



科研项目申报

——阅读 写作 思考 梦想

辜向东

重庆大学语言认知及语言应用研究基地

重庆大学外国语学院英语系

剑桥大学外语考试部

2018年4月南京 昆明 广州



内容提要

- ◆ 整体引介
- ◆ 以团队2014年国家社科基金重点项目书为例
- ◆ 学术与人生的思考



阅读 写作 思考 梦想

To read, write, think and dream

- ◆ 我们来列一下过去的两周我们每个人都做了些什么
- ◆ 看看我们都阅读了什么，读了多少，写了多少，写了什么，思考了多少，思考了什么，梦想了什么？



A "decade of dedication"
is more important than
giftedness in achieving
success in a given field
of learning.

—Benjamin Bloom

国家社科基金项目是哪一年开始设立的？



1991



1991

初为人母
责任与危机



1996

川大硕士
重大助教



2001

上海交大
读博

2006

美国UCLA
访问学者



2011

英国剑桥访学
学术研究顾问

2016

博导 奶奶
我心依旧





国家社科基金项目：从一般项目到重点项目

- ◆ 一般项目：大学英语四、六级考试反拨效应历时研究（**07**BYY030）
- ◆ 青年项目：高考英语试卷内容效度历时研究（1977-2007）（**08**CYY013）
- ◆ 重点项目：基于证据的四六级、雅思、托福考试效度对比研究（**14**AYY010）

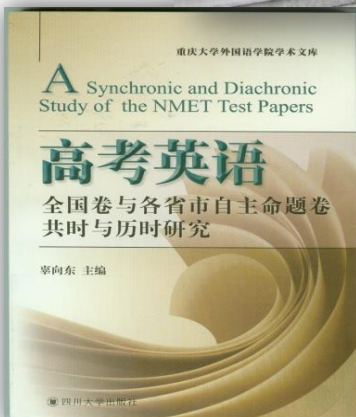


关注国家大事 并联系自己的专长



国家社科基金
重点项目

基于证据的
CET & IELTS & TOFEL
考试效度对比研究



教学、科研与日常生活融为一体
科研选题会无处不在



教学、科研与日常生活融为一体

科研选题会无处不在

- ◆ 书评：从一般期刊到SSCI期刊
- ◆ 综述：从一般期刊到SSCI期刊
- ◆ 实证研究：从一般期刊到CSSCI期刊



书评：从一般期刊到SSCI期刊

1. 辜向东, 2003, 适合英语专业本科生使用的英语语言学教材——《英语语言学导论》[J], 《重庆大学学报》社科版(6): 60-61.
2. 辜向东、向朝红, 2004, 给人以鱼, 一日食鱼; 授人以渔, 终身得鱼——《应用语言学: 研究方法与论文写作书评》[J], 《外国语言文学研究》(2): 84-86.
3. Zheng Y. & Gu X., 2016, Profiling Language Assessment Over the Last Half A Century: A Review of Talking About Language Assessment: The LAQ Interviews[J], *Language Assessment Quarterly* 13(3): 277-282.
4. Gu X. & Yu C., 2017, Review of Second Language Assessment and Mixed Methods Research[J], *Language Testing* 34 (3): 435-438.



综述：从一般期刊到SSCI期刊

1. 辜向东, 2004, 过去十年反拨效应研究综述：问题与趋势[J], 《外国语言文学研究》（4）：69-89。
2. 刘晓华、辜向东, 2013, 国内外反拨效应实证研究二十年回顾[J], 《外语测试与教学》（1）：4-17。
3. 辜向东、郑宇静, 2015, 语言测试的社会属性：《语言评测季刊》十年名家访谈录及启示[J], 《中国外语》（1）：67-74。
4. Zhao H. & Gu X., 2016, China Accreditation Test for Translators and Interpreters (CATTI): Test review based on the language pairing of English and Chinese[J], *Language Testing* 33 (3): 439-446.



实证研究：从一般期刊到CSSCI期刊

1. 辜向东，2001，词缀、大纲、教材、教法[J]，《重庆大学学报》社科版（3）：79-80。
2. 辜向东，2003，走出大学英语阅读技能认识上的误区[J]，《外语界》（4）：60-65。
3. 辜向东，2007，大学英语课堂教学的特征——兼论大学英语四、六级考试对课堂教学的影响[J]，《西安外国语大学学报》（4）：39-45。
4. 辜向东、肖巍，2012，英语专业本科生的进口与出口距《大纲》要求有多远？[J]，《中国外语》（1）：4-9，15。
5. 辜向东、洪岳、杨瑞锦，2017，“大量阅读输入”教学模式下英语专业学生的阅读态度——基于学生学习日志的历时质性研究[J]，《外语与外语教学》（3）：68-77。



学术资源在手边 人是最重要的资源

- ◆ 国际英语语言教育研究基金会官网 www.tirfonline.org

The International Research Foundation for English Language Education

- ◆ 国际语言测试协会官网 <http://www.iltaonline.com/>

International Language Testing Association

- ◆ 现代语言教育期刊官网 *The Modern Language Teaching Journal*

<http://onlinelibrary.wiley.com/doi/10.1111/modl.v100.S1/issuetoc>

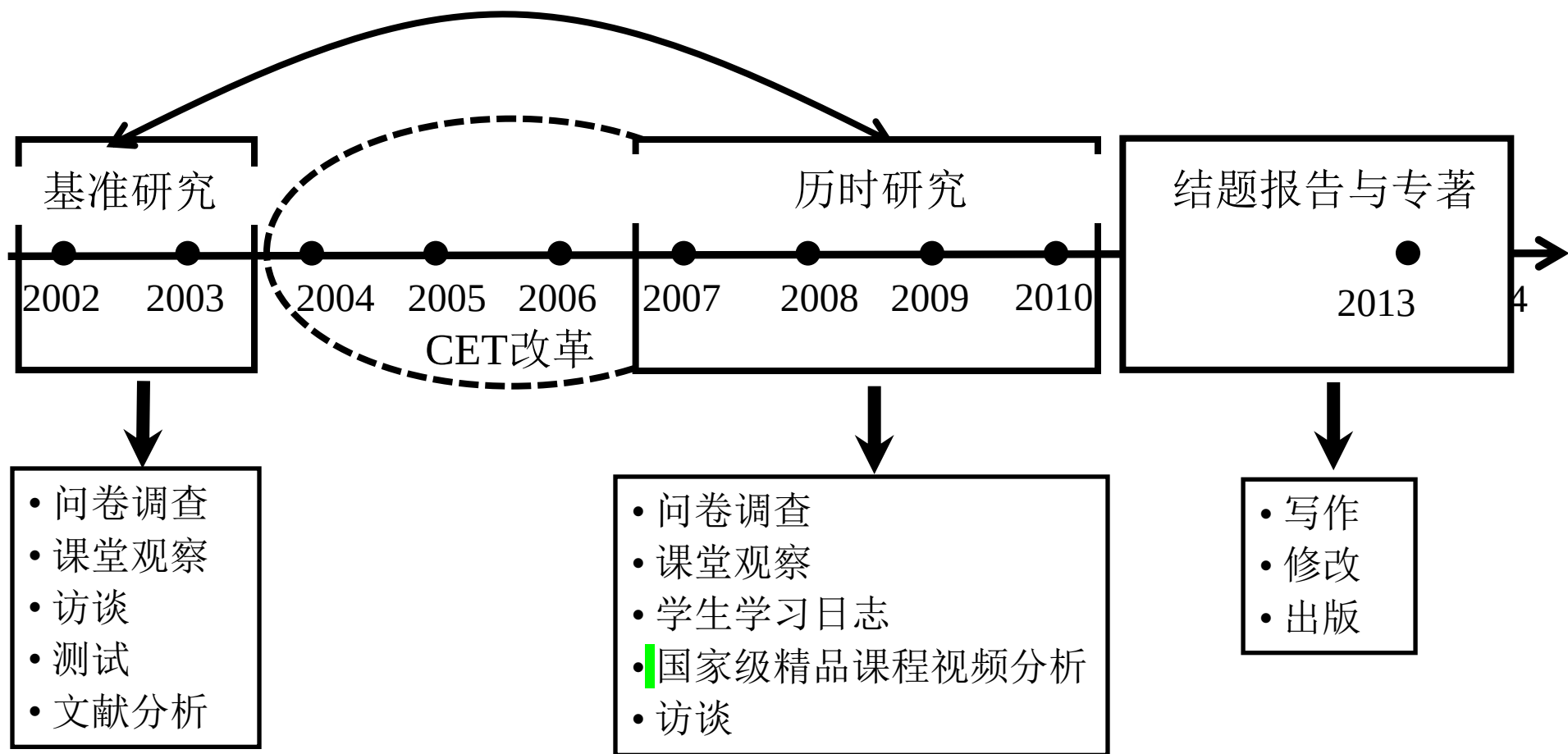
- ◆ Glenn Fulcher教授个人网页

<http://languagetesting.info/features/2016/review.html>



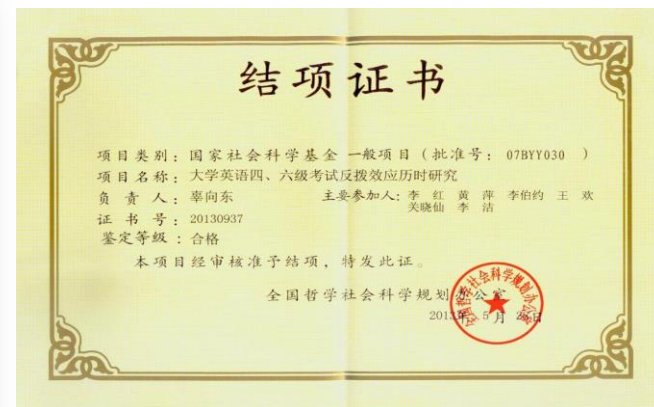
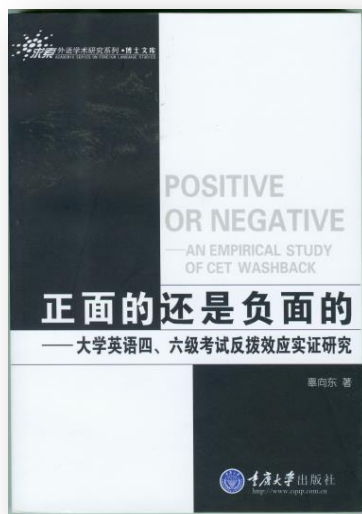
教考结合，捕捉瞬间灵感

项目：国家社会科学基金一般项目《大学英语四、六级考试反拨效应历时研究》，2007.3-2009.12（07BYY030）。





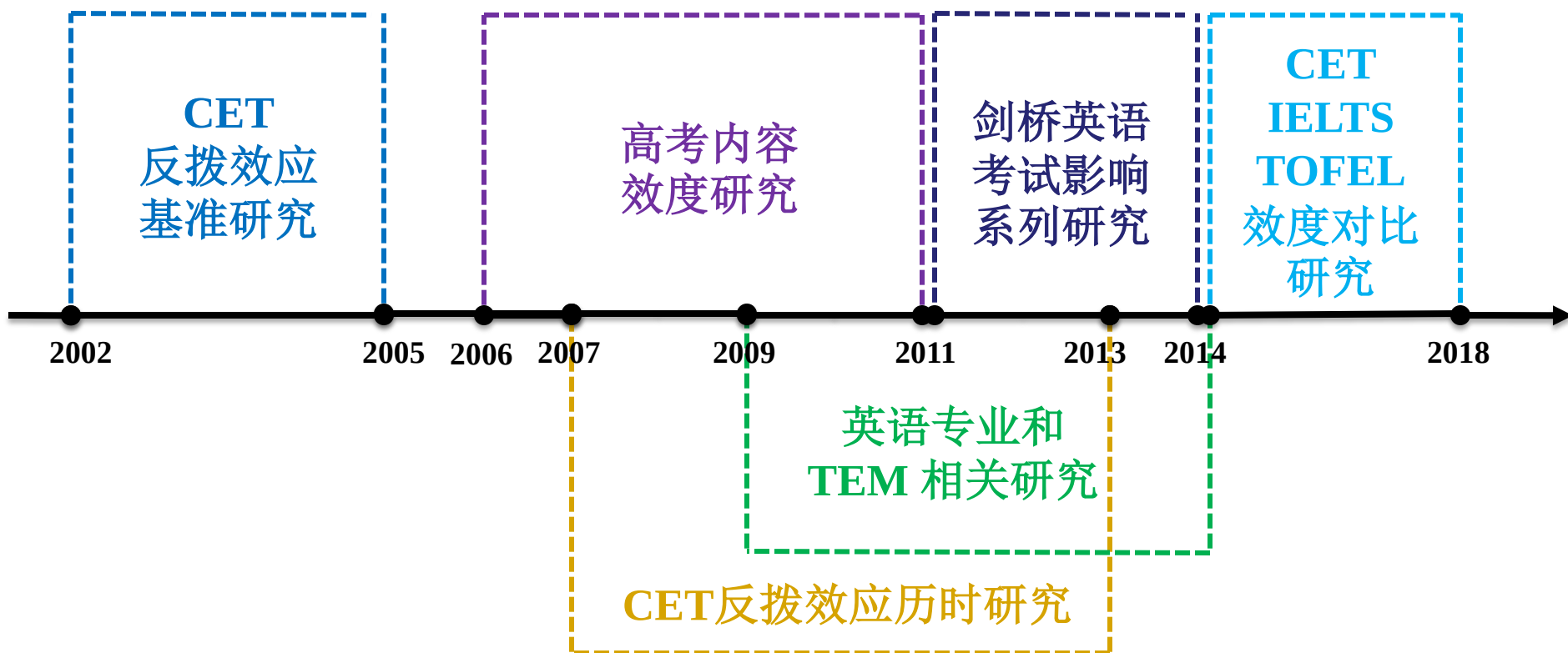
走过十年春夏秋冬（2007-2016）





关注改革与发展，做好前期研究准备

- ◆ 项目：**高考**英语全国卷与各省市自主命题卷共时与历时研究，2006.3-2007.12，重庆大学人文社科青年教师科研启动资助专项（CDSK2006-19）。
- ◆ 辜向东、高晓莹，2007，**高考**英语全国卷与各省市自主命题卷**书面表达**题共时与历时分析（2002-2006）[J]，《外国语言文学研究》（4）：39-47。
- ◆ 辜向东、高晓莹，2007，2007年**高考**英语全国卷与各省市自主命题卷**书面表达**题共时研究[J]，《中国考试》（12）：28-33。
- ◆ 辜向东，2008，《高考英语全国卷与各省市自主命题卷共时与历时研究》[C]，成都：四川大学出版社。
- ◆ 高晓莹、辜向东，2008，**高考**英语自主命题卷**书面表达**题的特色与问题[J]，《中小学外语教学》（3）：18-22；载辜向东（主编）《高考英语全国卷与各省市自主命题卷共时与历时研究》[C]。成都：四川大学出版社，233-242。
- ◆ 项目：**高考**英语试卷内容效度历时研究（1977-2007），2008.6-2010.12，**国家社会科学基金青年项目**（08CYY013）。





基于证据的四六级、雅思、托福考试 效度对比研究

- ◆ 2014.2.12-2.14 第1-6 稿 （个人草拟项目书）
- ◆ 2014.2.15-2.16 第 7 稿 （硕士生团队第一轮修改反馈）
- ◆ 2014.2.17-2.18 第 8-9 稿 （硕士生团队第二轮修改反馈）
- ◆ 2014.2.20-2.21第10-12稿 （课题组和硕士生团队共同修改反馈）
- ◆ 2014.2.22-2.25 定稿 （硕士生团队一起修改定稿）



修改反馈

- ◆ 第一轮（2014.2.15）肖巍 彭莹莹 李亚果 钟瑜
- ◆ 第二轮（2012.16）杨宏波 刘晓华 雷雪梅 赵护林 肖巍
钟瑜 李亚果 杨瑞锦 彭康州 赵路
- ◆ 第三轮（2014.2.17-2.18）董曼霞 颜巧珍 范劲松 Yu Guoxing
徐建 李智
- ◆ 第四轮（2014.2.20-2.21）肖巍 李清华 赵护林 Xi Xiaoming
刘晓华 田洁 申钰文 徐建 颜巧珍 钟瑜 杨瑞锦 赵路
- ◆ 第五轮（2014.2.22）钟瑜 杨宏波 肖巍 刘晓华
- ◆ 第六轮（2014.2.23）田洁 钟瑜 肖巍
- ◆ 第七轮（2014.2.24）颜巧珍



国家社会科学基金项目通讯评审意见表

评价指标	权重	指标说明	专家评分							
选题	3	主要考察选题的学术价值或应用价值，对国内外研究状况的总体把握程度。	10分	9分	8分	7分	6分	5分	4分	3分
论证	5	主要考察研究内容、基本观点、研究思路、研究方法、创新之处。	10分	9分	8分	7分	6分	5分	4分	3分
研究基础	2	主要考察课题负责人的研究积累和成果。	10分	9分	8分	7分	6分	5分	4分	3分
综合评价	是否建议入围		A. 建议入围... B. 不建议入围							
备注										
评审专家（签章）：										

说明：1.本表由通讯评审专家填写，申请人不得填写。项目登记号和项目序号不填。

2.请在“评价指标”对应的“专家评分”栏选择一个分值画圈，不能漏画，也不能多画，权重仅供参考；如建议该课题入围，请在“综合评价”栏 A 上画圈，不建议入围的圈选 B。“备注”栏可简要填写需要说明的其他事项或不填写。本表须评审专家本人签字或盖章有效。



课题名称修改

- ◆ 中英美三项大规模、高风险考试对比研究
- ◆ 基于证据的四、六级、雅思、托福考试效度对比研究
- ◆ 基于证据的四六级、雅思、托福考试效度对比研究
- ◆ 四六级、雅思、托福考试效度对比研究
- ◆ 基于证据的大学英语四、六级与雅思和托福考试的效度对比研究
- ◆ 基于证据的四六级、雅思、托福考试效度对比研究



课题名称第六稿反馈举例（钟瑜 肖巍）

课题名称：基于证据的四、六级、雅思、托福考试效度对比研究。

批注 [U1]: 标题中“四、六级”用顿号感觉有些奇怪。

主文档修订和批注

批注 [U1]

User

顿号很怪！能不能叫大学英语四、六级、雅思、托福？
CET 的全称。

批注 [U2]

User

还是觉得“大学英语四、六级”更正式。雅思和托福是音译，没问题，但 CET 是 college **en** test，意译也没问题。四、六级只是简称。



课题名称第11稿反馈举例（钟瑜 肖巍）

国家社会科学基金项目《课题论证》活页

课题名称：四六级、雅思、托福考试效度对比研究

1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。

批注 [U1]: 个人觉得这个题目更清晰，研究方法不一定要在题目中体现，但研究的主题恰是现在这个题目所涵盖的。

国家社会科学基金项目《课题论证》活页

课题名称：四六级、雅思、托福考试效度对比研究

1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。

批注 [U1]: 题目似乎缺少亮点，是没有专业术语的大白话。会不会让人觉得不够专业？
加个副标题“基于证据的效度验证”会不会好些？



研究现状述评第一稿和第六稿开头

课题名称：中英美三项大规模、高风险考试对比研究

1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。

1.1

中英美三项大规模、高风险考试简介

大学英语四、六级考试 (CET) 是我国教育部高等教育司主办的一项全国性考试，旨在测试在校大学生的英语综合应用能力，为实现大学英语课程教学目标发挥积极作用。人数直线上升，现在每年有上千万在校大学生参加该考试。2005 年至 2013 年，教育部对考试形式和成绩报道方式等一系列改革，本课题研究的 CET 是指此次改革后的考试。考试形式为笔试，考试时间为 120 分钟。考试成绩采用满分为 710 分的计分体制。考试分为两种类型：学术类和普通类。学术类考试主要对考生的英语水平是否满足进行大学或研究生学习的要求。本课题研究的 IELTS 指学术类，考试分为四部分。考试时间总计约 165 分钟。考试成绩采用 1-9 分的等级制。

托福考试 (TOEFL) 是由美国教育考试服务中心 (Education Testing Service, ETS) 主办的国际性英语能力考试。托福考试 1964 年首次在美国

课题名称：基于证据的四、六级、雅思、托福考试效度对比研究

1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。

1.1 本课题国内外研究现状述评

四、六级、雅思、托福考试简介

四、六级 (College English Test Band 4 & Band 6, CET)、**雅思 (International English Language Testing System, IELTS)**、**托福 (Test of English as a Foreign Language, TOEFL)** 是中、英、美三国最具影响的三项大规模、高风险考试。三个国际期刊 *Language Testing*、*Language Assessment Quarterly* 和 *System* 有对这三项考试专门的引介和综述 (Cheng & Zheng, 2008; Alderson, 2008; Zhang & Edler 2009; Chalhoub-Deville & Turner, 2000)。

四、六级 是我国教育部高等教育司主办的一项全国性教育考试，其目的是准确测量我国在校大学生的英语综合应用能力，为大学英语课程教学目标服务。该考试 1987 年实施，目前每年有上千万考生。本课题研究的四、六级是 2013 年最新改革后的四、六级，考试内容包括写作、听力理解、阅读理解和翻译。考试总时间为 130 分钟。成绩采用满分为 710 分的计分体制。

雅思 是由剑桥英语考试部 (Cambridge English Language Assessment)、英国文化委员会及澳洲教育国际开发署联合开发的为打算到英语国家学习、工作或定居的人们设置的英语水平考试。雅思 1989 年创办，目前每年有超过 200 万人次参加该考试。全球超过 170 个国家 8,000 余所院校机构、政府部门和职业机构认可雅思成



研究现状述评第六稿开头反馈举例（钟瑜 肖巍）

课题名称：基于证据的四、六级、雅思、托福考试效度对比研究

1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。

1.1 本课题国内外研究现状述评

• 四、六级、雅思、托福考试简介

四、六级 (College English Test Band 4 & Band 6, CET)、雅思(International English Language Testing System, IELTS)、托福(Test of English as a Foreign Language, TOEFL) 是中、英、美三国最具影响的三项大规模、高风险考试。三个国际期刊 *Language Testing*, *Language Assessment Quarterly* 和 *System* 有对这三项考试专门的引介和综述 (Cheng & Zheng, 2008; Alderson, 2008; Zhang & Edler 2009; Chalhoub-Deville & Turner, 2000)。

四、六级是我国教育部高等教育司主办的一项全国性教育考试，其目的是准确测量我国在校大学生的英语综合应用能力，为大学英语课程教学目标服务。该考试 1987 年开始实施，目前每年有上千万考生。本课题研究的四、六级是 2013 年最新改革后的四、六级，考试内容包括写作、听力理解、阅读理解和翻译。考试总时间为 130 分钟，试题满分为 710 分。

雅思是由剑桥英语考试部、英国文化委员会及澳洲教育国际开发署联合开发的为打算到英语国家学习、工作或定居的人们设置的英语水平考试。雅思 1989 年创办，目前每年有超过 200 万人次参加该考试。全球超过 170 个国家 8,000 余所院校机构、政府部门和职业机构认可雅思成绩。雅思分为学术类和普通类。本课程

批注 [U1]: 标题中“四、六级”用顿号感觉有些奇怪。

删除的内容: 成绩采用

删除的内容: 的计分体制

删除的内容: Cambridge English

国家社会科学基金项目《课题论证》活页

课题名称：基于证据的四、六级、雅思、托福考试效度对比研究

1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。

1.1 本课题国内外研究现状述评

• 四、六级、雅思、托福考试简介

四、六级 (College English Test Band 4 & Band 6, CET)、雅思(International English Language Testing System, IELTS)、托福(Test of English as a Foreign Language, TOEFL) 是中、英、美三国最具影响的三项大规模、高风险考试。三个国际期刊 *Language Testing*, *Language Assessment Quarterly* 和 *System* 有对这三项考试专门的引介和综述 (Cheng & Zheng, 2008; Alderson, 2008; Zhang & Edler 2009; Chalhoub-Deville & Turner, 2000)。

四、六级是我国教育部高等教育司主办的一项全国性教育考试，其目的是准确测量我国在校大学生的英语综合应用能力，为大学英语课程教学目标服务。该考试于 1987 年实施，目前每年考生上千万。本课题研究的四、六级是 2013 年最新改革后的四、六级，其考试内容包括写作、听力理解、阅读理解和翻译，考试总时间为 130 分钟，成绩采用满分 710 分的计分体制。

雅思是由剑桥英语考试部 (Cambridge English Language Assessment)、英国文化委员会及澳洲教育国际开发署

批注 [U1]: 顿号很怪！能不能叫大学英语四、六级、雅思、托福？CET的

批注 [U2]: 还是觉得“大学英语四、六

批注 [U3]: 可注明“以下简称四、六

批注 [U4]: 最具？理由？如果非要边

批注 [U5]: 三份业界顶级国际 SSCT

删除的内容: 有上千万

删除的内容: 。

删除的内容: 计

删除的内容: 。

删除的内容: 为

批注 [U6]: 有意

批注 [U7]: 不搭界。Cet 和托福都

删除的内容: 有超过 200 万人次参加

删除的内容: 全球超过 170 个国家

删除的内容: 指

删除的内容: 考试分为

删除的内容: 四部分。



研究现状述评第11稿开头反馈举例（刘晓华 颜巧珍）

国家社会科学基金项目《课题论证》活页

课题名称：四六级、雅思、托福考试效度对比研究

1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。

1.1 本课题国内外研究现状述评

• 效度定义及理论演进

效度 (validity) 是测试评价中最重要的考虑因素 (APA/AERA/NCME 1985)。较早的效度定义：“指一项测试是否测量了它所要测量的东西” (Kelly 1927: 14)。自 20 世纪 60 年代以来，语言测试与评价的研究一直围绕效度展开 (Kunnan 1998)，效度理论也从“分类效度观” (APA/AERA/NCME 1966; Lado 1961) 发展到“整体效度观” (Messick 1989, 1996)。分类效度观认为效度可分为校标关联效度、内容效度、构念效度等多种类型 (APA/AERA/NCME 1974)，其验证操作性强，但比较零散，且不够全面，没有考虑收集分数意义的价值和分数使用的后果等方面的证据。上世纪 80 年代末，Messick (1988, 1989) 提出整体效度的概念，给出了具有突破意义的效度定义，“经验证据和理论依据在多大程度上支持分数的解释与使用，对这个问题进行的综合评价

批注 [XL1]: 整个课题的论证给我的感觉还是停留在根据 Bachman 的理论对三个考试的效度分别同时进行研究，而没有说怎么进行比较。

带格式的：缩进：首行缩进：2 字符

删除的内容：是 Kelly (1927: 14) 提出的，“效度问题

删除的内容：取得了重大发展 (李清华 2006; 韩宝成、罗凯洲 2013)，发生了

删除的内容：的演进

删除的内容：，效度包括很多种类，如：

删除的内容：等。

删除的内容：分类效度观使效度的

删除的内容：易于

删除的内容：。

删除的内容：确定了构念的核心地位

课题名称：基于证据的效度验证理论视角的大学英语四、六级，雅思，托福考试对比研究

1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。

1.1 本课题国内外研究现状述评

• 效度理论的演进

效度 (validity) 是测试评价中最重要的考虑因素 (APA/AERA/NCME 1985)，其定义由 Kelly (1927: 14) 较早提出，“效度问题指一项测试是否测量了它所要测量的东西”。自 20 世纪 60 年代以来，语言测试与评价的研究一直围绕效度展开 (Kunnan 1998)，效度理论取得了重大发展 (李清华 2006; 韩宝成、罗凯洲 2013)，发生了从“分类效度观” (APA/AERA/NCME 1966; Lado 1961) 到“整体效度观” (Messick 1989, 1996) 的演进。分类效度观认为效度包括很多种类，如：校标关联效度、内容效度、构念效度 (APA/AERA/NCME 1974) 等。这一效度观使效度验证易于操作，但不够全面，没有考虑收集分数意义的价值和分数使用的后果等方面的证据。Messick (1988, 1989) 提出整体效度观，“经验证据和理论依据在多大程度上支持分数的解释与使用，对这个问题进行的综合评价就是效度”。这种“一元多维”的效度整体观明确了效验的对象不仅包括测试本身，还应包括测试结果的解释和使用 (Davies 2003)，把效度涵盖的范围由分数意义扩展到相关性、使用、价值和后果 (Shepard 1993)，同时确定了构念的核心地位 (Alderson & Banerjee 2001) (调整至句末，整句和前文的逻辑性更强。但前面未提及构念，这里提出是否合适?)。

删除的内容：(

删除的内容：)

删除的内容：。效度的

批注 [S1]: 此句不通

删除的内容：分类

删除的内容：的

删除的内容：比较零散，且

删除的内容：。

删除的内容：上世纪 80 年代末，

删除的内容：的概念

删除的内容：给出了具有突破性的效度定义，

删除的内容：而是

删除的内容：确定了构念的核心地位 (Alderson & Banerjee 2001)，

删除的内容：的

删除的内容：、复杂

删除的内容：，

删除的内容：。



研究现状述评第13稿开头反馈举例（刘晓华 肖巍）

课题名称：基于证据的四六级、雅思、托福考试效度对比研究
1. 本课题国内外研究现状述评，选题的价值和意义。2. 本课题研究的主要内容、基本观点、研究思路、研究方法、创新之处。3. 前期相关研究成果，开展本课题研究的主要参考文献。限 4000 字以内。
1.1 本课题国内外研究现状述评
• 效度理论
效度(validity)是测试评价中最重要的考虑因素(APA/AERA/NCME1985)。较早的效度定义“指一项测试是否测量了它所测量的东西”(Kelly 1927: 14)。自 20 世纪 60 年代以来，语言测试与评价的研究一直围绕效度展开(Kunnan1998)，效度理论取得了从“分类效度观”到“整体效度观”的重大发展。分类效度观认为效度可分为校标关联效度、内容效度、构念效度等多种类型(APA/AERA/NCME 1966, 1974; Lado 1961)。其验证操作性强，但比较零散，且未考虑收集分数使用等方面的证据。整体效度观给出了具有突破意义的效度定义，即“经验证据和理论依据在多大程度上支持分数的解释与使用，对这个问题进行的综合评价就是效度”(Messick, 1988, 1989, 1996)。这种“一元多维”的效度观确定了构念的核心地位(Alderson & Banerjee 2001)，明确了效度验证的对象是测试结果的解释和使用(Davies 2003)。
• 效度验证模式
Messick 的整体效度观对语言测试的开发与研究产生了深远影响，但该理论高度概括、过于抽象，使得效度验证缺乏可操作性。近年来语言测试学家根据该理论提出了一些更具体的效度验证框架：(1)“交际语言能力”模型和“测试方法层面”框架(Bachman1990)为效度验证开启了新视角。Bachman et al. (1995)

删节的内容：(APA/AERA/NCME 1966; Lado 1961)

删节的内容：(Messick, 1988, 1989, 1996)

批注 [XL1]: 跟前面的重复了占字数

删节的内容：(APA/AERA/NCME 1966, 1974)

说明：1. 本表由通讯评审专家填写，申请人不得填写。项目登记号和项目序号不填。

2. 请在“评价指标”对应的“专家评分”栏选择一个分值画圈，不能漏画，也不能多画，权重仅供参考；如建议该课题入围，请在“综合评价”栏 A 上画圈，不建议入围的不做标记。“备注”栏可简要填写需要说明的其他事项或不填写。本表须评审专家本人签字或盖章有效。

国家社会科学基金项目《课题论证》活页

课题名称：基于证据的四六级、雅思、托福考试效度对比研究

1. 本课题国内外研究现状述评，选题的价值和意义

1.1 本课题国内外研究现状述评

• 效度理论

效度(validity)是测试评价中最重要的考虑因素(APA/AERA/NCME 1985)。较早的效度定义指“一项测试是否测量了它所测量的东西”(1927: 14)。自 20 世纪 60 年代以来，效度理论取得了从“分类效度观”(APA/AERA/NCME 1966; Lado 1961)到“整体效度观”(Messick 1988, 1989, 1996)的重大发展。分类效度观认为效度可分为校标关联效度、内容效度、构念效度等多种类型(APA/AERA/NCME 1966, 1974)。其验证操作性强，但比较零散，且未考虑收集分数使用等方面的证据。整体效度观给出了具有突破意义的效度定义，即“经验证据和理论依据在多大程度上支持分数的解释与使用，对这个问题进行的综合评价就是效度”。这种“一元多维”的效度观确定了构念的核心地位(Alderson & Banerjee 2001)，明确了效度验证的对象是测试结果的解释和使用(Davies 2003)。

• 效度验证模式

整体效度观对语言测试的开发与研究产生了深远影响，但该理论高度概括、过于抽象，使得效度验证缺乏可操作性。近年来，语言测试学家根据该理论提出了一些更具体的效度验证框架：(1)“交际语言能力”模型和“测试方法层面”框架(Bachman 1990)。Bachman et al. (1995)运用该框架，对 CPE、FCE、TOEFL

删节的内容：指

批注 [U1]: 太过绝对吧？可以删去，接着说效度理论的发展即可。

删节的内容：语言测试与评价的研究一直围绕效度展开(Kunnan 1998)，

批注 [U2]: 上一段并未强调 messick，只是夹注了一下。

从“分类效度观”(APA/AERA/NCME 1966; Lado 1961)到“整体效度观”(Messick 1988, 1989, 1996)。故，此段开头应该接着谈整体效度观而非 messick。

删节的内容：Messick 的

批注 [U3]: 可删，或改为“发展了该理论”/“基于该理论”。

批注 [U4]: 先列举框架名称，再介绍。原文一上来就介绍，不协调。加上“该框架”，也设几个字，但语句通畅很多。

批注 [U5]: 应该简介 CLA 和 test method facet 的组成部分，与下面几个框架体例一致。

删节的内容：为



研究现状述评定稿

1.1 本课题国内外研究现状述评

· 效度理论

效度 (validity) 是测试评价中最重要的考虑因素 (APA/AERA/NCME 1985)。较早的效度定义为“一项测试是否测量了它所要测量的东西” (Kelly 1927:14)。自 20 世纪 60 年代以来, 语言测试与评价的研究一直围绕效度展开 (Kunnan 1998), 效度理论取得了从“**分类效度观**”到“**整体效度观**”的重大发展。**分类效度观** (Lado 1961) 认为效度可分为效标关联效度、内容效度、**构念效度**等多种类型 (APA/AERA/NCME 1966, 1974)。其验证操作性强, 但比较零散, 且未考虑收集分数使用等方面的证据。**整体效度观**给出了具有突破意义的效度定义, 即“对经验证据和理论依据在多大程度上支持分数的解释与使用进行的综合评价就是效度” (Messick 1989:13)。这种“一元多维”的效度观确定了**构念**的核心地位 (Alderson & Banerjee 2001), 明确了效度验证的对象是测试结果的解释和使用 (Davies 2003)。

· 效度验证模式

整体效度观给语言测试的开发与研究带来了重大变革, 但该理论高度概括、过于抽象, 使效度验证缺乏可操作性。近年来更多的语言测试学家根据该理论提出了一些具体的效度验证框架: (1) “**交际语言能力模型**”和“**测试方法层面框架**” (Bachman 1990) 为效度验证开启了新视角, Bachman et al. (1995) 运用该框架对 CPE、FCE、TOEFL 三项考试所考查的能力和测试任务特征做了分析, 并对框架进行了完善。(2) “**测试有用性框架**” (Bachman & Palmer 1996) 涵盖信度、**构念效度**、真实性、交互性、考试影响和可行性六个质量属性, 进一步阐释了 Messick 的效度理论。该框架可操作性强, 但质量属性之间的关联不甚明确 (韩宝成、罗凯洲 2013)。(3) “**基于论证的效验模式**” (Kane 1992) 与整体效度观一脉相承, 包括两个步骤: 提出效验观点、收集有关证据。Chapelle et al. (2008) 运用该模式论证了 TOEFL iBT 的效度。(4) “**测试使用论证框架**” (Bachman 2003) 发展了 Kane 的效度论证观。该框架遵循“事实→主张”的推理机制, 包含构建与评价两个过程 (Bachman & Palmer 2010)。不过其架构 (后果、决策、解释、测试记录) 比较抽象, 能否成为指导测试开发与使用的新范式有待检验。(5) “**基于证据的效验框架**” (Weir 2005) 从社会认知视角出发, 涵盖五个方面的效验证据: 基于理论的效度、环境效度、评分效度、效标关联效度和后果效度, 可操作性较强, 并在剑桥主体证书考



选题的价值和意义第9稿反馈（董曼霞）

1.2 选题的价值和意义

• **选题的学科理论与实践价值：**理论上，本研究将验证基于证据的效度验证理论框架在效度实证研究中的可行性，并进一步构建语言测试效度对比研究的模型。实践上，通过对比三项考试的效度，丰富和发展考试对比研究领域的文献，为类似研究提供借鉴。

• **选题的社会和现实意义：**雅思考试的分数等级已与《欧洲语言共同参考框架：学习、教学、评估》（欧洲理事会，2001）中相应的语言能力等级实行了匹配，雅思与托福的分数也等值化。本研究将参照国际公认的语言能力等级进行四、六级考试的分数等值化，为推动我国统一的语言能力等级共同量表的制定提供参考依据，也为教育、人事部门及广大的考试利益相关者提供入学、就业、交流等决策依据。

2.1 本课题研究的主要内容

效度验证理论包括受试特征、环境效度、认知效度、评分效度、后果效度和效标关联效度（Weir, 2005; Khalifa & Weir, 2009）。本课题拟从该理论出发，结合语言测试使用论证的要素和主张（Bachman, 2005; Bachman & Palmer, 2010），对四、六级，雅思，托福三项考试进行全面深入的效度对比研究。具体内容和研究问题如下：

- 1)→受试特征：三项考试涉及受试的哪些生理、心理和体验特征？
- 2)→环境效度：三项考试测试任务的环境和操作对所有受试是否公平？
- 3)→认知效度：完成三项考试测试任务的认知过程和交互活动是否真实？
- 4)→评分效度：三项考试的分数在多大程度上是可靠的？

批注 [LU19]: 这点有些迷惑，好像与本研究不是很相关，是不是表述的原因？作为一个读者，我很希望在这一部分清楚知道进行三项测试效度对比究竟有何实际意义。读后感觉本课题效度对比研究的现实意义就是参考雅思与托福的分数的等值化，四六级考试也应根据《欧框》实现分数等值化。研究实际意义应该不仅局限于此，可考虑将实际意义适度放宽。选题的意义对读者很重要，我想很多读者都会关注这一点的，评审专家应该也是这样。所以能否把这一块再好好理一下。





选题的价值和意义第11稿反馈（杨瑞锦 赵 路）

1.2 选题的价值和意义

Graddol (2006) 指出, 英语已经成为了全世界的语言。在我国, 英语教学是外语教育中规模最大的部分。由于教学层次复杂、考试种类繁多, 相应的英语教学与考试大纲对语言能力的描述和等级划分存在分歧, 给教学的组织、考试分数的解释和考试结果的使用等带来诸多不便 (方绪军等 2008)。又由于考试对个人、机构乃至整个社会都可能产生重要影响, 考试本身的质量如何以及考试使用是否得当也成为全社会关注的焦点。近年来, 关于制定全国统一的语言能力等级量表 (韩宝成 2006; 杨惠中等 2012) 和语言测试标准 (范劲松、金艳 2010) 的呼声越来越高。而两者的制定都需要相应的理论建构与实践探索。这正是本选题的价值和意义。

· **学科理论与实践价值:** 理论上, 构建语言测试效度对比研究模型。实践上, 丰富和发展考试对比研究领域的文献, 为类似研究提供思路和方法上借鉴。

批注 [54]: 本部分已经分条陈述了选题的价值和意义, 前面的一大段陈述显得有些多余, 可以直接切掉, 留下分条的表述即可, 如果感觉不妥可以将分条内容适当扩充。

带格式的: 字体颜色: 红色

1.2 选题的价值和意义

Graddol (2006) 指出, 英语已经成为了全世界的语言。在我国, 英语教学占据外语教育中最大的部分。由于教学层次复杂、考试种类繁多, 相应的英语教学与考试大纲对语言能力的描述和等级划分存在分歧, 给教学的组织、考试分数的解释和考试结果的使用等带来诸多不便 (方绪军等 2008)。又由于考试对个人、机构乃至整个社会都可能产生重要影响, 考试本身的质量如何以及考试使用是否得当也成为全社会关注的焦点。近年来, 关于制定全国统一的语言能力等级量表 (韩宝成 2006; 杨惠中等 2012) 和语言测试标准 (范劲松、金艳 2010) 的呼声越来越高。而两者的制定都需要相应的理论建构与实践探索。这正是本选题的价值和意义。

· **学科理论与实践价值:** 理论上, 构建语言测试效度对比研究模型。实践上, 丰富和发展考试对比研究领域的文献, 为类似研究提供思路和方法上借鉴。

· **社会和现实意义:** 本研究有助于推动我国语言测试开发与研究的国际化, 为我国制定统一的语言能力等级量表和适合我国国情的语言测试标准提供实证参考数据, 为教育、人事部门及广大利益相关者提供入学、就

国内外研究现状评述

删除的内容: 为弥补现有有效度研究中存在的不足, 本课题拟以 Bachman & Palmer (2010) 的“测试使用论证”框架为出发点, 对四六级、雅思、托福三项考试进行全面系统的效度对比研究。

删除的内容: 是

删除的内容: 规模

批注 [U10]: 语言能力等级量表应该涵盖初级到高级, 而本研究只涵盖了我国较高层次的 CET 考试做研究, 所以这样说是不是有点大了或者说略显牵强了些。



选题的价值和意义第11稿反馈（刘晓华 肖巍）

1.2 选题的价值和意义

Graddol (2006) 指出, 英语已成为世界性的语言。在我国, 英语教学是外语教育中规模最大的部分。由于教学层次复杂、考试种类繁多, 相应的英语教学与考试大纲对语言能力的描述和等级划分存在分歧, 给教学的组织、考试分数的解释和考试结果的使用等带来诸多不便 (方绪军等 2008)。又由于考试对个人、机构乃至整个社会都可能产生重大影响, 考试本身的质量如何以及考试使用是否得当也成为全社会关注的焦点。近年

删除的内容: 经... 为了全... 界性的语...

来, 关于制定全国统一的语言能力等级量表 (韩宝成 2006; 杨惠中等 2012) 和语言测试标准 (范劲松、金艳 2010) 的呼声越来越高。而两者的制定都需要相应的理论建构与实践探索。这正是本选题的价值和意义。

批注 [XL3]: 我仍然不太理解考试对比和制定统一的语言能力等级量表之间的关系。考试对比可能对语言测试标准的制定有一定启示。

• 学科理论与实践价值: 理论上, 构建语言测试效度对比研究模型。实践上, 丰富和发展考试对比研究领域的文献, 为类似研究提供思路和方法上借鉴。

1.2 选题的价值和意义

英语是国际上最通用的语言。在我国, 英语教育在外语教育中规模最大。由于教学层次复杂、考试种类繁多, 语言能力的描述和等级划分标准不一, 给教学的组织、考试分数的解释和考试结果的使用等带来诸多不便 (方绪军等 2008)。此外, 由于考试对个人、机构乃至整个社会都可能产生重要影响, 考试本身的质量如何以及考试使用是否得当也成为全社会关注的焦点。近年来, 关于制定全国统一的语言能力等级量表 (韩宝成 2006; 杨惠中等 2012) 和语言测试标准 (范劲松、金艳 2010) 的呼声越来越高。然而两者的制定都需要相应的理论建构与实践探索, 这正是本选题的价值和意义所在。

• 学科理论与实践价值: 理论上, 构建语言测试效度对比研究模型。实践上, 丰富和发展考试对比研究领域的文献, 为类似研究提供思路和方法上借鉴。

• 社会和现实意义: 本研究有助于推动我国语言测试开发与研究的国际化, 为制定我国统一的英语语言能力等级量表和适合我国国情的语言测试标准提供实证参考数据, 为教育、人事部门及广大利益相关者提供入学、就业、流动等决策依据, 同时有望提升我国自行开发的英语考试在国际的认可度。

2.1 本课题研究的主要内容

极具影响力的是 Weir (2005) 提出的基于证据 (evidence-based) 的效度验证理论框架。该框架包括受试特征、环境效度、认知效度、评分效度、后果效度和效标关联效度 (Weir 2005; Khalifa & Weir 2009)。本课题拟从该理论框架出发, 对四六级、雅思、托福三项考试进行全面深入的效度对比研究。具体内容和研究问题如下:

1) 受试特征: 三项考试涉及受试的哪些生理、心理和体验特征?

对比研究

删除的内容:

批注 [U8]: 不是术语, 翻译用意何在?

删除的内容: (coherent and forceful validity arguments)

批注 [U9]: 没改过来?

批注 [U10]: 不用指出也知道吧?

删除的内容: Graddol

删除的内容: (2006) 指出,

删除的内容: 已经成为了全世界的语言

删除的内容: 学

删除的内容: 相应的英语教学与考试大纲在

删除的内容: 上存在分歧

删除的内容: ,



选题的价值和意义定稿

1.2 选题的价值和意义

• 学科理论与实践价值

理论上，验证基于证据的效度验证框架在考试效度对比研究中的可行性，并进一步构建更加科学合理的语言测试效度对比研究模型。**实践上**，通过对比三项考试的效度，形成将三项考试关联起来的论证。这不仅可以丰富考试对比研究领域的文献，而且能为类似的研究提供思路和方法上的借鉴。

• 社会和现实意义

一方面，本研究有助于推动我国语言测试开发与研究的国际化，有望提升我国自行开发的英语考试在国际上的认可度，为教育、人事部门及广大利益相关者提供入学、就业、流动等决策依据；**另一方面**，由于我国英语教学层次复杂、考试种类繁多，近年来关于制定我国统一的语言能力等级量表（韩宝成 2006；杨惠中等 2012）和语言测试标准（范劲松、金艳 2010）的呼声越来越高，而本研究中的雅思和托福均已实现与国际公认的语言能力标准（ACTFL1986；CEFR 2001）对接，其开发与使用也遵循了国际公认的语言测试标准（ETS 2002；UCLES 2013），因此三项考试的效度对比研究有望为制定我国统一的语言能力等级量表和语言测试标准提供参考数据。



研究的主要内容第9稿国内同行反馈（范劲松）

2) 三项语言考试的简介固然重要，但似乎放在开头不太合适。我觉得在第一部分向后面放一下，稍微介绍即可（因为这三项语言考试都广为人知，省下的文字可以多一些研究背景的介绍）。开头宜稍微说明一下研究背景（如近些年基于论据argument和证据evidence的效度理论的发展、新的效度研究方法的进展等等）。从一开始能够抓住reviewer的眼睛似乎是最好的策略。

3) 我个人觉得Cyril Weir的框架固然不错，也就是所谓的socio-psychological framework of test validation，可以选择作为本研究的guiding theoretical framework，但是是否同时也可以把本研究放置在更大的研究背景下？那就是argument-based validation (Kane, 1992, 2001; 也包括Bachman, 2005; Bachman & Palmer, 2010)。我个人觉得，目前这三项语言考试尽管有很多的validaiton studies，但是比较缺乏的是将这些studies组织起来形成coherent and forceful validity arguments。Chapelle等人2008年的著作是TOEFL考试采用argument-based validation的代表。但是据我所知，其他两项考试似乎都没有类似的研究。如果放置到这些背景下，再选择Weir的框架作为理论框架，是不是会更厚实一些？



研究的主要内容第9稿国外同行反馈（Xiaoming Xi）

Hi Xiangdong,

This sounds like very interesting research, and really ambitious!

I would be very interested in participating in this research.

I have one suggestion though: I think Weir's social-cognitive framework is not as influential as Bachman's or Chapelle et al's adaptations of Kane's work. I've attached a draft chapter for a book volume where I compare these different frameworks. I would suggest that you consider using Bachman's AUA or Chapelle et al's validity argument framework. I'm also sending you another chapter I recently published that might be helpful.

Best,

Xiaoming



研究的主要内容第11稿学生反馈（申钰文 钟瑜）

2.1 本课题研究的主要内容 本课题拟以 Bachman & Palmer (2010) 的 AUA 框架为基础，对四六级、雅思、托福三项考试进行全面深入的效度对比研究。具体研究问题如下： 1) 测试的使用和依据测试结果所作的决定对利益相关群体产生的影响是否是正面的？ 2) 基于测试所作的决定是否尊重群体的价值观，符合相关的法规，且对所有的利益相关者公平？ 3) 对于所测的语言能力的诠释是否是有意、公正、可推广、相关且充分？ 4) 对于不同的测试任务、测试环境以及考试群体，考生表现的记录与考生能力的描述是否一致？	批注 [微软用户3]: 研究的内容和问题是否能体现课题的价值和意义，是否能达到预期的目标？研究的主要问题是本申请书的核心，也是亮点所在，也是审核整个课题的可行性的重要依据。
2.2 基本观点 本课题的基本观点是：三项考试都是以英语为外语或二语的大规模、高风险语言考试，受试将接受或正在接受高等教育，其效度应该既有较大的相似性，也存在一定的差异。而实际情况如何，需要进行全面深入的效度对比研究。	批注 [微软用户4]: 这个也比较重要，但这里陈述是否过于简单，是否可以对研究有个大胆的预测。
2.3 研究思路和方法 本课题拟以Bachman & Palmer (2010)的AUA框架为出发点，结合四六级、雅思、托福考试实际，运用质性和量化相结合的研究方法，对比这三项考试的效度。具体研究思路和方法如下：	批注 [微软用户5]: 这种罗列方法有些分数，其实这几种方法之间的内在逻辑性是很强的，怎么将其中的内在逻辑性呈现出来也是要考虑的。

2.1 本课题研究的主要内容 本课题拟以 Bachman & Palmer (2010) 的 AUA 框架为基础，对四六级、雅思、托福三项考试进行全面深入的效度对比研究。具体研究问题如下： 1) 测试的使用和依据测试结果所作的决定对利益相关群体产生的影响是否是正面的？ 2) 基于测试所作的决定是否尊重群体的价值观，符合相关的法规，且对所有的利益相关者公平？ 3) 对于所测的语言能力的诠释是否是有意、公正、可推广、相关且充分？ 4) 对于不同的测试任务、测试环境以及考试群体，考生表现的记录与考生能力的描述是否一致？	批注 [U12]: 之前的版本，不是用的“基于证据的 (evidence-based) 效度验证模式”吗？怎么换了？如果有更换，前面在阐述不同的验证模型的时候，应该重点突出本研究要用的这个模型的内容和优势，简述其他模型。 批注 [U13]: 研究问题一定紧扣项目选题，这几个问题是否涵盖三大考试效度对比研究最主要的内容？
2.2 基本观点 本课题的基本观点是：三项考试都是以英语为外语或二语的大规模、高风险语言考试，受试将接受或正在接受高等教育，其效度应该既有较大的相似性，存在一定的差异。而实际情况如何，需要进行全面深入的效度对比研究。	批注 [U14]: 基本观点是否应该与研究问题相对应？是否应该是对研究问题的答案的猜想？ 删除的内容: 也



研究的主要内容第11稿学生反馈（肖巍 刘晓华）

2.1 本课题研究的主要内容 本课题拟以 Bachman & Palmer (2010) 的 AUA 框架为基础，对四六级、雅思、托福三项考试进行全面深入的效度对比研究。具体研究问题如下： 1) 测试的使用和依据测试结果所作的决定对利益相关群体产生的影响是否是正面的？ 2) 基于测试所作的决定是否尊重群体的价值观，符合相关的法规，且对所有的利益相关者公平？ 3) 对于所测的语言能力的诠释是否是有意、公正、可推广、相关且充分？ 4) 对于不同的测试任务、测试环境以及考试群体，考生表现的记录与考生能力的描述是否一致？ 2.2 基本观点 本课题的基本观点是：三项考试都是以英语为外语或二语的大规模、高风险语言考试，受试将接受或正在接受高等教育，其效度应该既有较大的相似性，也存在一定的差异。而实际情况如何，需要进行全面深入的效度对比研究。 2.3 研究思路和方法 本课题拟以 Bachman & Palmer (2010) 的 AUA 框架为出发点，结合四六级、雅思、托福考试实际，运用质性和量化相结合的研究方法，对比这三项考试的效度。具体研究思路和方法如下：	批注 [U11]: 我对这个框架不太熟。觉得这个框架的思维太超前了，更多涉及到社会、伦理而不是考试和考试内容本身。没有 weir 那个好懂。Reviewer 如果缺乏背景，看到这里，一定会想：这些东西跟“效度”（传统意义上的）有什么联系？如果是 weir 那个，因为每一种效度都带上了“效度”两个字，可能让人容易接受一些。 批注 [U12]: 若不改，就把与 aua 框架相关的方面用黑体突出。不然，不熟悉的人实在难以看出理论与问题、问题与方法之间的关联。 批注 [U13]: 正如上面所说，weir 那个，你可以把每种方法与相应的效度对应起来。这个你怎么对应？ 我个人比较喜欢 weir 那个理论，更实在，对考试本身关注更多（虽然也考虑考后效度）。这正如语言学与边缘分支的关系一样。虽然语言具有认知、社会、心理等多重维度，但语言本体才是核心。对于一项考试，当然需要考虑到后果，但考试本身才是关键。
---	--

2.1 本课题研究的主要内容 本课题拟以 Bachman & Palmer (2010) 的 AUA 框架为基础，对四六级、雅思、托福三项考试进行全面深入的效度对比研究。具体研究问题如下： 1) 三项考试的使用和依据测试结果所作的决定对利益相关群体产生的影响是否是正面的？ 2) 基于三项考试所作的决定是否尊重群体的价值观，符合相关的法规，且对所有的利益相关者公平？ 3) 对三项考试所测的语言能力的诠释是否是有意、公正、可推广、相关且充分？ 4) 三项考试在不同的测试任务、测试环境以及考试群体的情况下，考生表现与考生能力的描述是否一致？	删除的内容: 测试 删除的内容: 测试 删除的内容: 对于 删除的内容: 对于 删除的内容: 的记录 批注 [XL4]: 研究问题主要在谈效度论证，而忽略了对比
---	--



研究的主要内容定稿

2.1 本课题研究的主要内容

本课题拟从 Weir (2005) “**基于证据的效度验证框架**”出发，从六个方面对四六级、雅思、托福进行较全面深入的考试效度对比研究。具体内容和研究问题如下：

- 1) **受试特征**：三项考试涉及受试的哪些生理、心理和体验特征？
- 2) **环境效度**：三项考试测试任务的环境和操作对所有受试是否公平？
- 3) **认知效度**：受试完成三项考试测试任务的认知过程和交互活动是否真实？
- 4) **评分效度**：三项考试的评分及考试分数在多大程度上是可靠的？
- 5) **后果效度**：三项考试对受试的心理状态和学习过程产生了什么影响？
- 6) **效标关联效度**：三项考试的分数是否一致性较高？是否可以进行等值？

这些具体内容和研究问题最终指向并回答一个总的问题：**三项考试的效度有何异同？**



基本观点第6稿反馈（肖巍）

2.2 基本观点

本课题的基本观点是：四、六级考试的目的在于准确地衡量我国在校大学生的英语综合应用能力。雅思考试的目的是测试为进入英联邦国家接受高等教育的非英语国家学生的英语水平和能力。托福考试的目的是衡量考生在学术环境中运用四项英语技能的综合能力。三项考试都是以英语为外语或二语的的大规模、高风险语言考试，考生均将接受或正在接受高等教育，三者效度应该既有较大的相似，也存在一定差异。而实际情况是否如此，需要进行全面深入的对比研究。

2.3 研究思路和方法

本课题拟以Weir（2005）的基于证据的效度验证理论框架为基础，结合Bachman（2005）的测试使用论证的要素和主张，运用定性和定量的研究方法对比三项考试的效度，具体思路和方法如下：

删除的内容：测试语

删除的内容：而且受试都是以英语为外语或第二语言，并

删除的内容：或处于高等教育层次

删除的内容：，其

删除的内容：之处

删除的内容：明显的

删除的内容：对其效度



基本观点定稿

2.2 基本观点

尽管四六级、雅思、托福这三项考试的目的、性质、构念、分数解释和结果使用等诸多方面存在不同，但三项考试都是以英语为外语或二语的大规模、高风险语言考试，受试将接受或正在接受高等教育，三项考试应该具有可比性，三者的效度应该既有较大的相似性，也存在一定的差异。而实际情况是否如此，有待进行全面深入的实证研究。



研究方法第7稿反馈（刘晓华 肖巍）

1)→**文献法**：课题组将首先通过大量的文献研读与研讨，结合四、六级、雅思、托福考试实际，对现有理论和框架进行充实与完善，构建科学合理的考试对比研究框架，并研制具有可操作性的考试对比测评细目。

2)→**专家判断**：课题组将运用构建的考试对比研究框架和研制的考试对比测评细目，通过专家判断，对三项考试的听力、阅读、写作和口语技能进行考试内容评定。为确保专家评定的内外部一致性，课题组将采用多人、多次评定相结合的方法。

3)→**测试法**：用三项考试的样题或真题对同一组大样本受试进行测试，将测试结果进行相关性、因子结构、分数等值等的分析。

4)→**问卷调查**：对同一组大样本受试在三项考试后进行问卷调查。

5)→**深度访谈**：对参与三项考试的部分受试进行深度半结构化访谈。

6)→**有声思维**：对一组小样本受试采用有声思维法，研究他们考试中的认知过程和策略使用。

7)→**学习日志**：对一组小样本受试采用学习日志法，跟踪他们三项考试的备考过程及考试对他们的影响。

2.4 创新之处

本课题的创新之处体现在以下四个方面：

1)→**选题新颖**：将国内很有影响的考试与国际权威的考试进行全面、系统、深入的效度对比研究，在我国尚属首例。

批注 [X11]：这些方法哪些回答上面的哪些问题？是否所有的研究问题这些方法都能回答到？比如评分效度需要对评分员的评分过程、评分员内部和评分员之间的一致性进行研究，但是研究方法中并未体现。我建议最好针对每一个研究问题逐一罗列使用的方法，那样更加清晰别人也能看到研究的可行性。

1)→**文献法**：课题组将首先通过大量的文献研读与研讨，结合四、六级、雅思、托福考试实际，对现有理论和框架进行充实与完善，构建科学合理的考试对比研究框架，并研制具有可操作性的考试对比测评细目。（构建基于证据的效度对比研究框架）

2)→**专家判断**：课题组将运用构建的考试对比研究框架和研制的考试对比测评细目，通过专家判断，对三项考试的听力、阅读、写作和口语技能进行考试内容评定。为确保专家评定的内外部一致性，课题组将采用多人、多次评定相结合的方法。（研究环境效度）

3)→**测试法**：用三项考试的样题或真题对同一组大样本受试进行测试，将测试结果进行相关性、因子结构、分数等值等的分析。（研究效标关联效度）

4)→**问卷调查**：对同一组大样本受试在三项考试后进行问卷调查。（研究考生特征）

5)→**深度访谈**：对参与三项考试的部分受试进行深度半结构化访谈。（研究考生特征）

6)→**有声思维**：对一组小样本受试采用有声思维法，研究他们考试中的认知过程和策略使用。（研究认知效度）

7)→**学习日志**：对一组小样本受试采用学习日志法，跟踪他们三项考试的备考过程及考试对他们的影响。（研究考后效度）

带格式的：字体：加粗

带格式的：字体：加粗

带格式的：字体：加粗

带格式的：字体：加粗

带格式的：字体：加粗

带格式的：字体：加粗



研究方法第9稿国外同行反馈 (Guoxing Yu)

↵

2.2 基本观点↵

本课题的基本观点是：三项考试都是以英语为外语或二语的大规模、高风险语言考试，受试将接受或正在接受高等教育，其效度应该既有较大的相似性，也存在一定的差异。而实际情况如何，需要进行全面深入的效度对比研究。↵

↵

2.3 研究思路和方法↵

本课题将以基于证据的效度验证理论 (Weir 2005) 为基础，结合四、六级，雅思，托福考试实际，运用定性和定量的研究方法，对比这三项考试的效度。具体研究思路和方法如下：↵

- 1) **文献法**：课题组将首先通过大量的文献研读与研讨，对现有理论和框架进行充实与完善，构建科学合理的考试对比研究框架，并研制具有可操作性的考试对比测评细目。(构建语言测试效度对比研

究模型)↵

- 2) **专家判断**：课题组将运用构建的语言测试效度对比研究模型，通过专家判断，对三项考试的听力、阅读、写作和口语技能的测试内容和测试的认知过程进行评价。为确保专家评定的内外部一致性，课题组将采用多人、多次评定相结合的方法。(研究环境效度和认知效度)↵
- 3) **测试法**：用三项考试的样题或真题对同一组大样本受试进行测试，用概化理论、项目反应理论、结构方程模型等多变量方法分析数据。(研究效标关联效度和评分效度)↵
- 4) **问卷调查**：对同一组大样本受试在三项考试后进行问卷调查。(研究受试特征和后果效度)↵

批注 [G8]: I understand that this may be typical for China social science funding council application, but I really feel CNY200K is far from enough...

批注 [G9]: WORD ORDER.

批注 [G10]: It would be good if you can do a diagram to indicate the relationships between these sub-projects and how they contribute to the overall understanding of "validity" (or "validities") or comparability of the three tests...

Who are the participants (test takers, and experts)? And how many would be involved in each sub-project?..

批注 [G11]: Interesting, but how would the experts judge cognitive process involved or required in the three tests and their comparability. It would be a massive project for this alone...



研究方法第9稿和第11稿学生反馈（董曼霞 申钰文）

2.3 研究思路和方法

本课题拟以Bachman & Palmer (2010)的AUA框架为出发点，结合四六级、雅思、托福考试实际，运用质性和量化相结合的研究方法，对比这三项考试的效度。具体研究思路和方法如下：

批注 [微软用户5]:这种罗列方法有些分散，其实这几种方法之间的内在逻辑性是很强的，怎么将其中的内在逻辑性呈现出来也是要考虑的。

本课题将以基于证据的效度验证理论为基础，结合测试使用论证的要素和主张，运用定性和定量的研究方法对比三项考试的效度。具体研究思路和方法如下：

1) **文献法：**课题组将首先通过大量的文献研读与研讨，结合四、六级，雅思，托福考试实际，对现有

批注 [LU21]:研究思路可否考虑画一个研究思路图更直观清晰。然后再说明每一个步骤采用了什么方法。

研究思路和方法定稿

2.3 研究思路和方法

本课题拟分五个阶段，采用七种方法收集三项考试六个方面的效度证据（见图 1）。

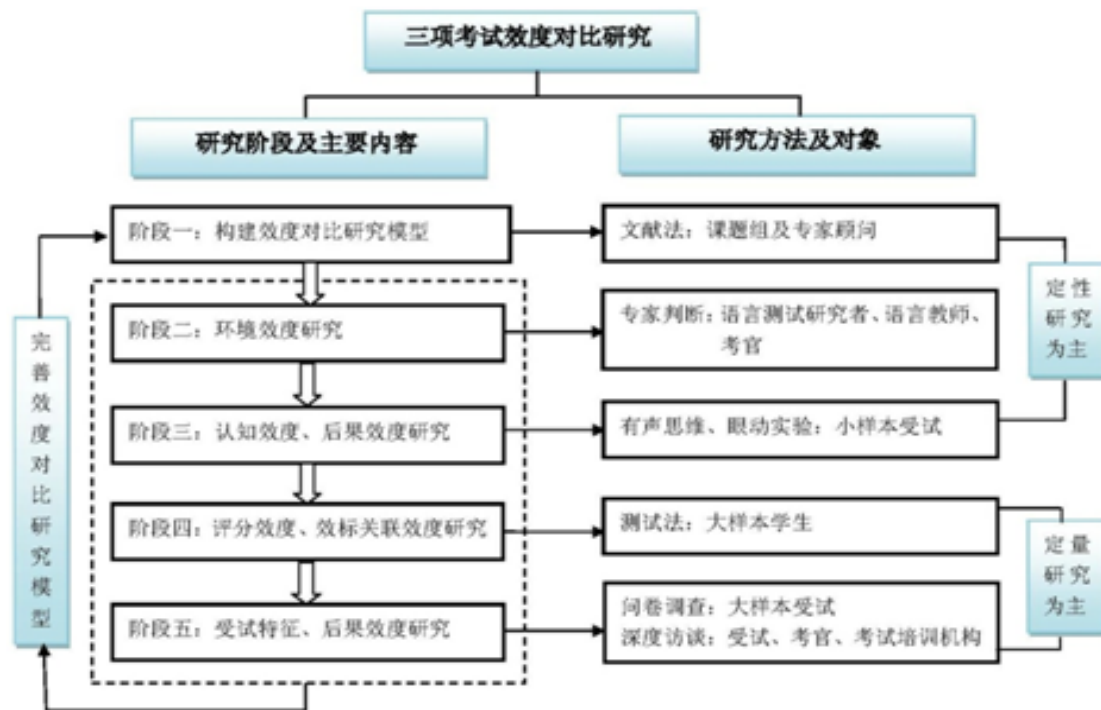


图 1 四六级、雅思、托福考试效度对比研究总体设计

具体方法和思路描述如下：

- 1) **文献法**：通过研读与研讨大量文献，对现有理论和框架进行梳理，进一步构建科学合理的考试效度对比框架和细目，并在实际操作中不断完善。
- 2) **专家判断**：专家组运用构建的框架和细目评测三项考试的任务与构念的对应关系。为确保专家评定的内外部一致性，采用多人交叉和多次评定相结合的方法。



创新之处第9稿反馈（董曼霞）

2.4 创新之处

本课题的创新之处体现在以下四个方面：

- 1) → **选题新颖**：将国内很有影响的考试与国际权威的考试进行全面、系统、深入的效度对比研究，在我国尚属首例。
- 2) → **内容全面**：研究内容既涉及三项考试各技能的测试任务、环境与操作，也包括受试的特征、完成测试任务的认知过程、交互活动、考试影响和预期目的等。
- 3) → **方法多样**：包括文献法、专家判断、测试法、问卷调查、深度访谈、有声思维和学习日志，定性与定量相结合。
- 4) → **意义深远**：本课题对三项考试的效度进行对比研究，不仅有助于充实、丰富和发展语言测试领域相关的理论建构和实践探索，而且有助于推动我国语言测试开发与研究的国际化，并为我国制定统一的语言能力等级量表，为教育、人事部门及广大的考试利益相关者提供决策或参考依据。

批注 [LU22]：创新之处，我听过一堂国家社科评审专家专门关于社科项目的申报的讲座，提到创新之处一般3点就可以了，而且一定能说明你的独到之处。

批注 [LU23]：换成“研究视角新”是否好点？

批注 [LU24]：删除这一部分。意义深远好像不是创新之处。放在研究意义中似乎更符合。



创新之处定稿

2.4 创新之处

- 1) **选题新颖**：在我国，将国内极具影响的考试与国际权威考试进行较全面系统深入的效度对比研究尚属首例。
- 2) **内容全面**：研究内容不仅包括效度对比框架的理论建构，而且涉及三项考试效度的实证对比，具体内容涵盖受试特征、测试任务、环境与操作、完成测试任务的认知过程、交互活动、分数解释、考试影响等。
- 3) **方法多样**：除效度对比研究中常用的测试法和专家判断，本课题还将采用文献法、问卷调查、深度访谈、有声思维和眼动实验，特别是有声思维和眼动实验被首度用于测试效度对比研究。



前期相关研究成果第11稿反馈（申钰文 钟瑜 肖巍）

3.1 前期相关研究成果

• **课题负责人前期相关研究成果：**在《外语界》、*Cambridge Research Notes* 等国内外学术期刊发表论文70余篇，其中CSSCI和核心期刊论文28篇；在国际语言测试学会年会、中国英语教学研究会年会等国内外学术会议上宣读论文50余篇，其中特邀发言、大会发言和专题研讨发言11次。出版国家社科结题专著1部（上下卷），

批注 [微软用户6]：这个方面可以分两部分阐释，一是课题负责人前期成果总述，二是罗列与此课题相关的成果。这样才能让人明确课题申请人着实在这方面研究的基础和积淀。

• **课题负责人前期相关研究成果：**在《外语界》、*Cambridge Research Notes* 等国内外学术期刊发表论文70余篇，其中CSSCI和核心期刊论文28篇；在国际语言测试学会年会、中国英语教学研究会年会等国内外学术会议上宣读论文50余篇，其中特邀发言、大会发言和专题研讨发言11次。出版国家社科结题专著1部（上下卷），英文专著1部，学术文集1部。

批注 [U17]：“前期相关成果”应该是与这个研究相关的测试、效度等方面的研究吧。而您这里写的是您目前已有的主要学术成果，是否需要说明您的研究方向，以及它们跟这个研究的相关性呢？另外，课题组其他主要成员中是否也有相关领域专家，若他们有相关研究，在此列出应能更有效地凸显团队有做好这个项目的能力。或者，一方面从总量上凸显成员做科研的能力，另一方面从相关性上凸显前期已做大量相关研究。

3.2 开展本课题研究的主要参考文献

1. → AERA, APA, & NCME. (1985). *Standards for Educational and Psychological Testing*. Washington, DC: APA.
2. → Bachman, L. F., Davidson, F., Ryan, K. & Choi, I. C. (1995). *An Investigation into the Comparability of Two*

3.1 前期相关研究成果

• **课题负责人前期相关研究成果：**在《外语界》、*Cambridge Research Notes* 等国内外学术期刊发表论文70余篇，其中CSSCI和核心期刊论文28篇（其中效度研究论文xx篇，cssci和核心期刊论文xx篇），在国际语言测试学会年会、中国英语教学研究会年会等国内外学术会议上宣读论文50余篇（效度研究论文宣读xx篇），其中特邀发言、大会发言和专题研讨发言11次（效度研究论文发言xx次）。出版国家社科结题专著1部（上下卷），英文专著1部，学术文集1部。

批注 [U15]：加个《课程·教材·教法》？

批注 [U16]：在效度研究发了多少才能说明你在效度研究上的实力。总共多少篇文章，只是你的总体科研实力。



前期相关研究成果第一次提交

3.1 前期相关研究成果⁴

由于篇幅限制，我们无法一一列出前期相关研究成果的名称。在此用一个较直观的图来呈现课题负责人及课题组成员前期的相关研究成果形式、作者排序、发表的级别、被引用的频次以及效度研究的成果数量，以期能较简略地呈现我们前期已有的相关研究成果（见图2）。⁴

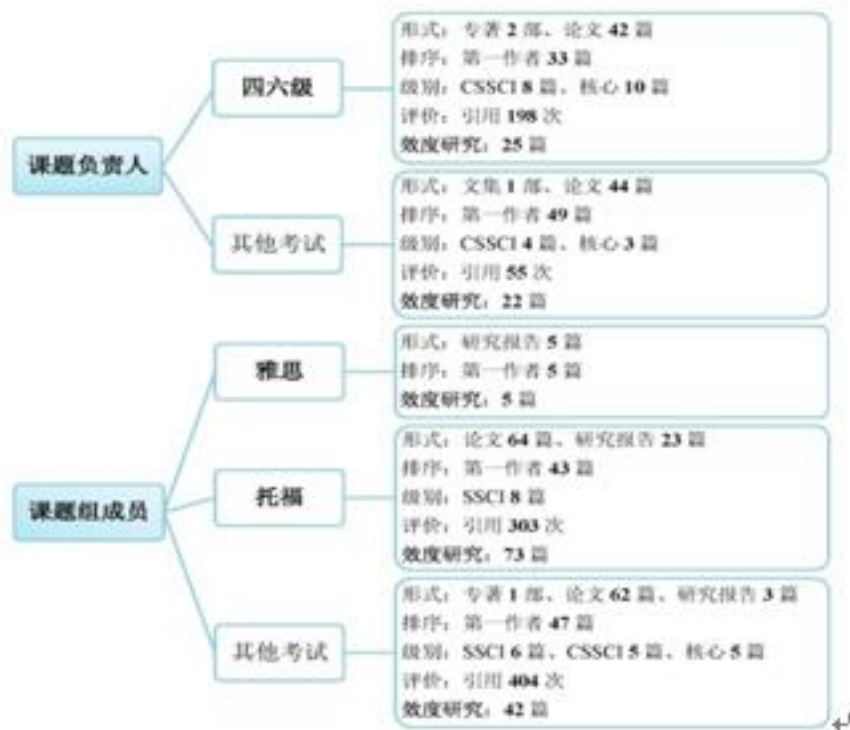


图2 课题负责人和课题组成员前期相关研究成果⁴



前期相关研究成果定稿

3.1 前期相关研究成果

· **课题负责人前期相关研究成果：**四六级专著2部、论文42篇，其中效度研究25篇、第一作者33篇、CSSCI 8篇、核心10篇；其他考试文集1部、论文40篇，其中效度研究22篇、第一作者33篇、CSSCI 4篇、核心3篇。代表性成果有：

- 1) 怀疑与误解——评《文汇报》文章“要素还是应试？”（论文，唯一作者，CSSCI）
- 2) 在大学英语四六级考试中增加快速阅读的必要性（论文，唯一作者，CET 改革采纳）
- 3) 正面的还是负面的一大学英语四六级考试反拨效应实证研究（博士论文专著，独著）
- 4) 高考英语全国卷与各省自主命题卷共时与历时研究（学术文集，主编）
- 5) 剑桥商务英语在中国的影响（研究报告，唯一作者，剑桥大学外语考试部采纳）
- 6) 大学英语四六级考试反拨效应历时研究（国家社科项目结题专著，第一作者，主笔）
- 7) 改革后 CET 听力测试语篇输入与预期回答任务特征分析（论文，第一作者，核心）
- 8) 大学英语四级考试快速阅读部分内容效度研究（论文，第一作者，CSSCI）
- 9) 大学英语教师对 CET 认识的反拨效应历时研究（论文，第一作者，CSSCI）
- 10) CET 二十年写作试题分析与研究（论文，第一作者，CSSCI）

· **课题组前期相关研究成果：**雅思和托福考试论文 64 篇、研究报告 28 份，其中效度研究 78 篇、第一作者 48 篇、SSCI 8 篇；其他考试专著 1 部、论文 62 篇、研究报告 3 份，其中效度研究 42 篇、第一作者 47 篇、SSCI 6 篇、CSSCI 5 篇、核心 5 篇。代表性成果有：

- 1) Test for English Majors (TEM) in China (论文，第一作者，SSCI)
- 2) Evaluating analytic scoring for the TOEFL® Academic Speaking Test (TAST) for operational use. (论文，唯一作者，SSCI)
- 3) Validating TOEFL® iBT speaking and setting score requirements for ITA screening (论文，唯一作者，SSCI)
- 4) How do we go about investigating test fairness? (论文，唯一作者，SSCI)
- 5) Lexical diversity in writing and speaking task performances (论文，唯一作者，SSCI)
- 6) The cognitive processes of taking IELTS academic writing task one (研究报告，剑桥大学外语考试部采纳，第一作者)
- 7) 语言测试的标准研究：回顾、反思与启迪（论文，第一作者，CSSCI）
- 8) 大学专业英语的测试信度与效度研究（论文，唯一作者，CSSCI）
- 9) 语言测试构念效度研究（专著，独著）
- 10) 复杂结构语言测试信度的多元概化分析——以 CET-6 数据分析为例（论文，第一作者，核心）



主要参考文献第10稿自查发现英文全是专著

3.2 开展本课题研究的主要参考文献

1. AERA, APA, & NCME. (1985). *Standards for Educational and Psychological Testing*. Washington, DC: APA.
2. Arcuino, C. (2013). *The relationship between the test of English as a foreign language (TOEFL), the international English language testing system (IELTS) scores and academic success of international master's students*. Doctoral dissertation, Fort Collins, Colorado: Colorado State University.
3. Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
4. Bachman, L. F., Davidson, F., Ryan, K. & Choi, I. C. (1995). *An Investigation into the Comparability of Two Tests of English as a Foreign Language-The Cambridge-TOEF Comparability Study*. New York: Cambridge University Press.
5. Chapelle, C., Enright, M. K., & Jamieson, J. (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York and London: Routledge.
6. Messick, S. (1989). Validity. In R. L. Linn (ed.). *Educational Measurement* (3rd edition). New York: Macmillan.
7. Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.
8. Khalifa, H. & Weir, J. C. *Examining Reading: Research and practice in assessing second language reading*. Cambridge: Cambridge University Press.
9. TOEFL iBT @Research Insight Series Vol.1-Vol.6 Retrieved from http://www.ets.org/toefl/research/ibt_insight_series/
10. IELTS Research Reports & Online Series (1998-2014), retrieved from <http://www.ielts.org/researchers/research.aspx>



主要参考文献定稿

3.2 开展本课题研究的主要参考文献⁴¹

- [1] AERA, APA, & NCME. (1985, 1999). *Standards for Educational and Psychological Testing*. Washington, DC: APA.⁴¹
- [2] Bachman, L.F., & Palmer, A. (2010). *Language Assessment in Practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.⁴¹
- [3] Cho, Y. et al. (2013). Investigating the effects of prompt characteristics on the comparability of TOEFL iBT™ integrated writing tasks. *Language Testing*, 30(4), 513-534.⁴¹
- [4] *IELTS Research Reports & Online Series* (1998-2014), retrieved from <http://www.ielts.org/researchers/research.aspx>⁴¹
- [5] Messick, S. (1989). Validity. In R. L. Linn (ed.). *Educational Measurement* (3rd edition). New York: Macmillan.⁴¹
- [6] *TOEFL iBT® Research Insight Series*. Vol.1-Vol.6 Retrieved from http://www.ets.org/toefl/research/ibt_insight_series/⁴¹
- [7] Weir, J. C. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan. ⁴¹
- [8] Weir, J. C. et al. (2013). *Measured Constructs*. Cambridge: Cambridge University Press.⁴¹
- [9] Xi, X. (2007). Validating TOEFL® iBT speaking and setting score requirements for ITA screening. *Language Assessment Quarterly*, 4(4), 318-351.⁴¹
- [10] Allami, H., & Aghajari, J. (2014). Pragmatic knowledge assessment in listening sections of IELTS tests. *Theory and Practice in Language Studies*, 4(2), 332-340.⁴¹
- [11] 范劲松, 金艳. (2010). 语言测试标准研究: 回顾、反思和启迪. 《外语界》第一期: 82-91. ⁴¹
- [12] 韩宝成, 罗凯洲. (2013). 语言测试效度及其验证模式的嬗变. 《外语教学与研究》第三期: 411-425. ⁴¹
- [13] 金艳, 张晓艺. (2013). 技能综合对语言测试构念效度的影响——培生英语考试与大学英语六级网考的对比研究. 《外语电化教学》第六期: 3-10. ⁴¹
- [14] 李清华. (2007). 《语言测试与效度验证——基于证据的研究方法》述介. 《现代外语》第二期: 214-217. ⁴¹
- [15] 杨惠中 & Weir, C. (1998). 《大学英语四六级考试效度研究》. 上海: 上海外语教育出版社. ⁴¹



主要学术简历学生反馈（田洁）

1. 课题负责人的主要学术简历、在相关研究领域的学术积累和贡献

主要学术简历：课题负责人是重庆市人文社科重点研究基地重庆大学语言认知及语言应用研究基地主任兼语言测试研究所所长，专职研究员，重庆大学外国语学院教授，硕士生导师（本校英语专业无博士点），2012年10月起被剑桥大学外语考试部聘为兼职高级研究员。上海交通大学博士毕业，师从大学英语四、六级考试创始人杨惠中教授（2001.3-2005.3），已经从事了十余年大学英语四、六级考试研究，特别是内容效度和反拨效应研究。2005年9月至2006年9月，在美国加州大学洛杉矶分校做访问学者，师从世界著名语言测试学家Lyle F Bachman，他是最早在我国引介其语言测试理论和学术思想的学者之一。2011年10月至2012年10月，课题负责人获国家留学基金委全额奖学金资助，在剑桥大学外语考试部做访问教授期间，与效度和发展研究部主任Dr Nick Saville及其团队密切合作，进行剑桥主体英语考试在中国的影响的研究。课题负责人目前与世界著名语言测试学家Cyril J Weir同为剑桥大学外语考试部兼职高级研究员，有较多的学术联系与交流，非常熟悉其社会认知效度验证理论和基于证据的效度验证模式。

删除的内容：我们没有

删除的内容：。

删除的内容：剑桥大学外语考试部

删除的内容：研究



主要学术简历定稿

2. 课题负责人的主要学术简历、在相关研究领域的学术积累和贡献⁴

主要学术简历：课题负责人是重庆市人文社科重点研究基地重庆大学语言认知及语言应用研究基地主任兼语言测试研究所所长、专职研究员、重庆大学外国语学院教授、硕士生导师（我校本专业无博士点）。2012年10月起被剑桥大学外语考试部聘为兼职高级研究员。上海交通大学博士毕业，师从大学英语四六级考试创始人杨惠中教授（2001.3-2005.3）。现已从事大学英语四六级考试研究十余年，重点研究其内容效度和反拨效应。2005年9月至2006年9月，课题负责人在美国加州大学洛杉矶分校做访问学者，师从世界著名语言测试学家Lyle F. Bachman教授，是最早在我国引介其语言测试理论和学术思想的学者之一。2011年10月至2012年10月，课题负责人获国家留学基金委全额奖学金资助，在剑桥大学外语考试部做访问教授期间，与效度和发展研究部主任Nick Saville博士及其团队密切合作，进行剑桥主体英语考试在中国的影响研究。课题负责人目前与世界著名语言测试学家Cyril J. Weir教授同为剑桥大学外语考试部兼职高级研究员，有较多的学术联系与交流，非常熟悉其社会认知效度验证理论和基于证据的效度验证模式。⁴



主要参加者

专业 职务	学位	研究专长	工作单位
教授	博士	专门用途英语	重庆大学
高级 研究员	博士	教育测量与 语言测试	普林斯顿大学 (美国)
教授	博士	语言测试	重庆邮电大学
高级讲师 (副教授)	博士	语言测试	布里斯托大学 (英国)
讲师	博士	语言测试	复旦大学
博士生	在读博士	心理测量 与统计	南京师范大学



经费预算

序号↵	经费开支科目↵	金额（元）↵	序号↵	经费开支科目↵	金额（元）↵
1↵	资料费↵	30,000↵	7↵	专家咨询费↵	40,000↵
2↵	数据采集费↵	100,000↵	8↵	劳务费↵	40,000↵
3↵	差旅费↵	127,000↵	9↵	印刷费↵	10,000↵
4↵	会议费↵	30,000↵	10↵	管理费↵	3,000↵
5↵	国际合作与交流费↵	↵	11↵	其他费用↵	↵
6↵	设备费↵	20,000↵	合计↵	400,000↵	
年度经↵ 费预算↵	年份↵	2014 年↵	2015 年↵	2016 年↵	2017 年↵
↵	金额（元）↵	60 000↵	120 000↵	140 000↵	80 000↵



极致人生

最好的思路与方法是无时无刻不思考与行动

- ◆ 56岁UCLA读博的黑人老太太
- ◆ 有四个孩子却在全职工作的剑桥同事
- ◆ 评了教授还去读博的重大同事
- ◆ 读硕期间恋爱、结婚、生孩子、申CSC奖学金国外攻博的学生
- ◆ 45岁生日凤凰涅槃
- ◆ 女人要团结一切可以团结的力量
-



56岁UCLA读博的黑人老太太





有四个孩子却在全职工作的剑桥同事





评了教授还去读博的重大同事



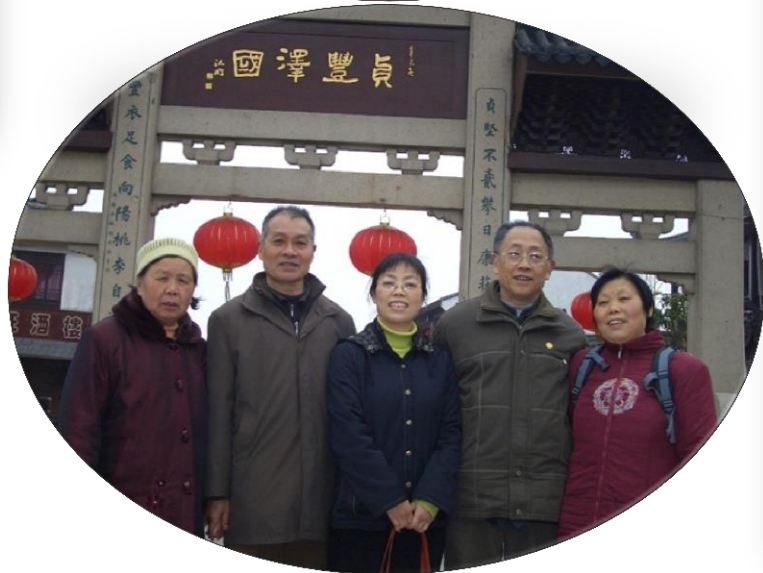
读硕期间 恋爱 结婚 生孩子 申CSC奖学金国外攻博的学生





45岁生日凤凰涅槃





和谐生态

女人要团结一切
可以团结的力量





◆ **Reading** is to the mind what exercise is to the body.

——Central Library Cambridge

◆ To strive, to seek, to find, and **not to yield.**

——Alfred Tennyson

For A Better Tomorrow

WHY NOT YOU

Today is a new day!

Many will seize this day.

Many will live it to the fullest.

Why not you?



JUST SAIL

As we sail through life,

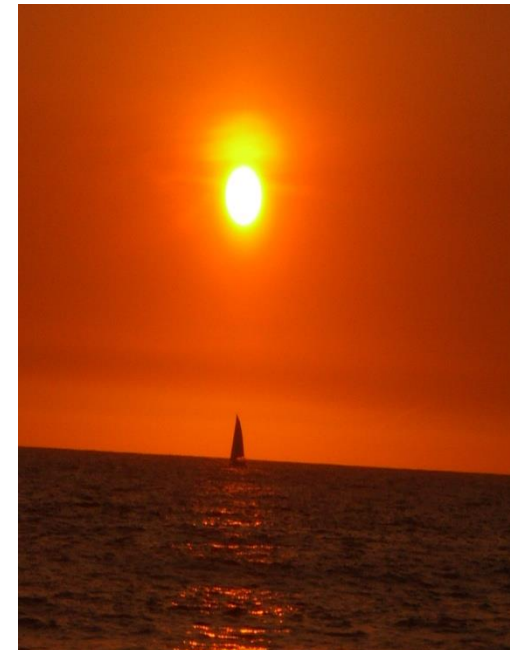
Don't avoid storms and rough waters.

Just let it pass.

Just sail,

Always remember,

Calm seas never make skillful sailors.



Dreams

---- Langston Hughes

Hold fast to dreams

For if dreams die

Life is a broken-winged bird

That cannot fly.

Hold fast to dreams

For when dreams go

Life is a barren field

Frozen with snow.

千万不能丢掉梦想，

因为假如梦想破灭

生活就像小鸟断了翅膀

再也不能展翅飞翔

千万不能丢掉梦想，

因为梦想一旦离去

生活就像冬天的田野那样荒凉

冰封大地白雪茫茫



Thank you!

新浪博客：辜向东-人生GU事

微信公众号：人生GU事

微信好友：fudounainai

邮箱：xiangdonggu@263.net