同时具备这两种能力,或者是考生的表达基本功从整体而言尚有欠缺。与后面八级测试的相关分析数据相比,四级的客观题与主观题之间相关系数大大低于八级的同类相关数据,这是因为大多数院校的作文训练课程在三年级开设,故而基础阶段学生这方面的水平相对较低。

另外,我们还可以从意义水平这个角度看出这次测试的有效度。众所周知,当意义水平小于0.05时,表示偶然因素很小,具有统计学意义,而当意义水平小于0.01时,则表示非常有意义。由表6可知,所有大题之间的相关因素的意义水平都是零,这表明了他们之间的相关情况都是很有意义的,这也表明了,所有这些测试内容都是必要的。

4.2 对八级考试客观题项目分析数据的分析

4.2.1 总体分析