

doi:12.3969/j.issn.1672-0598.2012.04.024

雅思与华威阅读测试对比研究^{*}

冯 兰

(重庆交通大学 外国语学院, 重庆 400074)

摘 要:雅思和华威考试都是专为拟进入以英语为教学目标语言的欧美高等院校进行学习深造的学生提供的尺度参照性考试。本文从文本表现形式、难度系数、主题选择、阅读理解任务的真实性、题目数量、题型类型等方面对两者阅读测试部分的效度和信度进行了对比分析。

关键词:雅思;华威;阅读测试;效度;信度

中图分类号:H310.4 **文献标志码:**A **文章编号:**1672-0598(2012)04-0135-06

一、引言

众所周知,雅思考试(IELTS: International English Language Testing)由剑桥大学考试委员会、英国文化协会及澳大利亚教育国际开发署共同举办,专为拟进入以英语为教学语言的欧美院校进行深造的学生和移民至以英语为母语(或官方语言)的国家的人士进行英语水平测试而设立的国际化考试体系,考试分学术类(Academic)和移民类(General training)两种,考试内容包括听力、阅读、写作和口语四个部分,是当今世界最负盛名的考试之一。自1990年4月雅思在中国推广以来,每年数百万考生报名参考,其中相当一部分是计划本科毕业后出国留学的在校大学生,规模和数量之大可见一斑。

相对雅思而言,华威英语语言测试(WELT: Warwick English Language Test)晚于雅思进入中国(2008年)。华威英语语言测试于1980年由英国著名大学华威大学创办,由英国考试委员会语言专家设计考试题目,华威英语语言测试因其设

计的科学性和严谨性,成为除雅思、托福考试外被英国大学广泛认可的权威性考试。华威英语语言测试(WELT)考试时长2小时45分钟,考试题型分语法、写作、阅读三部分。针对不同的考试对象,华威英语语言测分为G类(code G)和FP类(code F)。前者针对申请攻读大学本科课程和硕士研究生课程的考生,针对适用于大学预科学生。

阅读测试是雅思和华威及其它英语测试的重要组成部分。根据笔者统计,在1994—2011年之间,中国知网数据库中分析、研究两种考试的文章达300多篇,但却鲜见对两种语言测试进行对比研究的文章,而专门针对两类考试中同一题型进行对比研究的文章更是少见。因此,本文拟对雅思考试学术类与华威语言测试G类中的阅读测试部分进行比较。

二、雅思、华威成绩等级标准

语言学术界对语言测试进行了不同类型的分

* [收稿日期]2012-04-15

[作者简介]冯兰,女,重庆人;硕士,重庆交通大学外国语学院讲师,主要从事语言测试及第二语言习得研究。

类。Valette(1977)将语言测试分为四类:学能测试(aptitude test)、阶段成绩考试(progress test)、成绩测试(achievement test)和水平测试(proficiency test)。其中,学能测试细又分为分班测试(placement test)和诊断性测试(diagnostic test)(Harrison,1983; Alderson et al,1995)。按上述标准划分,雅思和华威属于水平测试。同时,雅思和华威也属于尺度参照性考试(criterion-referenced assessments),二者参照一定的尺度对学生进行考核。学生的成绩不是通过与其他考生的比较,而是通过与某种特定的尺度进行比较而得到体现,学生的成绩必须达到某一特定尺度方能及格。这种特定尺度的选择或制定是尺度参照考试的关键(杨惠中、金艳,2001)。在成绩的评判上,雅思和华威采用不同的评分方式对考生的语言能力进行评判:雅思采取分数制(4.5~9.0分),华威实行ABCD制。

表1 雅思、华威考试分数对照表

雅思	华威考试
9.0	
8	AAA
7	AAB/ABB/BBB
6.5	BBC/BCC
6	CCC
5.5	DCC
5.0	DDC/DDD
4.5	

从表1可以看出,雅思成绩分为八个等级,最低分4.5分,最高分9.0分,最终提供给考生的成绩为各项(听力、阅读、写作、口语)得分总和的平均值。华威的计分方式则是针对考试的三种题型(语法、写作、阅读)分别给出等级(每项得分分为ABCD四个等级),三部分汇集在一起则为其最终成绩。表1还显示,雅思得分等级离散度大,反映考生个体差异更清楚。因此,雅思更能体现考生个体间成绩的差异。华威语言测试虽能反映考生在语言各方面的强项和弱项,但等级离散度小,难以区分个体间语言能力的差异。著名语言测试家Heaton(1991)认为,英语测试的发展趋势是使用

数字来评定评分等级,以此区分成绩离散度。由此看来,雅思的评分体系较华威评分体系更科学合理。

三、雅思阅读测试与华威阅读测试对比

下文将采用雅思和华威语言测试官方网站所提供的考试样卷作为数据分析样本。雅思和华威中的阅读部分测试时间相当,要求在一个小时内完成三篇学术类小散文的阅读和答题。

(一)表面效度

Harrison(1983)认为表面效度指试卷表面形式的可信度和公众对考试的接受程度。同时表面效度也体现为一种文本表现形式(Cumming and Berwick,1995:p3)。文本形式分为线性文本和非线性文本两种。线性文本以纯语言形式,即语句和段落来表达文本思想,非线性文本则在纯语言文本的基础上,通过穿插标题(或小标题)、前言、表格、图表等形式表达文本主题(Nuttall2002:p45)。笔者通过对2012年1至3月雅思和华威官方网站上各自三套测试真题的分析和研究得出以下数据。

表2 雅思阅读中非线性文本类型

	数量	百分比
文章标题	11	57
图表	7	30
节标题	3	13
总数	23	100

表2显示,雅思阅读文本属典型的非线性文本。三篇阅读文本中文章标题占57%,图表占30%,节标题占13%。相反,华威阅读试题从头至尾使用Times New Roman字体格式,中间没有任何变化,字体格式严肃、正式,从视觉上给读者以学术化氛围的浓厚感觉(Sanchez,2003)。而雅思阅读文本从形式上看,文字颜色有深有浅,文本中既有表格又有图片,形式丰富多样,表面效度更趋明显。由此可见,雅思阅读测试表面效度高于华威。

(二)文章长度、答题数量和易读性

表3 文本的定量比较

	雅 思			华 威		
	1	2	3	1	2	3
文章数量	1	2	3	1	2	3
文章长度	1 065	882	635	650	650	650
段落数量	7	9	7	6	4	10
相关关联问题数量	14	14	12	11	11	8
易读性	33.58	40.38	34.5	56.83	27.74	53.43

表4 易读性对照表 (Source-Michigan English Language Test Battery, 1996)

易读性分数	阅读难易程度
90 ~ 100	很简单
80 ~ 90	简单
70 ~ 80	较简单
60 ~ 70	标准
50 ~ 60	较难
30 ~ 50	难
0 ~ 30	非常难

笔者对2012年1月雅思和华威两套试题中六篇阅读文章进行了对比分析。从表3看出,雅思考试的文章长度在635~1065之间,而华威考试的文章长度非常一致,每篇都是650词。相反,雅思的三篇文本长度由长至短,词数依次递减。在易读性方面,表3表明,雅思考试的难度系数在33.58~40.38之间,属于“难”范畴(参见表4)。华威考试在难易程度安排的顺序上体现了以下规律:第一篇和第三篇难度系数分别为56.83和53.43,属于“较难”范畴,第二篇为27.74,属于“非常难”范畴。通过表4得知,易读性分数越低,阅读文章的难度就越高。华威测试的这种设计方式能否真正测试读者的阅读能力有待我们进一步思考。

语言学家Armstrong(1980)曾做过一项实验,分析高等院校管理学专业期刊杂志的易读性。实验结果表明(参见表5),在分析样本中,一本期刊易读性为54.3,属于“较难”范畴,三本期刊介于

20.2~29.4之间,属于“非常难”范畴,七本期刊介于31.7~46之间,属于“难”范畴。由此可见,63%的高等院校学术期刊属于“难”这一范畴,只有10%属于“较难”范畴和27%属于“非常难”范畴,而雅思考试的难易度都均属“难”的范畴。通过对比可以看出,雅思考试的文章难易度基本符合高等院校专业学习中的文本难易度,而华威考试的第一和第三篇不能完全反映学术文章的复杂性。

表5 管理学专业期刊易读性

期刊名称	易读性	期刊名称	易读性
Administrative Science Qtlly	20.2	Advanced Management Journal	46.0
Harvard Business Review	31.7	System and Procedures Journal	32.8
Academy of Management Journal	38.7	Business Horizons	29.4
California Management Review	32.6	Personnel	35.5
Industrial Relations	23.3	Supervisory Management	54.3
		MEAN SCORE	34.45

(三) 文本真实性与文本主题

考试文本的真实性是反映文本内容可信性的重要标准。考试的真实性包含测试材料真实性和测试任务的真实性,要求测试应尽可能地复制出现实生活中语言使用的各种特征(Bachman, 1995)。其中测试材料的真实性指测试中的阅读文章是否直接来源于现实生活中的文本素材,如报纸、杂志、互联网等,而不是指为了某一测试目的而专门杜撰和编写的阅读文章。华威阅读文本皆标注了来源,但雅思阅读文本更像专为考生撰写的学术文章,未提供出处来源。相比之下,华威考试更具真实性。

除文本材料的真实性外,阅读文本主题也是值得关注的重要因素。主题是阅读材料所表达的主要思想,对主题的分析可以帮助读者发现阅读材料话题所覆盖的范围是否宽广,是否带有命题者的主观偏向。

表6 雅思阅读测试主题统计

	Passage 1	Passage 2	Passage 3
Test 1	蝙蝠研究	科技发展	学习与心理
Test 2	日本塔	食物成本	农村交通
Test 3	蚂蚁智慧	人口与基因	欧洲森林保护
Test 4	金字塔研究	渔业研究	噪声影响

笔者选择雅思和华威考试2012年1至4月考题中的阅读文本素材进行了对比分析。从表6看出,雅思阅读主题包含自然科学类、人文社科和科技类,范围较广,具有较大的覆盖性。华威阅读

主题主要集中在生态学和环境学两个方面,涉及面较窄。一般而言,任何一种语言阅读测试的主题设计不应仅给熟悉某一领域的考生提供便利(Alderson (2000)),更应充分考虑内容图式(content schemata)给考生带来的影响(Cohen (1994))。显而易见,雅思考试在文本主题选择上充分考虑了大众考生的学科背景。

(四) 阅读题型、题目数量及阅读理解任务的真实性

表7是笔者对雅思和华威2012年1月的两套阅读考题做出的分类统计。

表7 阅读理解任务分类统计

阅读结构	问题类型	雅思问题数量	华威问题数量
泛读 (理解文章大意)	多项选择题	40	19,23,28
	配对题	5	
	简答题		1, 2, 3, 4, 5
	多项选择题	6,15,16,17	16,18,20,22,24,26
精读 (理解具体内容)		7,8,9,10,11,12,	
	配对题	13,14,22,23,24, 25,26,27	
	正误判断题/未提及	18,19,20,21,36, 37,38,39	
		1,2,3,4,29,30,31, 32,33	
词汇识别	多项选择题		6,7,8,9,10,11, 12,13,14,15 (25)
衔接手段	多项选择题		17,27
文体识别	多项选择题		21,29,30
功能识别	选择题	34,35	

表7显示,雅思和华威考试的测试题型和题目数量存在较大差距。雅思考试的阅读题型包括多项选择、配对、正误判断、选择四种,华威考试只包括简答和多项选择两种。在华威阅读试题中,1至5题测试考生的精读技巧,考查考生查询文本具体信息的能力,其余25题为多项选择题。虽然多项选择题具有评分信度高、实施容易(Weir 1990),且可用来测试考生辨认定事实细节、因果关系、逻辑推理、正误判断和辨别区分能力的优点(刘建达,1994),但其缺点也较明显:利于作弊、猜测几率高、负向反拨大、结构效度低(Hughes,

1989:p60-62; Alderson, 2000:p211-212; Heaton, 2000:p27)。

雅思考试大量采用匹配题(占60%)和正误判断题(占20%),多项选择题只占总量的12.5%(见表7)。雅思考试采用标题段落匹配题不但能考查考生对语篇的宏观概括能力,而且能考察考生获取具体信息的能力。正误判断题的优点在于语篇选择和命题容易,能考查考生多个层次的阅读能力(Heaton, 2000:p114)。为避免考生在答题过程中50%对50%的猜测概率发生,雅思专门设计了“未提及”这一题型。这一题型对于大多

数中国学生而言,往往容易混淆“错误”和“未提及”这两个概念。因此,这对测试学生的阅读理解能力具有极大的效度。

在阅读题目的数量上,两者之间也存在一定差异,雅思为40题,华威为30题。任平、但武钢(2009)认为,当阅读测试题目数量达到50时,信度为0.7,当题目增加至75时,信度能提高到0.78。题数达到50以后,随着题数的增加,信度虽有上升,但上升幅度会愈来愈小。由此可见,雅思阅读试题信度接近0.7,华威则低于0.7。因此,雅思阅读无论在效度还是信度方面都大于华威。

在语言测试中,除阅读材料应具有真实性外,阅读理解任务也应具有真实性。语言学家 Bachman(1995)曾对“真实性”做出如下定义:“考生所承担的任务应该是现实生活中相互作用的交际活动,而不是对考官的刺激做出的典型的、例行的考试反应。”在华威考试中,阅读理解任务的真实性得到了体现。表7还显示,在考查学生对衔接手段和文章体裁的识别方面,17、27题两题侧重对承上启下的写作方法进行考查,而21、29、30题则侧重考查考生对不同体裁文章修辞手法的了解。相比之下,雅思则主要关注考生对阅读技巧的使用。因此,华威考试的阅读理解任务比雅思更具真实性。

四、结语

通过对雅思、华威两类考试中阅读测试部分的对比分析,我们清晰地发现,雅思阅读素材为非线性文本,华威为线性文本;雅思选材广泛、题目数量多、题型多样,但文本材料和阅读理解任务缺乏真实性;华威阅读文本虽素材真实,皆有文本来源,但阅读素材来源狭窄,题型单一,题数偏少,难度系数低。从整体上看,雅思阅读测试部分在信度和效度上高于华威。笔者认为,雅思、华威虽各有其优缺点,但都有相互借鉴、相互改进的空间:雅思在采用非线性阅读文本的同时,亦需进一步加强考试材料和阅读理解任务的真实性;华威在难易梯度设计、文本主题多元化方面尚存不足,还有一定的改进空间。

[参考文献]

[1] ALDERSON, J. C., CLAPHAM, C. and WALL, D. Language Test Construction and Evaluation[M]. Cam-

bridge: Cambridge University Press, 1995.

- [2] ALDERSON, J. C. Assessing Reading [M]. Cambridge: Cambridge University Press, 2000.
- [3] ALDERSON, J. C. Testing in EAP: Progress? Achievement? Proficiency? In G. M. BLUE, J. MILTON, and J. SAVILLE eds. Assessing English for Academic Purposes [M]. Switzerland: Peter Lang AG, 2000: 21-48.
- [4] ANON, English Proficiency Test Preparation [EB/OL], 2000. <http://www.apc.edu.au/TestPrep.html>. (Accessed 18th December 2003).
- [5] ANON, A-Z of Methodology [EB/OL], 2000. <http://uk.cambridge.org/elt/ces/methodology/skimmingscanning.htm>. (Accessed 1st January 2004).
- [6] ANON, No date, The Fog Index and Readability Formulas [online]. Available from: <http://www.klariti.com/plain-english/Fog-Index-Readability-Formulas.html>. (Accessed 24th December 2003).
- [7] ARMSTRONG, J. S. Unintelligible Management Research and Academic Prestige[J]. Interfaces, 1980, 10(2): 80-86
- [8] BACHMAN, L. F., DAVIDSON, F., RYAN, K., CHOI, I. C. An Investigation into the Comparability of Two Tests of English as a Foreign Language-The Cambridge-TOEFL Comparability Study [M]. Cambridge: Cambridge University Press, 1995.
- [9] BACHMAN, L. F. and PALMER, A. S. Language Testing in Practice [M]. Oxford: Oxford University Press, 1996.
- [10] CAMBRIDGE, IELTS 3 Examination Papers from the University of Cambridge Local Examinations Syndicate [M]. Cambridge: Cambridge University Press, 2002.
- [11] Carrell, P. L. Some issues in studying the role of schemata, or back-ground knowledge, in second language comprehension [J]. Reading in a Foreign Language, 1983, 1(2): 81-92.
- [12] COHEN, A. D. Assessing Language Ability in the Classroom[M]. 2nd ed. USA: Heinle & Heinle Publishers, 1994.
- [13] CUMMING, A., and BERWICK, R. Validation in Language Testing[M]. Avon: Multilingual Matters, 1995.
- [14] DAVIES, A. Principles of Language Testing[M]. Oxford: Basil Blackwell, 1990.
- [15] EUROPEAN ASSOCIATION FOR INTERNATIONAL EDUCATION, Proficiency in Second Language as a Critical Factor for Success: The Perspective of an international Test Maker, Educational Testing Service

- (ETS) [EB/OL]. 2002. <http://www.eaie.org/pdf/conf2002/706.pdf>. (Accessed 17th December 2003).
- [16] Gough, P. B. One second a reading [C] // E J Kavanagh & C. Mattingly. *Language by Ear and Eye*. Cambridge, MA: MIT Press; 1972;331-358.
- [17] HARRISON, A. *A Language Testing Handbook* [M]. London: The MacMillan Press Limited, 1983.
- [18] HEATON, J. B. *Writing English Language Tests* [M]. New Edition. Harlow: Longman, 1991.
- [19] Heaton, J. B. *Writing English Language Tests* [M]. Beijing: Foreign Language Teaching and Research Press, 2000.
- [20] HENNING, G. *A Guide to Language Testing* [M]. Massachusetts: Newbury House Publishers, 1987.
- [21] Hughes, A. *Testing for Language Teachers* [M]. Cambridge: Cambridge University Press, 1989.
- [22] MICHIGAN ENGLISH LANGUAGE BATTERY, MELAB Technical Manual [M]. Michigan: University of Michigan, 1996.
- [23] MITCHELL, R. *Lecture Notes-Week 6 Assessment of Language Proficiency* [M]. MAALLT, Southampton University, 2003.
- [24] Nuttall, C. *Teaching reading skills in a foreign language* [M]. Shanghai: Shanghai Foreign Language Education Press, 2002.
- [25] Rumelhart, D. E. *Toward an interactive model of reading* [C] // S. Dornic. *Attention and Performance*. New York: Academic Press, 1977, 6: 573-603.
- [26] SANCHEZ, M. *Web-Safe Fonts for your Site* [EB/OL], 2003. <http://www.webpromotionguru.com/articles/s329.shtml> (Accessed 24th December 2003).
- [27] SPENCE-BROWN, R., No date, *The real world and the language tester: considerations of authenticity and interactiveness in the design and assessment of language tests* [EB/OL]. <http://202.245.103.49/ken-shu/Rs/robyn.htm> (Accessed 24th December 2003).
- [28] Smith, F *Understanding Reading* [M]. New York, NY: Holt, Rinehart and Winston, 1971.
- [29] UNIVERSITY OF WARWICK, *Warwick English Language Test* [EB/OL], 2003. <http://www.warwick.ac.uk/fac/soc/CELTE/products/welt/index.shtml>. (Accessed 14th December 2003).
- [30] VALETTE, R. M. *Modern Language Testing* [M]. 2nd Ed. USA: Harcourt Brace Javanovich Inc, 1977.
- [31] Weir, C. J. *Communicative Language Testing* [M]. Prentice Hall International (UK) Ltd., 1990.
- [32] Weir, C. J. & H. Khalifa. *A cognitive processing approach towards defining reading comprehension* [J]. *Cambridge ESOL: RESEARCH NOTES*, 2008, 31: 2-10.
- [33] 任平,但武刚. 六级阅读的雅思化改革对其阅读试题设计的启示[J]. *中国电力教育*, 2009(9):195-196.
- [34] 刘建达. *测试方法对阅读测试的影响* [D]. 广州: 广东外语外贸大学硕士学位论文, 1994.
- [35] 杨惠中,金艳. *大学英语四、六级考试分数解释* [J]. *外语界*, 2001(1).

(责任编辑:夏东,朱德东)

A Comparative Study of Reading Comprehension Tests between IELTS and WELT

FENG Lan

(School of Foreign Language, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: The International English Language Testing System (IELTS) and Warwick English Language Test (WELT) are both criterion-referenced tests which are designed to assess whether a candidate has sufficient linguistic proficiency to study in the medium of English at an undergraduate or postgraduate level. This paper analyzes and compares the reliability and validity of the sections of reading comprehension tests of the two from such perspectives as authenticity, question number and types and so on based on expression forms, difficult coefficients, theme selection and reading comprehension task of the texts.

Key words: IELTS; WELT; reading comprehension test; validity; reliability