

International Conference on Language Testing and Assessment

语言测试与评价国际研讨会

November 27-29 2015

2015年11月27-29日

主办单位 (Hosted by):



广东外语外贸大学 Guangdong University of Foreign Studies



教育部考试中心 National Education Examinations Authority

承办单位 (Organized by):



教育部人文社科重点研究基地 外国语言学及应用语言学研究
National Key Research Center for Linguistics and Applied Linguistics

协办单位 (In association with):

高等教育出版 High Education Press

广东外语外贸大学教师发展中心 Center for Teacher Development of GDUFS

赞助单位 (Sponsored by):

英国文化委员会 British Council

外语教学与研究出版社 Foreign Language Teaching and Research Press

北京外研讯飞教育科技有限公司 Beijing FIF Education Technology Co. Ltd

批改网 www.pigai.org Pigai - automatic English Essay Scoring Service

The Organizing Committee

Chair

LIU Jianda 刘建达

Director of the National Key Research Center for Linguistics & Applied Linguistics

Vice President of Guangdong University of Foreign Studies

Co-Chairs

REN Wei 任伟

The National Key Research Center for Linguistics & Applied Linguistics

XU Hai 徐海

Deputy Director of the National Key Research Center for Linguistics & Applied Linguistics

Members

HAN Jiaxun 韩家勋 *National Education Examinations Authority*

Head of the National Foreign Language Assessment Framework Project Team

WU Sha 吴莎 *National Education Examinations Authority*

Deputy Head of the National Foreign Language Assessment Framework Project Team

YANG Jing 杨静

Deputy Director of the National Key Research Center for Linguistics & Applied Linguistics

Logistics

XIAO Yuanling 肖沅陵

PHD candidates from Class 2015

Postgraduate students in language testing

LIU Hongying 刘洪英

文科基地 2015 级博士生

语言测试方向硕士研究生

Website Designer

LI Xia

李霞

ZHENG Changqing

郑长青

Student Assistants

HUANG Weiyang 黄玮莹 ZHOU Yanqiong 周艳琼 PENG Chuan 彭川

CAO Linlin 曹琳琳 FAN Tingting 范婷婷 Wang Hua 王华

DING Jiaolong 丁娇龙 XIE Fang 谢芳 Zhou Yan 周燕

YE Wei 叶薇 LIU Yang 刘洋 ZOU Run 邹润

XU Liu 徐柳 YAN Shenglan 晏盛兰

Conference Schedule

FRIDAY, NOV. 27	
09:00-21:00	Registration (Lobby of TB No.7 七教一楼大厅)
SATURDAY, NOV. 28	
08:30-09:00	Opening Ceremony (Lecture Hall, Lower Ground Floor, GDUFS Library 图书馆负一层报告厅)
09:00-09:20	Photo Taking (Outside GDUFS library 图书馆门外)
09:20-10:30	Keynote Speech (1) Prof. LIU Jianda et al. (Lecture Hall, Lower Ground Floor, GDUFS Library 图书馆负一层报告厅)
10:30-10:40	Tea Break (Hall of Lower Ground Floor, GDUFS Library 图书馆负一层大厅)
10:40-11:20	Keynote Speech (2) Prof. Nick Saville (Lecture Hall, Lower Ground Floor, GDUFS Library 图书馆负一层报告厅)
11:20-12:00	Keynote Speech (3) Prof. Barry O’Sullivan (Lecture Hall, Lower Ground Floor, GDUFS Library 图书馆负一层报告厅)
12:00-14:00	Lunch (First Floor, Poly Hotel, 保利山庄一层)
14:00-17:50	Parallel sessions (TB No.7 七教)
18:30	Conference Dinner (Easeland Hotel, 逸林酒店)
SUNDAY, NOV. 29	
08:30-12:20	Parallel sessions (TB No.7 七教)
12:20-14:00	Lunch (First Floor, Poly Hotel, 保利山庄一层)
14:00-14:40	Keynote Speech (4) Prof. CHENG Liying (Lecture Hall, Lower Ground Floor, GDUFS Library 图书馆负一层报告厅)
14:40-15:20	Keynote Speech (5) Prof. John de Jong (Lecture Hall, Lower Ground Floor, GDUFS Library 图书馆负一层报告厅)
15:20-15:40	Tea Break (Hall of Lower Ground Floor, GDUFS Library 图书馆负一层大厅)
15:40-16:40	Invited Forum Prof. GUI Shichun (Lecture Hall, Lower Ground Floor, GDUFS Library 图书馆负一层报告厅)
16:40-17:00	Closing ceremony (Lecture Hall, Lower Ground Floor, GDUFS Library 图书馆负一层报告厅)

Contents

CONTENTS	1
CONFERENCE PROGRAM	2
ABSTRACTS OF KEYNOTE SPEECHES	21
KEYNOTE SPEECH 1	21
KEYNOTE SPEECH 2	22
KEYNOTE SPEECH 3	23
KEYNOTE SPEECH 4	24
KEYNOTE SPEECH 5	25
INVITED FORUM.....	26
ABSTRACTS OF PARALLEL SESSIONS	27
SATURDAY SESSIONS	27
ROOM 301.....	27
ROOM 302.....	32
ROOM 311.....	36
ROOM 313.....	41
ROOM 316.....	46
ROOM 317.....	51
ROOM 318.....	56
ROOM 422.....	60
ROOM 423.....	66
ROOM 523.....	73
ROOM 714.....	78
ROOM 715.....	85
SUNDAY SESSIONS.....	89
ROOM 301.....	89
ROOM 302.....	94
ROOM 311.....	100
ROOM 313.....	105
ROOM 316.....	109
ROOM 317.....	114
ROOM 318.....	120
ROOM 422.....	125
ROOM 423.....	130
ROOM 714.....	136
ROOM 715.....	141
AUTHOR INDEX	148

Conference Program

FRIDAY, Nov 27th, 2015			
09:00-21:00	Registration (报到注册)	Location: Lobby of Teaching Building No.7 第七教学楼一楼大厅	
SATURDAY, Nov 28th, 2015			
08:30-09:00	Opening Ceremony Chair LIU Jianda <i>Guangdong University of Foreign Studies</i>	Opening address by the President of Guangdong University of Foreign Studies; The President of the National Education Examinations Authority	图书馆负一层报告厅 Lecture Hall, Lower Ground Floor, GDUFS Library
9:00-9:20	Photo Taking (图书馆门外)		

09:20-10:30	Keynote Speech (1) Chair Prof. REN Wei	Prof. LIU Jianda, Prof. WU Zunmin, Prof. HE Lianzhen, Prof. JIN Yan, Prof. ZENG Yongqiang, Prof. ZOU Shen, Prof. HAN Baocheng, Prof. YAN Ming, Prof. MU Lei Developing Scales for Chinese EFL Learners	图书馆负一层报告厅 Lecture Hall, Lower Ground Floor, GDUFS Library
10:30-10:40	Tea Break (图书馆负一层大厅)		
10:40-11:20	Keynote Speech (2) Chair Prof. Antony Kunnan	Prof. Nick Saville Developing a learning ladder to set language standards	图书馆负一层报告厅 Lecture Hall, Lower Ground Floor, GDUFS Library
11:20-12:00	Keynote Speech (3) Chair Prof. JIN Yan	Prof. Barry O'Sullivan Testing Teachers: Issues and solutions	
12:00-14:00	Lunch (First Floor, Poly Hotel, 保利山庄一层)		
14:00-17:50	Parallel sessions (Teaching Building No.7)		

Room 301 THEME Development of Language Proficiency Scales Chair 刘芹	14:00-14:30	金艳、揭薇 Developing an English speaking scale for teaching, learning and assessment in China: Principles, procedures and progress
	14:30-15:00	揭薇 解读英语口语能力描述语：基于口语量表描述语库的研究
	15:00-15:30	王淑花 Developing Operational Framework and Descriptors of Pragmatic Effectiveness
	15:30-16:00	泰中华、刘芹 多维度大学生英语写作能力标准构建
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	贾国栋、王晓彤 大学英语写作能力标准的制定与实施--以中国人民大学为例
	16:50-17:20	黄玮莹 Development of Scales of Oral Pragmatic Competence
Room 302 (Panel 3.1) THEME Validity and validation Chair FAN Jinsong	14:00-14:30	李清华 Validity Issues of Second Language Dynamic Assessment
	14:30-15:00	FAN Jinsong ZHENG Yongyan & ZHAO Guanfang Building a Validity Argument for a School-Based English Test: Evidence from Construct and Predicative Validity
	15:00-15:30	江进林 基于 Coh-Metrix 的大学英语六级与托福阅读语篇难度研究
	15:30-16:00	牟欣 学术英语大规模读写测试的效度研究
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	CHEN Yangfang On the Validity and Discrimination of Cet-4 after Reform——Cet-4 performance of **

		University in Dec. 2007 as an Example
	16:50-17:20	施雅俐 Mixed-method Research in Language Testing: History, Practice and Trend
	17:20-17:50	张妙 A Study of NMET Writing Test Items of Guangdong Province from 2007 to 2014
<u>Room 311</u> <u>THEME</u> Reform and development of China's English testing Chair 刘辅兰	14:00-14:30	GU Xiangdong 50 Years of High-stakes Exams in China—A Personal Journey
	14:30-15:00	陈文存 An Analysis of the Problems of 'Scientific Design' of English Test Contents of College Entrance Examination
	15:00-15:30	刘辅兰 Writing prompt convergence in high-stakes tests: Exploring alternative rhetorical functions and objects of enquiry
	15:30-16:00	徐莎莎 Consensus and divergence: Voices of test-takers and raters under a high-stakes language test
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	陶百强 Reform China's Gaokao item writing mechanism—From secluded item writing to an online test item submission system
	16:50-17:20	HUANG Rui New National College Entrance Examination and Basic Strategy for English Academic Test Evaluation
	<u>Room 313</u>	14:00-14:30

<u>THEME</u> Construct in language testing and scales Chair 王巍巍	14:30-15:00	肖立宏 从促进人全面发展的角度来看英语语言测试----以 2015 年高考英语北京卷为例
	15:00-15:30	YE Wei Investigating the story continuation writing task
	15:30-16:00	许艺、王巍巍、伍志伟、穆雷 交替传译策略能力的构成研究
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	王巍巍、许艺、穆雷 “语言能力”与“口译能力”刍议
	16:50-17:20	LIU Jianzhu A Study on the CEFR-oriented Construct of Interpreting Competence
	17:20-17:50	ZHU Meifang A Preliminary Study on the Scales of Chinese Middle School Students' Strategic Competence in Oral English Communication
<u>Room 316</u> <u>THEME</u> Classroom Assessment Chair TANG Xiongying	14:00-14:30	Saefurrohman Classroom Assessment Practices Of Indonesian Junior High School Teachers In English As Foreign Language (EFL) Classes
14:30-15:00	WEI Wei, HU Rong Competing Feedback From Summative Assessments And Classroom Instructions On English Major Students' Use Of Reading Comprehension Strategies	
15:00-15:30	LI Xuelian Assessment For Learning In Classroom: Students' Management Of Their Own Learning Goals	
15:30-16:00	XU Yueting Feasibility And Challenges Of Assessment For Learning In Large Classes	

	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	TANG Xiongying Turning Assessment into Mediated Learning Experience -- Assessing EFL Students' Independent Learning by Learning Logs
	16:50-17:20	刘春燕 高校外语课程中的任务型语言评价探索
<u>Room 317</u>	14:00-14:30	YUAN Shuhou Formative assessment studies in English teaching in China: review and reflection
<u>THEME</u> Formative Assessment	14:30-15:00	CHEN Qiuxian The cultural appropriateness and appropriation of formative assessment in the Chinese sociocultural context
	15:00-15:30	FAN Meng 形成性评估在非英语专业研究生英语写作教学中的应用研究
Chair YUAN Shuhou	15:30-16:00	April Jiawei Zhang The impact of formative assessment on motivation and achievement in English classes in Chinese primary schools
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	CHEN Chao On foreign language teachers' assessment literacy in portfolio assessment
	16:50-17:20	CAI Wenyan 大学英语教师口语课堂形成性评价能力研究---教学支架使用情况调查
<u>Room 318</u>	14:00-14:30	ZHANG Xinling, ZHU Lin A Study into the Effect of Task Type on Linguistic Features of CFL Learners' Writings
<u>THEME</u>	14:30-15:00	伍志伟、许艺、穆雷

Test tasks and test development 范香娟		翻译专业本科成就测试口译部分题型研究
	15:00-15:30	范香娟 外国留学生语用能力测试
	15:30-16:00	Reima Al-Jarf What Teachers Should Know about Vocabulary Tests for EFL Freshman Students
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	Sercan SAGLAM Over testing or double testing
	16:50-17:20	张惠玲 测试方法对语法测试的影响.
Room 422	14:00-14:30	HAN Ye Exploring Chinese EFL learners' cognitive processing of written corrective feedback: A case study
THEME	14:30-15:00	WANG Fang, HE Junjie, WU Linqiong A Correlation Study on Key Positioning and Test Performance with Language Test Multiple Choice Items
Focusing on test takers	15:00-15:30	ZHANG Xinling, ZHOU Yan, ZHANG Siyu Chinese EFL Learners' Writing Process in Reading-to-write and Reading-listening-writing Integrated Writing Tasks
Chair HAN Ye	15:30-16:00	XU Yiyi, ZHANG Ping The Effect of Priming Position and Learners' Proficiency Level on Contextualized Word Association
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	TIAN Jie, GU Xiangdong, ZHENG Yujing A Verbal Protocol Analysis of Test-takers' Processes of Taking Reading Test of Cambridge

		English: Advanced
	16:50-17:20	DENG Yanping, YANG Ruijin The Application of Think Aloud Protocol in the Reading Strategy of New HSK Band 5 Reading Test
	17:20-17:50	YANG J GU Xiangdong HONG Yue An Introspective Study on Test-takers' Process of Keywords Composition in New HSK-5
<u>Room 423</u>	14:00-14:30	BAI Ying A Students' Perspective on Intended Consequences of a High-stakes Test in China and Its Effect on Test Performance
<u>THEME</u>	14:30-15:00	亓鲁霞 Applying the Dynamic System Theory to an analysis of test stakes
Testing	15:00-15:30	HAO Chenqian, CHENQiuxian Could the consequences of test be removed? -- Perspectives from Chinese EFL students of tertiary level
Consequences	15:30-16:00	张放 Why do school policies on CELST test preparation differ? --- An investigation into the washback of CELST through examining school administrator attitudes and school policies
Chair	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
高霄	16:20-16:50	黄妍琪、肖巍、赵路、辜向东 剑桥商务英语考试的测试影响——以考生为视角
	16:50-17:20	高霄 学术素养培养为导向的语言测试反拨效应研究
	17:20-17:50	王华 Tracking students' change of language ability during their preparation for NMET
	14:00-14:30	ZHENG Ying, ZHUANG Yanyan, YAN Youyang Investigating the Practice of the CEFR outside Europe: A Case Study on English Writing

<u>Room 523</u> <u>THEME</u> CEFR and Beyond Chair ZHENG Ying		Assessment in China
	14:30-15:00	YAN Qiaozhen, GU Xiangdong, YU Chengyuan An Investigation of Chinese English Majors' English Language Competence with Reference to the CEFR
	15:00-15:30	PENG Chuan Aligning Chinese EFL learners with the CEFR
	15:30-16:00	张晓蕾、张萍 Sentential Context Priming in Word Association
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	ZHANG Xiaoyi ZOU Shaoyan Effects of the Formulation of CEFR Descriptors on Scale Comprehensibility and Calibration
	16:50-17:20	HU Wenting, HE Cencheng An Empirical research on Correlation of Language Elements and Language Level in Net Text
	17:20-17:50	甘凌 Ethics in Language Testing: Review and Implication
<u>Room 714</u> <u>THEME</u> Language testing and teaching Chair ZHOU Jiming	14:00-14:30	PENG Yuanling Design a Rubric Aligned with Intended Learning Outcomes to guide Business Case Study Presentation
	14:30-15:00	LI Tian Exploring the Comparability of TEM4 Reading Tests and English-major Students' Textbooks in One of China's International Studies University
	15:00-15:30	ZHOU Jiming From rigidly copying to flexibly implementing: lessons learned from an institutional-level assessment change
	15:30-16:00	Chan Kwong Tung

		The Application of Bloom's Taxonomy in the Teaching and Assessment of Reading Comprehension: the case of Chinese Language in Singapore
	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	BIN Ke, ZHOU Yan A Comparative Study on the Influences of Different Types of Integrated Writing Tasks on Chinese EFL Learners' Source Use
	16:50-17:20	姚雯静 Unguided assessors along the way---A case study on tertiary EFL teachers' conceptions and practices of assessment in one college from China
	17:20-17:50	陈璇 A Research and Design on the MTI Interpretation Teaching Resources Platform
<u>Room 715</u>	14:00-14:30	WANG Yihong Impact of Integrating an Automated Assessment Tool into English Writing on Teachers
<u>THEME</u>	14:30-15:00	MA Rui, GU Gang The Impact of Automated Writing Evaluation on EFL Writing
Information Technology In Language Testing	15:00-15:30	Roxanne Wong Designing an automated essay scoring system
	15:30-16:00	WU Jinyu Computerized Integrated Writing: Introduction and Application
Chair WANG Yihong	16:00-16:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	16:20-16:50	LIU Li, GAO Ying Confirmatory Factor Analysis of Speaker's Oral English Level Scale Based on Identity
	16:50-17:20	杨洋 跨文化交际能力测量工具研究综述

18:30	Conference Dinner (Easeland Hotel, 逸林酒店)	
SUNDAY, Nov 29th, 2015		
08:30-12:20	Parallel sessions (Teaching Building No.7)	
<u>Room 301</u> <u>THEME</u> Development of Language Proficiency Scales Chair 冯莉	08:30-09:00	HE Lianzhen, CHEN Dajian, MIN Shangchao Theoretical backing and practice reasoning of listening ability scale
	09:00-09:30	严明、冯莉 翻译能力描述语开发研究——兼谈中国英语能力（翻译）量表子项目设计
	09:30-10:00	PAN Mingwei ZOU Shen Scaling exemplary writing activities: An SFL text typology perspective
	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	唐莉莉 中国大学生英语口语互动能力等级量表研究
	10:50-11:20	程欣、陈桦 A Comparative Study on English Pronunciation Descriptors Used Abroad and at Home
	11:20-11:50	赵护林、王巍巍、许艺、穆雷 On the Collection of Descriptors of China Common Framework for English Language Proficiency (Interpreting)
	11:50-12:20	周艳琼 The Development of EFL Reading Strategy scale of Chinese College Students

<u>Room 302</u>	08:30-09:00	ZHU Zhengcai & ZHANG Lidong A Validation Framework for the National English Proficiency Scale of China
	09:00-09:30	李廉 An EFA-based Study of Simulating Construct Validity of Objective Listening Test
	09:30-10:00	ZOU Shaoyan, ZHANG Xiaoyi An Investigation into the Context Validity of CEFR Writing Descriptors
	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	李乔娜 The comparison of content validity study on cloze tests between the NMET of Guangdong and Nation (2011-2015)
	10:50-11:20	杨秀芬 The Comparative Study of the Validity between the Senior High School Entrance English Examination and the TOFEL Junior Standard Test —— An Evidence-based Approach
	11:20-11:50	张晓丹 The comparison of content validity study on reading comprehension between the NMET of Guangdong and Nation (2011-2015)
<u>THEME</u> Validity and validation Chair ZHANG Lidong	11:50-12:20	龚晨、张萍 A Study on the Context Effects in Word Association Task
	<u>Room 311</u>	张敏 标准设定应用于大学英语听力成就测试的可行性分析
	09:00-09:30	张锦涛、吴苓 美军国防语言水平测试体系建设及启示
	09:30-10:00	肖奕、吴宜静、王凯旋 朱汉雄 国内语言测试公平性研究的现状与对策
<u>THEME</u> Reform and development of	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)

China's English testing Chair 张锦涛	10:20-10:50	LIU Xiurong 独立学院ESP测评体系探究——以金城学院为例
	10:50-11:20	YANG Zhiqiang, YANG Ruijin, MENG Lei A Review of English Speaking Test Studies in China over the Past Three Decades
	11:20-11:50	程蒙蒙 关于制定国家外语考试质量标准的研究及建议
Room 313 THEME Construct in language testing Chair HUO Yongshou	08:30-09:00	HUO Yongshou The Ineffability of Language Ability and Philosophical Considerations in Language Testing
	09:00-09:30	ZHU You yi Decoding and Reconstructing of Textual Signs
	09:30-10:00	赵辑 文化资本视角下的外语测试研究
	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	盛艳玲 社会语言能力相关文献综述
	10:50-11:20	陈灿 从描述语角度探究跨文化交际能力的构成要素
	11:20-11:50	LIU Qian An Analysis of Semantic Structure of “Can Do” Descriptors for Textual Competence
Room 316 THEME Classroom Assessment	08:30-09:00	徐嗣群 理工类英语课程形成性考核评价体系的构建与实践——以南方科技大学为例
	09:00-09:30	YANG Maoxia SHAO Zheng Action Research on Formative Assessment of Research Project Course
	09:30-10:00	LI Danli, LU Xiaoxian Innovation in EFL Creative Writing through Focus on Assessment for Learning

Chair 徐嗣群	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	GUO Ying A Study on Dynamic Qualitative Assessment for Ecological Classroom of Foreign Languages
	10:50-11:20	ZHU Qiyun Chinese Students' Perceptions Of Innovative Assessment Tasks In An EFL Writing Curriculum
	11:20-11:50	邵争、杨茂霞 PBL 课堂评估中模糊评分的应用
<u>Room 317</u> <u>THEME</u> Formative Assessment	08:30-09:00	黄剑、罗少茜 形成性评价概念重构视角下的教师评价能力标准研究
	09:00-09:30	赵中宝 An Investigation into the Impact of Formative Assessment on College Oral English Teaching and Learning
	09:30-10:00	王伟强 Understanding Students' Formative Self-Assessment Practices in the Chinese EFL Writing Class: Two Case Studies
	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	刘洋、李英春 A Study of Education Students' Perceptions of Formative Assessment
Chair 黄剑	10:50-11:20	Tatiana Ershova Language teachers' target language skills: assessing students' writing
	11:20-11:50	Maggie Ma

		Using formative assessment to facilitate self-regulated learning: assessment practices and student perceptions in an EFL writing classroom
	11:50-12:20	Nguyen Thi Minh Loan Implementing and evaluating the use of e-portfolios in FL instruction
<u>Room 318</u> <u>THEME</u> Test tasks and test development Chair QIU Xuyan	08:30-09:00	Muhiddin Timuçin Denктаş Let's roll up our sleeves writing out our own rubric for oral assessment
	09:00-09:30	QIU Xuyan Taking into account topic familiarity in assessment design
	09:30-10:00	LI Aike, Brechtje Post L2 Development of Speech Rhythm Learners of English Pronunciation Assessment
	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	LIU Li, JIA Guodong The role of speech content in the assessment of candidate oral performance
	10:50-11:20	LV Jiantao Developing a Computer-aided Speaking, Listening and Writing English Test at a University Level: Validity and Reliability Evidence from Response Data
	11:20-11:50	林椿、肖云南 两种评分方法的评分效应比较研究
	11:50-12:20	童扬芬 Language Test Fairness: Conception and Framework
<u>Room 422</u> <u>THEME</u>	08:30-09:00	LIU Liandi Interactional patterns of Chinese EFL Learners in the Paired Speaking Test
	09:00-09:30	CAI Yuyang Using the Trait–State–Occasion Model to examine changes in graduate students' metacognitive strategic competence development

Focusing on test takers Chair 张聪	09:30-10:00	张剑、罗柳萍 The Vocabulary Use of Chinese Candidates of Different Levels in Oral English Examinations
	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	张聪 Multimodal Analysis of Interviewers' Pragmatic Identity in English Oral Testing
	10:50-11:20	YANG Ji, GU Xiangdong, LIANG Yansong An Introspective Study on Test-takers' Process of Picture Composition in the New HSK-5
	11:20-11:50	LI Xueping A closer look at the cognitive processes of NMET reading
Room 423 THEME Diagnostic Testing Chair XIE Qin	08:30-09:00	孟亚茹、马晓梅 EFL Listening Diagnostic Assessment Based on Cognitive Diagnostic Approaches
	09:00-09:30	ZHAO Guanfang, FAN Jinsong Diagnosing EFL writing difficulties in Chinese secondary schools
	09:30-10:00	Carolyn Westbrook, Yvonne Merry, Alexey Korenev, Tanya Ershova Formative assessment of language teachers' ESP: developing a scale for diagnostic assessment of teachers' classroom language proficiency
	10:00-10:20	Tea break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	OYEDOKUN-ALLI, WASIU ADEMOLA A Diagnostic and Prognostic Approach to Error Analysis of Ab(use) of Modifiers among Yoruba Bilinguals
	10:50-11:20	XIE Qin Diagnostic assessment and tracking of English academic writing of Hong Kong university students
	11:20-11:50	范婷婷 Construction and Validation of Diagnostic Proficiency Scales for Chinese EFL Readers in

		High School
Room 714	08:30-09:00	CHEN Zhuo, LIN Yi Evaluation Form For College English Oral Presentations
	09:00-09:30	Reima AL-Jarf Issues in Assessing the Speaking Skill in EFL
	09:30-10:00	ZU Daqing 浸入式教学理念下外教口语课教学模式改革与英语专业体验式教习
	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)
	10:20-10:50	LI Zhiying 交际性语言测试对高中英语教学的引导作用研究
	10:50-11:20	LUO Sha Effects of Peer Assessment on EFL Writing in MOOCs: The Learners' Perceptions
	11:20-11:50	杨雅平 Pragmatic Competence and Lessons for University English Teaching, Language Testing and Assessment
THEME Language testing and teaching		
Chair LUO Sha		
Room 715	08:30-09:00	YANG Manzhen, LU Jiantao, ZHU Yuanyuan, ZHANG Decheng, LIANG Li Implementing Assessment for Learning through an Internet Forum:Feedback from both EFL Learners and Tertiary English Teachers
	09:00-09:30	LIANG Li, YANG Manzhen, LU Jiantao, ZHU Yuanyuan, ZHANG Decheng Recognizing and measuring self-regulated EFL learning within an online formative assessment module
	09:30-10:00	LI Song, HE Cencheng An Empirical Research on Relevance of Subjective and Objective Questions in English IBT: A Case Study of IBT at Beijing University of Technology
THEME Information Technology In Language Assessment		
Chair HUANG Liyan		

	10:00-10:20	Tea Break (Lobby of Teaching Building No.7 第七教学楼一楼大厅)	
	10:20-10:50	HUANG Liyan, CHENG Fenxiang Designing and Validating the Test Tasks of a Computer-based Speaking-Listening Test	
	10:50-11:20	XU Ying, ZENG Yongqiang Using Generalizability Theory (GT) and Many-facet Rasch Model (MFRM) to investigate Computerized English Listening and Speaking Test (CELST) rating	
	11:20-11:50	肖云南、林椿 Methods of Scaling and Equating in Constructing Item Bank of Computer-based English Placement Test	
	11:50-12:20	CAO Linlin Automatic and human rating differences in CELST in Guangdong NMET: A many-facet Rasch analysis	
12:20-14:00	Lunch (First Floor, Poly Hotel, 保利山庄一层)		
14:00-14:40	Keynote Speech (4) Chair Prof. GU Xiangdong	Prof. CHENG Liying Clarifying Classroom Assessment Purposes and Uses through Classroom Assessment Tasks and Environment	图书馆负一层报告厅 Lecture Hall, Lower Ground Floor, GDUFS Library
14:40-15:20	Keynote Speech (5) Prof. WU Zunmin	Prof. John de Jong Developing scaled language learning objectives	
15:20-15:40	Tea Break (图书馆负一层大厅)		

15:40-16:40	Invited Forum Chair Prof. HE Lianzhen	Prof. GUI Shichun Rethinking English Education in China: Complexities and Perplexities	图书馆负一层报告厅
16:40-17:00	Closing Ceremony	Speech by Dr. YANG Jing (Vice Director of the National Key Research Center for Linguistics & Applied Linguistics at GDUFS)	Lecture Hall, Lower Ground Floor, GDUFS Library

Abstracts of Keynote Speeches

Keynote Speech 1

Developing Scales for Chinese EFL Learners

Prof. LIU Jianda	Guangdong University of Foreign Studies
Prof. WU Zunmin	Beijing Normal University
Prof. HE Lianzhen	Zhejiang University
Prof. JIN Yan	Shanghai Jiaotong University
Prof. ZENG Yongqiang	Guangdong Teachers College of Foreign Language and Arts
Prof. ZOU Shen	Shanghai International Studies University
Prof. HAN Baocheng	Beijing Foreign Studies University
Prof. YAN Ming	Heilongjiang University
Prof. MU Lei	Guangdong University of Foreign Studies

In September 2014, the China State Council issued the document “The Implementation Opinions of the State Council on Deepening the Reform of the Examination and Enrollment System”, stipulating the “construction of the assessment system of foreign language proficiency”. The system includes several major parts, among which developing scales for Chinese EFL learners is the most fundamental one. The scales will describe what Chinese learners and users of English can do and to a certain degree prescribe what they should be able to do with the target language. This talk will introduce the why, what, when, and how of the construction.

Biographic Sketch of Prof. LIU Jianda:

LIU Jianda is professor of the National Key Research Center for Linguistics and Applied Linguistics at Guangdong University of Foreign Studies, Guangzhou (China). He received his PhD in Linguistics and Applied Linguistics from the City University of Hong Kong. He has worked in designing, developing, and validating major language tests in China. His research interests include language testing, pragmatics, and second language acquisition. He has published many papers in different academic journals. His recent book entitled *Measuring Interlanguage Pragmatic Knowledge of EFL Learners* was published by Peter Lang. He is now in charge of the development of the China Scales of English Language Competence, a project sponsored by the Ministry of Education of PRC.

Keynote Speech 2

Developing a learning ladder to set language standards

Prof. Nick Saville Cambridge English Senior Management Team

This talk addresses the concept of Learning Oriented Assessment and makes the case for introducing a common framework of reference within educational reforms as a way of setting learning goals and assessing proficiency outcomes more effectively.

The challenge for educators is to ensure that language education systems equip young people with the functional communication skills that employers need in the workplace and that the learners themselves find useful and fulfilling in their private lives. Languages are essential ‘life skills’, not just academic subjects to be learned in school and then forgotten. This means there are “real world” demands that educationalists must respond to in order to set and raise proficiency standards over time – and this is particularly true of English which has evolved into a lingua franca on a global scale.

In general, high stakes assessments are often used for accountability purposes - a key concern for policy makers who use the results from tests to hold stakeholders to account: governments use them to evaluate schools; schools use them to evaluate teachers, and so on. High stakes language tests in this context also provide scores or certificates for making important decisions, e.g. the cut score for university entrance. Unfortunately the impact of this approach is often negative in that too much effort is put into achieving the required score or in passing the test and not enough into developing the important functional skills.

In this talk I suggest that assessment systems must play a more positive role, and that our focus should be on linking school learning with wider society more effectively. The challenge is to provide a systemic model which links all types and levels of assessment into a coherent model to improve both the learning processes and the learning outcomes. I will argue that to do this we need to adopt a learning-oriented approach that locates learning at the heart of all assessment contexts - from classroom tasks to progress tests, to high stakes exams, such as the gaokao. Teaching, learning and assessment must be aligned to the same set of goals, and thus to each other so that we can ensure that the large-scale, external examinations impact positively on learning and on the educational systems where they are used.

Large-scale language assessment is one dimension of this learning-oriented landscape: it describes progression on a vertical axis, through lower to higher levels of proficiency - and is quantitative, in that it is based on measurement principles. This contrasts with a second dimension – a horizontal one – which groups learners who are all at the same global proficiency level and explores how they differ from each other, for example in terms of their

individual cognition, their experience, and their learning needs. That is, it relates to qualitative differences between learners that is typically the domain of the classroom and the teacher.

In summary, we need a develop a widely understood interpretive framework – one that enables us to reference common standards that embody a learning ladder to ensure all types of assessment focus on the same goals, and report performance in terms of the same underlying scale. By treating language ability as a skill with social and professional value, we can provide a clear basis for aligning teaching and assessment to desired outcomes - and this can be used for understanding and measuring progression, metaphorically speaking, ‘up the learning ladder’.

Biographic Sketch of Prof. Nick Saville:

Nick Saville is a member of the Cambridge English Senior Management Team and is responsible for directing the work of the Research and Validation Group. He holds a PhD from the University of Bedfordshire in language test impact, as well as degrees in Linguistics and in TEFL from the University of Reading. He has specialised in language assessment since 1987. Before joining Cambridge in 1989, he taught at the University of Cagliari, Italy, and managed test development projects in Tokyo.

Keynote Speech 3

Testing Teachers: issues and solutions

Prof. Barry O’Sullivan British Council

For many years now, there has been a movement on behalf of a host of national education authorities to come to grips with a critical issue in language teaching. This issue is the language ability of the teaching cadre. Essentially, there are two questions to be answered. The first of these is: “what level of language does a teacher need in order to be able to teach that language to a specific grade or school?” The second is: “what level of language do our language teachers currently have?” With the rise of English as a medium of instruction (EMI) the questions are broadened to include a range of subjects, but the core remains the same, what level is required, where are they now? The answer to these questions are seen by education ministries as vital to their ability to make systematic medium and long term improvements to their systems.

There are a number of critical issues involved in establishing evidence of teachers’ language ability. From a ministry’ perspective the main consideration tends to be cost, as high quality internationally recognised tests are typically very expensive and not always suitable. On the teachers’ side, there is the question of trust. How will the test scores be used? Is there

a possibility that my career is in jeopardy? Should I be judged using a test that may not be fully appropriate for the purpose?

In this talk, I will present these issues, discussing them in relation to how they have been addressed in the past. Since the issue of test use is critical, I will focus on how the British Council has worked with ministries and teacher groups to ensure that the tests used are seen as being fully appropriate culturally and linguistically.

Biographic Sketch of Prof. Barry O’Sullivan:

Professor Barry O’Sullivan is currently working with the British Council in London as Head of Assessment Research & Development. His recent work includes the design, development and validation of a placement test to be used by the British in their centres across the world and the design, development and validation of a new business to business language test called Aptis.

Keynote Speech 4

Clarifying Classroom Assessment Purposes and Uses through Classroom

Assessment Tasks and Environment

Prof. CHENG Liying Queen's University

Undoubtedly, most of the information that students have about their learning, about themselves, and about their futures comes from classroom assessment. Similarly, most of what parents and teachers know about their children's learning comes from classroom assessment. It is through the day-to-day classroom assessment tasks and the environment teachers and students co-create that important decisions (purposes and uses) are understood, communicated, and reported. The ways teachers communicate their expectations to students, and the ways they provide feedback on how well these expectations are being met, help students form the concepts of what is important to learn and how good they are at learning it. Current debates about quality of classroom assessment continue to use validity and reliability arguments developed for large-scale testing. This plenary highlights the context dependence of classroom assessment in relation to large-scale testing, and discusses the intricate relationship between assessment and instruction through assessment tasks and environment in supporting student learning.

Biographic Sketch of Prof. CHENG Liying:

CHENG Liying is Professor and Director of the Assessment and Evaluation Group (AEG) at the Faculty of Education, Queen’s University. Her primary research interests are the impact of large-scale testing on instruction, the relationships between assessment and

instruction, and the academic and professional acculturation of international and new immigrant students, workers, and professionals to Canada. Since 2000, she has obtained research funding of more than one million Canadian dollars. In addition, she has conducted 150 conference presentations and has more than 100 publications in top-tier journals including Language Testing, Language Assessment Quarterly, Assessment in Education etc.

Keynote Speech 5

Developing scaled language learning objectives

Prof. John De Jong VU University Amsterdam & Pearson

The Common European Framework of Reference for Languages (henceforward CEFR) has provided a system for benchmarking learner progress that has become widely accepted in Europe and beyond; however, the real meaning of the levels and the limitations of the framework as originally set out have been widely ignored or misinterpreted. In addition, the original set of descriptors is very sparse, apart from those related to speaking in the upper middle area of the scale, and this limitation severely restricts its applicability in many situations.

This paper describes an ongoing research project at Pearson to develop the ‘Global Scale of English Learning Objectives’. This is a set of ‘Can Do’ descriptors covering the full range of English language skills from below A1 (no language ability) to C2 (full and fluent language ability), and is intended as a basis for standardizing the development of instructional and testing materials, and for learning, teaching, and assessment practice in general. It is inspired by and built on the basis of the CEFR published by the Council of Europe in 2001.

Pearson have considerably extended the range of descriptors, and have created a new scale to measure English language proficiency, the Global Scale of English, which is directly correlated to the CEFR but allows the observation and reporting of much smaller increments in English language proficiency.

This paper first explains the basic principles underlying the CEFR and Pearson’s development of the Global Scale of English and the Learning Objectives, also showing why this work was felt to be necessary. We go on to present the methodology for creating descriptors and for editing them to ensure they are precise, succinct and clear. We explain how the descriptors were scaled, with the help of many hundreds of teachers and ELT experts worldwide, and lastly, we briefly outline some of the intended applications for the Global Scale of English and Learning Objectives.

Biographic Sketch of Prof. John De Jong:

John De Jong is Senior Vice President Global Assessment Standards with Pearson and Professor of Language Testing at VU University Amsterdam. John obtained a Ph.D. in Educational Measurement from Twente University. He has specialised in empirical scaling and the development of internationally standardised reporting scales of language proficiency. He was involved in developing the Common European Framework for Languages.

Invited Forum

Rethinking English Education in China: Complexities and Perplexities

Prof. GUI Shichun Guangdong University of Foreign Studies

The debate around the competence/performance dichotomy has led to the growth of performance theory, language as code/signal, complex adaptive system and the usage-based theories which have significant contribution to the study of language teaching as well as language changes. The spread of English as a Lingua Franca triggers a number of problems for us to resolve, such as the differences of native speakers vs. nonnative speakers, the repositioning of English education in China, the characteristics of interlanguage (Basic Variety), ultimate attainment, use-based theories of classroom teaching, the understanding of language education by Chinese teachers of English, the social cognition and public opinions of English education, etc. This talk aims to draw the attention and discussion among education executives as well as the general public of Chinese teachers of English, rather than just inserting some ready-made conclusions; so that transformation of English education in China can be well found.

Professor GuiShichun, a well-known figure in the academic field in China, is credited with being the first scholar who introduced applied linguistics studies into China. He has established several other records of "the first in China," hence has long been called "the brave pioneering explorer" in his field in his country. For example, he established the first applied linguistics (AL) MA program in 1978, the first AL certificate program in 1980, and the first AL Ph.D. program in 1986. He also established and held the first national symposium on applied linguistics and English teaching in 1980 and the first international symposium on foreign language teaching in China in 1985. Further, in 1978, he also started advocating the application of international mainstream language testing theories and established the nation-wide English Proficiency Test (EPT) in China. Generally regarded as the founding father of language testing in China, he is also the advisor of the National College English Test committee.

Abstracts of Parallel Sessions

Saturday Sessions

Room 301

Developing an English speaking scale for teaching, learning and assessment in

China: Principles, procedures and progress

上海交通大学 金艳

上海交通大学/上海对外贸易学院 揭薇

With the extensive use of the Common European Framework of Reference for Languages (Council of Europe, 1991) in Europe and other parts of the world in recent decades, the importance of a language scale for describing and scaling language proficiency in a scientific, uniformed and principled way is being increasingly recognized. Such a scale is considered of urgent necessity in China in order to improve its English language education and bring it in line with international standards. In this presentation, we will report on the principles we have adopted, the procedures we are following and the progress we have made in the development and validation of the National English Scale—Speaking Scale (NES-SS).

A critical review of current English education policy and practice in China was conducted, with particular reference to oral English teaching and assessment, based on which we proposed the basic principles to be followed in the development of the NES-SS. First, the establishment of performance levels should take the context of oral English teaching and assessment in China into full consideration. Second, level descriptors should be informative, accessible and understandable to both researchers and practitioners of English language education in China. Third, special attention needs to be paid to the salient and observable features of Chinese learners' oral English proficiency. Based on the principles and after a detailed comparison of various approaches and methodologies in developing and validating language scales, we established the procedures for developing and validating the NES-SS. As the first step, we defined the construct of oral English proficiency based on a review of relevant theories and empirical studies. The outcome of the work at this stage was a framework for describing the oral English proficiency of Chinese learners. As a second step, we collected about 4000 existing descriptors of oral English proficiency, revised most of them, and re-wrote some new descriptors. Workshops and surveys were conducted to discard descriptors of poor quality, and improve the clarity and meaningfulness of the remaining

descriptors. The levels and categories assigned to each descriptor were also investigated qualitatively at this stage. The outcome of the work at the second stage is a descriptor pool with 794 descriptors of oral English proficiency. The third step towards establishing the NES-SS is to validate the levels and categories of each descriptor through quantitative analyses. The outcome will be a speaking scale carefully calibrated against a larger dataset as well as other sub-scales of the NES.

Language test validation is an iterative process. Similarly, the validity of a language scale needs to be continuously investigated among various groups of stakeholders, including for example, language policymakers, educators, teachers and assessors. It is therefore hoped that both human and financial resources will be made available to validate the speaking scale on an on-going basis after it is put into large-scale use.

解读英语口语能力描述语：基于口语量表描述语库的研究

上海交通大学/上海对外贸易学院 揭薇

One of the major tasks to develop an English Speaking scale is to use the "Can-do" descriptors to portray oral English proficiency. A quantitative analysis of descriptors is conducted in this study so as to provide research evidence for English Speaking scale development and validation. By collecting about 4,000 existing descriptors of oral English proficiency, to establish the initial descriptor pool and using corpus software and statistical software for word frequency and key word analysis. These studies looked into details the collocation of words, annotated, checked descriptor key words based on a review of semantic analysis theory, interpreted and explained the features of oral English proficiency descriptors, in order to have a systemic, in-depth analysis and research upon proficiency descriptors therefore link up the descriptor analysis from lexical level to the sentence level. The study attempts to make explicitly the syntactic structure of descriptors, and expatiate on the semantic feature of proficiency descriptors, which makes users further understand descriptors, grasps the collocation rules and principles among components, and thus provide evidence and guidance for the writing and revision of oral English proficiency descriptors.

Developing Operational Framework and Descriptors of Pragmatic Effectiveness

北京物资学院外语学院 王淑花

Pragmatic effectiveness is an intricate subject in the field of language proficiency and pragmatics. Hymes did not talk about pragmatic effectiveness directly, nor did other researchers of language proficiency. But in his communicative competence model Hymes defined “sociolinguistic competence” as the rules for using language appropriately in context (appropriateness). Bachman & Palmer defined pragmatic competence as knowing function knowledge and sociolinguistic knowledge. Purpura argues that the determination of what is meaningful or pragmatically appropriate, acceptable, natural or conventional depends on the underlying contextual, sociocultural, sociolinguistic, psychological or rhetorical norms, assumptions, expectations and presuppositions of the interlocutors in a given situation. Some pragmatic researchers consider appropriateness as the maxim below politeness, while still others take it as the supreme maxim of pragmatics.

This paper defines pragmatic effectiveness on the basis of the researches mentioned above, that is, pragmatic effectiveness means that to achieve a certain communicative function and the ideal communication goal, an language user expresses himself appropriately in the linguistic, situational, and sociocultural context. To be specific, he might be able to

- 1) obey the conversational rules and the politeness maxim, use the language to maintain social contact with his interlocutor according to the others' attributes;
- 2) use the knowledge of register appropriately and use either the formal or informal language appropriately according to the situation;
- 3) use the knowledge of genre appropriately to express his meanings;
- 4) use the knowledge of dialects appropriately to achieve his communicative purpose;
- 5) use the knowledge of idioms appropriately to communicate his implied meanings;

These five aspects may not develop simultaneously. The beginners may achieve his pragmatic effectiveness by obeying the conversational rules and politeness principle, while the higher language achievers may use more functional knowledge and sociolinguistic knowledge; use them more flexibly as well.

Based on this operational framework, the researcher will determine the key words to collect descriptors from the existing proficiency scales and academic papers. On the basis of the relative researches on writing and speaking, the factors that determine the levels of descriptors will be explored. Finally the descriptors will be made into questionnaires after being edited. The scale of pragmatic effectiveness will be established after the verification among teachers and learners.

多维度大学生英语写作能力标准构建¹

大连外国语大学 泰中华

上海理工大学 刘 芹

This research attempts a preliminary construction of the prototype criteria of multi-dimensional college English writing ability through analyzing relevant theoretical and empirical literature reviews. The criteria are theoretically based on textual linguistics, cognitive constructs of writing, communicative language ability theory, and ESP ability theory. Their empirical bases focus on writing tasks of major English tests both at home and abroad, coupled with Weigle's academic English writing framework. Through establishing and analyzing the interactional relations of writing text, contexts and writers, this research has constructed a writing ability criterion comprising a tri-dimension of textual construction, textual production and individual mechanism. Textual construction, based on Flower and Hayes' cognitive writing theory, includes three aspects of concept generating, concept organization, and goal setting, covering seven traits; textual production refers to major writing ability standards and relevant research findings both at home and abroad, and covers 55 traits; individual mechanism stems from Bachman's communicative language ability theory, inclusive of physical fitness and psychological fitness. The empirical part of this research is a questionnaire survey with 60 English writing ability traits as its content, covering 1494 students and 379 English teachers from 12 colleges in China. The findings fit well with the protocol model hypothesis. The creativeness of this research lies in the claim that the writing ability exhibits itself more throughout the whole writing activity including the textual construction and the writer's psychological and physical status than just in the writing text. A multi-factor and multi-dimension assessment is contributive to an all-round and objective evaluation of one's writing ability development, which provides a comprehensive and reliable feedback on English writing teaching. The model in question adopts a combination of formative assessment and summative assessment.

大学英语写作能力标准的制定与实施——以中国人民大学为例

中国人民大学外国语学院 贾国栋 王晓彤

英语是当前国际交流的主要通用语言，英语写作是国际交流过程中不可或缺的能力之一。随着我国更多地融入国际交流进程，培养大学生具有较强的英语写作能力已成为学校英语教学中的重要任务。因而，学校培养学生掌握英语写作能力具有现实性与长远性，是学校发展的重要组成部分。就学生个人发展而言，具备良好的英语写作能力一方面有利于个人能力地全面提升。

《中国人民大学英语写作能力标准》（简称《写作标准》）就是在这种背景下完成的。《写作标准》是参照《大学英语教学指南》和中国人民大学《本科人才培养改革路线图》制定的，目的是全面提高学生的英语基础写作和学术写作的能力。

《写作标准》对照《大学英语教学指南》中对写作能力培养的基础、提高和发展三个目标和中国人民大学本科生英语写作能力的实际情况设定为“英语写作（一）”、“英语写作（二）”和“学术英语写作”三个目标。《写作标准》的能力等级分为1到6级，与《欧洲语言共同参考框架》的A1至C2六个等级相对接。《写作标准》的能力描述从“体裁与风格”、“组织与发展”和“语言与文化”三个维度进行，每个维度都有由1级最初步的写作要求到6级最高级的要求所组成。

为有效检测学习效果，《写作标准》设计了“写作水平测试”版块，从测试内容、分项、形式、时间、对象、成绩转换与公布等方面进行了规定以保证测试的科学性。《写作标准》将对学校大学英语教学起到积极的推动作用，促使教学更加注重学生写作能力的培养。

Development of Scales of Oral Pragmatic Competence

广东外语外贸大学 黄玮莹

Pragmatic competence has been recognized as an indispensable part of Communicative competence. However, no scales of pragmatic competence have been developed and little attention is paid to pragmatic competence in scales of oral proficiency. Therefore, the current study aims to develop a scale of pragmatic competence, based on research in Interlanguage Pragmatics and speech acts theory. This paper presents the construct of the scale, the methodology applied and a pilot study of the scaling of descriptors.

Room 302

Validity Issues of Second Language Dynamic Assessment

南方医科大学 李清华

Dynamic assessment (DA) has found its way into second and foreign language (L2) assessment in recent years. It is argued that DA can be a useful tool to diagnose and promote L2 development, yet little attention has been paid to DA aimed at measuring L2 competence development. Validity issues L2 DA have yet been concerned by few researchers in language assessment circle. Therefore, this paper attempts to propose a tentative validity model of L2 DA based on the comments on the validity of classroom-based formative assessment and micro and macro validation models of L2 DA. In this model, the cornerstone issues of validity theory of L2 DA involve construct, consequence and fairness. Five essential questions to pursue in L2 DA validity are clarified. Potential sources of evidence include learner development, mediation, and learner responsiveness in the process of DA. Moreover, an argument-based approach is recommended to be applied in L2 DA validation.

Building a Validity Argument for a School-Based English Test: Evidence from Construct and Predicative Validity

Fudan University FAN Jnsong ZHENG Yongyan
Shanghai International Studies University ZHAO Guanfang

Recent developments in validation theory require that multiple strands of validity evidence be collected, weighed, and synthesized into an evaluative judgment or validity argument in support of test score interpretation and use (e.g., AERA, APA, & NCME, 2014; Kane, 2001, 2004). Drawing upon Chapelle et al.'s (2008) model of interpretive argument, this study was aimed at building a validity argument for a high-stakes school-based English proficiency test, focusing on two inferences in the model, i.e. explanation and extrapolation. To address the two inferences in the interpretive argument, the following two research questions were investigated: 1) What is the factor structure of this English language test? Is

this factor structure consistent with the relevant theories of language ability? 2) To what extent can students' test scores predict their performance in target language use (TLU) domains?

To address the first research question, confirmatory factor analysis (CFA) in Structural Equation Modeling was performed whereby five competing models were posited to represent the factor structure of the English test. These five models were then tested against the test data (N=4231) and compared with each other in terms of model-data fit, parameter estimates, and model parsimony. Results revealed that the higher-order factor model best fit the test data (S-B $\chi^2=763.43$, $df=73$, $p=0.00$, CFI=0.959, RMSEA=0.053). To address the second research question, a self-assessment (SA) questionnaire was developed and validated using the Rasch measurement model, which was intended to reflect test takers' language ability in TLU domains (Fan & Zhao, 2015, forthcoming). Data were collected from 250 students who had already taken the English test. The best fitting model, i.e. the higher-order factor model was imposed on the test data of the 250 students who completed the SA questionnaires. Results indicated that the model fit the sample data well. Then, CFA was applied to modeling the SA questionnaire data. In this analysis, four competing models were posited to represent the factor structure of the SA data, among which the higher-order factor model was found to demonstrate the best model-data fit (S-B $\chi^2=634.27$, $df=294$, $p=0.00$, CFI=0.923, RMSEA=0.068). Finally, structural regression modeling was utilized whereby the higher-order factor in the test data model was regressed on the higher-order factor in the SA data model. Results revealed that the structural regression model fit the data well, and the regression coefficient was 0.52, suggesting that students' test scores could predict their performance in TLU domains reasonably well.

Though focusing on only two inferences (i.e. explanation and extrapolation) in the interpretive argument, this research lent crucial support to the validity of this English language test. To portray a coherent and convincing validity argument narrative for this test, however, much more research is warranted. Future research is necessary to address the other inferences in the model of interpretive argument such as evaluation, generalization, and utility. In addition, more research is necessary to address the potential rebuttals which may weaken the claims made of test score interpretation and use.

基于 Coh-Metrix 的大学英语六级与托福阅读语篇难度研究

对外经济贸易大学英语学院 江进林

This study presents an analysis of the difficulty of CET6 and TOEFL reading texts with Coh-Metrix tool based on about 100 texts. The results show that Coh-Metrix can analyze text difficulty along eight functional dimensions (narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion, verb cohesion, connectivity, and temporality) through calculation of 53 lexico-grammatical and semantic features. It is found that the reading texts of CET6 are easier than TOEFL, and they have significant difference along four dimensions. CET6 reading texts exhibit higher narrativity (topics being more familiar and texts more oral) and connectivity (more explicit additive connectives, adversative and contrastive connectives, and comparative connectives), while lower word concreteness (words being more abstract) and referential cohesion (less word overlap) than TOEFL. The study prepares a relatively objective descriptive frame of reference for test development, reading material grading and text selection for students.

学术英语大规模读写测试的效度研究

北京工业大学 牟欣

Results of a test, especially high-stakes tasks are important tools to allocate social resources and improve fairness. Adopting “Educational Test Fairness Disciplines” published by Test Behavior Federal Committee and research result of Messick concerning evidential aspect. Kunnan(2000) built a theoretical frame including validity, access and justice. His theory is divided into five levels and validity is regarded the most important element among them.

Based on Rasch Model. This research aims at investigating the variance of final test score by BJUT 2013 and 2014 post-graduate students, and the validity of the test. Relative analysis includes structure, content and format; furthermore, whether some students are discriminated because of them.

The results proves the test is efficient in distinguish students ‘level but there are many problems existing.

On the Validity and Discrimination of Cet-4 after Reform

--- Cet-4 performance of ** University in Dec. 2007 as an Example

Shanghai International Studies University CHEN Yangfang

Pearson Correlation Analysis of cet-4 performance of **University in December 2007 show: Reading scores is the most related to the total scores of cet-4, followed by listening, writing ,and comprehensive competence scores. This does not agree with the original reform intention of testing the comprehensive performance. Frequency statistic of cet-4 performance show: due to the use of standardized instruments, of 3060 examinees, total scores fall on 203 points, with at most 47 total scores (15.7 on average) fall on one point; listening scores fall on 57 points, with at most 192 listening scores (53.68 on average) fall on one point; reading scores fall on 57 points, with at most 133 reading scores (53.68 on average) fall on one point; writing scores fall on 31 points, with at most 309 writing scores (98.71 on average) fall on one point; comprehensive performance fall on 19 points, with at most 471 comprehensive performance scores (161.05 on average) fall on one point. For such concentrated degree, there is room to be improved in the discrimination of cet-4 after reform.

Mixed-method Research in Language Testing: History, Practice and Trend

浙江大学 施雅俐

With a view to elucidating the history, practice and trend of MMR in Language Testing, the review conducts an analysis of 63 MMR articles collected through searching by key words within the scope of all the empirical publications in the two premier journals Language Testing and Language Assessment Quarterly from seven dimensions respectively as follows: year, theme, research purpose, research question, research model, integration manner and research tools. Results show that MMR studies in Language Testing: 1) were mainly published during 2010s, and the increase has doubled per decade from 1980s to 2010s; 2) have mainly discussed such themes as scoring and validation, and mainly investigated such language skills as speaking and writing; 3) have the purposes dominated by Triangulation 48% and Complementarity 33%, followed respectively by Expansion 14% and Development 5%; 4) have the purposes dominated by “each question per each phase” 59%, followed respectively

by “general MMR question” 25% and “quantitative question +qualitative question + MMR question” 13%; 5) have the model dominated by concurrent design 49% and sequential design 49%, and followed by complex design 2%; 6) have the integration manner dominated by “integration in interpretation” 73%, followed respectively by “integration in both analysis and interpretation” 16%, and “no integration” 1%; 7) have such quantitative tools as Rasch, ANOVA, Correlation, Regression and Factor analysis; and have such qualitative tools as questionnaire, interview, think-aloud, reflection and comments. Based on the above results, five implications are proposed.

A Study of NMET Writing Test Items of Guangdong

Province from 2007 to 2014

华南师范大学 张妙

This paper carries out a diachronical study on NMET writing test items of Guangdong Province from 2007 to 2014 to examine the content validity of these items. The findings reveals that the content validity is relatively high for the writing test has basically met the requirements of the course syllabus and the test specifications, except that the form of input and the principle of interactiveness are still expected to adjust to the course syllabus. Based on these findings, this paper puts forward some suggestions for future improvement in the design of NMET writing test items of Guangdong Province and for English writing teaching in senior high school English classrooms.

Room 311

50 Years of High-stakes Exams in China—A Personal Journey

Chongqing University GU Xiangdong

In this talk, I will use my personal anecdotes to show how high-stakes exams, including English exams could have far-reaching impact on individuals' lives. My talk will focus on snapshots of my own test-taking experiences of high-stakes exams from the entrance exams for junior and senior middle schools, through Gaokao of two times, to the entrance exams for

MA and PhD programs. My story will contextualize the huge changes in education and assessment in China over the last 50 years. It might be a very good example to help people at home and abroad to understand the role and status of high-stakes exams in the social, political, cultural and educational transitions in China over the last 50 years. I will illustrate how far we have come and suggest what the future may hold for the necessity of standards, transparency, accountability and fairness of high-stakes exams in China.

An Analysis of the Problems of ‘Scientific Design’ of English Test Contents of College Entrance Examination

西华师范大学 陈文存

This paper argues that a scientific design of the English test contents of college entrance examination should be based on the legal basis, reflect the nature of language and language learning and follow the elementary principles of language testing. Based on the three aspects, the paper analyses the English test contents of the current college entrance examination. The author claims that the contents of test papers should be based on the English capacity demand of college studies and the national curriculum standard. However, there exist some problems in the paper contents on both aspects, esp. the test can't meet different demands from colleges of different levels. Generally, the current paper can reflect the nature of language and language learning. Great progress has been made in following the principles of language testing, but some deficiencies can be found in testing thinking ability in English, interaction and authenticity of testing tasks. There should be more contents for testing candidates' critical thinking. The paper highlights that the scientifically designed test contents should only let the candidates who learn English according to the laws of foreign language learning and have acquired remarkable language ability get high scores, while the test-oriented learning way can't lead to high scores.

Writing prompt convergence in high-stakes tests: Exploring alternative rhetorical functions and objects of enquiry

江西师范大学 刘辅兰

The writing prompts in high-stakes tests have become an issue of concern due to their possible effect on test takers' writing performance and potential washback on writing pedagogy. In the present study, the prompts (n=120) from three of these tests, TOEFL, IELTS

and TEM4 (a language test in China), were investigated for two elements, their specific rhetorical function and object of enquiry, to examine whether these prompts displayed certain patterns. Results revealed that both of these elements converged around a narrow set. For prompts which requires a standard essay as responses (hereafter essay prompts), evaluation (44%) and hortation (23.3%) were the most common rhetorical functions among the 60 prompts. And three sociocultural themes: education, technology and current social issues. dominated the objects of enquiry. For prompts which require responses other than a standard essay (hereafter non-essay prompts), the confinement in the ranges of the 60 prompts is also manifest. These findings may indicate that the prompts in high-stakes tests are converging around a restricted set of themes which may be generating prescriptive responses that fail to appreciate the full range of linguistic and argumentative possibilities normally expected in academic writing. Here, we propose that a wider range of writing prompts with alternative rhetorical functions and objects of enquiry be considered in high-stakes tests in order to facilitate students' writing and thinking development.

Consensus and divergence: Voices of test-takers and raters under a high-stakes language test

浙江财经大学外国语学院 徐莎莎

It is against the backdrop of higher education system in China that the present study is carried out, aiming to better understand different stakeholders' reactions to the writing tasks of a high-stakes English test - Graduate School Entrance English Examination (GSEEE). The present study elaborates on the revision of the GSEEE test format and task design in 2005 and the writing tasks in particular, which are expected by policymakers and test designers to facilitate positive washback on learning and teaching of English. The present study places the test-driven preparation practices towards the GSEEE writing tasks at the center, and offers valuable insights into the interaction among perceptions, test-driven preparation practices and products. Test preparation, as an overt dimension of washback on learning, are scrutinized from perspectives of two direct stakeholders (test-takers and raters). A mixed method design was employed, using questionnaire and semi-structured interview to collect data from 623 test-takers and 84 raters from a GSEEE marking center. Test-takers' perception of task demand and corresponding preparation practices were collected, as well as raters' opinion of the effectiveness and appropriateness of preparation strategies and activities. In terms of specific patterns of preparation practices for the GSEEE writing tasks, five factors emerged and were identified as memorizing practice, test familiarization, comprehensive learning,

skills development and drilling practice. Mann-Whitney U test and thematic content analysis of in-depth interview indicated that the consensus and divergence co-existed on the perceived usefulness of preparation activities between test-takers and raters. Furthermore, regression analysis by linking test scores demonstrated that these divergences had the potential to influence the effects of test preparation and alter direction of the intended effects of test revision. The research findings of the present study free the GSEEE test-takers from the mire of mechanical memorization practice, and substitute mechanical memorization for test familiarization and drilling practice, which are proven to be more appropriate in raters' view and effective in test-takers' interests. The present study demonstrates the importance of co-operation and communication among different stakeholders. In order to achieve the intended washback effect, a closer fit should be obtained between how test designers intend for their tests to bring about and how test-takers actually perceive and prepare for them. It is hoped that the detailed portrayal of test-takers' preparation experience and the raters' responses can help decision makers and test developers at central agencies to reflect on the nature and the impact of the GSEEE and other high-stakes language tests.

Reform China's Gaokao item writing mechanism

---From secluded item writing to an online test item submission system

江西金太阳教育研究院 陶百强

Gaokao, a high stakes test for college admission in China, is a "super" or largest-scale exam in the world, with as many as nearly 10 million examinees. The test developer and administrator, NEEA with Ministry of Education of China, will face a challenge that its Gaokao item bank calls for more items or tasks in terms of both quantity and quality with the popularization of nation-wide Gaokao reform piloted in Zhejiang Province and Shanghai in 2014, with examinees offered a second chance to sit the English test, developed by NEEA.

Although the traditional secluded item writing mechanism for Gaokao ensures confidentiality to a great extent, it is suffocated by disadvantages, such as inhumanity for secluded item writers (Li, 2014:2) and vulnerability to item flaws due to limited /no access to reference books, the Internet or corpora during high pressure item writing environment (Li & Tao, 2014). In contrast, foreign testing institutions, including the College Entrance Examination Center (CEEC) in Taiwan, China, Educational Testing Service (ETS) and Pearson, conducted an item writing mechanism open to the public. In mainland China ONLY

Zhejiang Province launched a Gaokao test item soliciting project in 2013 (The development of English subject test is commissioned to NEEA).

Based on the new challenge of increasing test items and new Gaokao policy, the author puts forward the following proposals to conduct a new mechanism (online test item submission system) to solicit test items to replace the traditional secluded and secretive item development mechanism.

A. Relatively-public item writing: If item solicitation was open to the public, it would be a tough job to screen out qualified test items from a plethora of items offered due to the huge population in China. Hence it is recommended that NEEA set limits to item submitters—testing agency is recommended to search the CNKI, the largest academic databank for Gaokao-related researchers in Gaokao item writing or NMET, and then send item-writing invitations, regardless of item writers' status and professional titles.

B. Quality management (Quality insurance and monitoring): Review submitted test items from the following facets: scientificity (conforming to item-writing guidelines), originality, fairness and standardization. The test item bank supervisor shall monitor the originality and freshness of test items in a dynamic manner, in order to retire then-original items or tasks but presently exposed to the public due to lapse of time .

C. Supporting measures: The country must make **regulations** and related **legislation** to ensure and strengthen test confidentiality and crack down on disclosure of confidential information and breach of related laws. Confidentiality is the first priority for item writing especially in China, a *Renqing-based society*^①. 3P principles, challenge system and a whistleblower program are proposed to ensure confidentiality during item writing process besides legal prosecution when violating the Criminal Law of China.

D. Reasonable remuneration for item writers: It's strongly recommended that testing institutions increase the testing investment to guarantee test item quality and related research funds. Item writing is a creative job, which should be respected and incarnated by reasonable remuneration.

New National College Entrance Examination and Basic Strategy for English Academic Test Evaluation

School of Foreign Languages, Jimei University HUANG Rui

The National College Entrance Examination in 2017 has decided that candidates can take English test beforehand. English subject has no longer had a united exam in the third year of the high-school; candidates can take part in social examination for twice in a year, or six times in three years in order to obtain the highest record for the college entrance exam. Because the key universities always ask for high record of English subject, candidates want to entrance those key universities they need spend more time and effort to improve their English. Otherwise, they may have no chance to enter those universities. Therefore, candidates may take English test in the first year of the high-school as soon as they finished entrance high-school exam; then they may spend more time in learning other subjects. As a whole, they need prepare scientifically about their learning plan on English during they are in junior in order to enter a key high-school and get a high English record. Thus, how to evaluate the English academic test? There are following several aspects: 1. Carrying out a multi-agent evaluation; 2. Paying attention to the learning process of students and development of the evaluation; 3. Putting forward to different implementation evaluation; 4. Taking full use of incentive functions of evaluation; 5. Making a decision of the practical evaluation contents according to the requirements of English subject; last but no least, 6. Using the network to evaluate students' ability of autonomous study effectively.

Room 313

What is tested?—Implications from our team's verbal protocol studies

Chongqing University GU Xiangdong, YANG Ruijin, MENG Lei, ZHONG Yu

Verbal protocol method is deduced from introspective method, which taps the subject's reflection on mental process, originating in the disciplines of philosophy and psychology. The subjects in verbal protocol studies must conduct a particular task while speaking out whatever thoughts pass through their minds (Ericsson & Simon, 1993). Since 2006, our team has

completed ten MA graduation theses and one BA thesis by employing the verbal protocol method. The verbal report methodology is employed in all these studies, but in different forms, i.e. introspection, immediate retrospection and retrospective interview. The method is proved to be effective in eliciting test-takers' test-taking processes, offering one optional way of validating tests from the qualitative perspective.

Tests in these studies include College English Test (CET), Test for English Majors (TEM), Hanyu Shuiping Kaoshi (HSK), The Test of English as a Foreign Language (TOFEL), International English Language Testing System (IELTS) and Cambridge English: Advanced (CAE). These studies focus on both test-takers' receptive abilities (i.e. listening and reading comprehension) in nine task types: long dialogue listening comprehension test, compound dictation test, multiple choice test, gapped text test, multiple-matching test, gap filling test, True/False/Not Given test, banked cloze test and vocabulary test; and productive abilities (i.e. writing and translation) in five task types: keywords composition test, picture composition test, Chinese-English sentence translation test, and Chinese-English and English-Chinese passage translation test.

By examining the verbal protocols, these studies identify the linguistic knowledge as well as various cognitive and metacognitive strategies employed by test-takers in test-taking processes, and find that some of the knowledge and strategies are the right constructs the tests intend to measure, such as grammatical knowledge, discourse knowledge for listening comprehension tests, clause-level, sentence-level and text-level information sources for banked cloze. Differences of test-takers' application of strategies across proficiency levels are also explored. The studies also find that there are some ability-irrelevant factors affecting test-takers' processes, which influence test-takers' performance and contaminate the validity of the tests to some extent.

Our studies have some weaknesses which might shed light on further studies using verbal protocol method. Firstly, a larger sample-sized study may be needed to confirm the study result. Secondly, it seems impossible that each subject verbalizes every single thought in his/her mind. As a result, a follow-up interview of each subject is needed. Thirdly, to some extent, the relevance and comprehensiveness of literature review part need to be improved. Fourthly, research frameworks should be derived from analyzing what the tests actually measure rather than simply adopting external criteria.

从促进人全面发展的角度来看英语语言测试

——以 2015 年高考英语北京卷为例

北京教育考试院 肖立宏

《国务院关于深化考试招生制度改革的实施意见》中要求“坚持育人为本，遵循教育规律。……培养德智体美全面发展的社会主义建设者和接班人。”在教育全球化、信息化、终生化以及当前人类所面临的由“科学知识型”转向“文化知识型”的转型期，教学目的、教学内容、教学方法、师生关系、教学评价等都应有所转变。教育的目的是实现人的整体发展，关注个人能力与潜力的全面挖掘；寻求个体之间的理解与生命的意义；关注人生经验，而不是狭隘的“基本技能”；促使学生养成批判性思维的能力，致力于人类文化的创造和改造，而不仅仅是现有文化的复制；侧重人文精神培养和融合。

一些高利害的考试，例如高考，对教学有着较强的引导作用。通过对考试内容的改革，能够引导教学关注人的全面发展，做到全科育人，全程育人。英语语言测试，除了考查考生的基本知识、基本技能之外，还要关注考生的情感、态度、价值观，关注考生思维能力、认知能力的发展。

本文依据《普通高中英语课程标准（实验）》和《2015年普通高等学校招生全国统一考试北京卷考试说明》，结合修订后的布鲁姆教育目标分类框架，从选材与试题两方面对2015年高考英语北京卷进行分析并提出改进意见。

Investigating Story Continuation Writing Task (SCWT)

Guangdong University of Foreign Studies YE Wei

Story continuation writing task (SCWT) has been established as a useful language learning task in the field of language acquisition, and proposal was made to expand its application to the assessment field (Wang & Wang, 2014). To explore the feasibility to its assessment use, this research presented an exploratory qualitative construct investigation by collecting and analyzing the TAP and retrospective interview data from twenty-one EFL writers of diverse proficiency levels and semi-structure interview from nine teachers, with the aim to explore the knowledge and skill types that facilitate the SCWT performance. The data thus collected was coded in view of the grounded theory and existing literature on writing process and Bachman and Palmer's (2010) communicative language ability (CLA) model. It was found that the SCWT demanded, among others, writers to acquire a basic understanding of the source story, which enables the extension of the given storyline, accuracy and diversity in word use, knowledge about the problem-solution structure embedded in narrative. It was

suggested that the SCWT, as an assessment task, promotes language learning by encouraging extensive reading and accumulation of a vast array of vocabulary.

交替传译策略能力的构成研究

广东外语外贸大学 许艺 王巍巍 伍志伟 穆雷

口译策略是口译能力中的重要组成部分。口译策略能力构成的确定对口译教学与培训、口译评估、口译资格认证和口译服务市场等方面起着至关重要的指导作用。然而，在口译研究中，口译策略尚未有统一分类，口译策略能力的构成也尚不明确。本文旨在通过文献研究法和对个案的模拟口译实验和深度访谈，基于扎根理论及认知策略和元认知策略理论，考察译员在完成交传任务时，运用了哪些策略，对交替传译的策略编码分析，探讨各策略之间的关系如何及其与口译水平之间关系，以构建交替传译策略能力理论结构模型。本文提出的交替传译策略能力理论模型可为中国英语能力量表（口译）研究、口译能力研究和口译测试研究提供理论基础，并为口译教学、口译教学大纲制定和口译教学评估方案制定等提供指导和参考。

“语言能力”与“口译能力”刍议

广东外语外贸大学 王巍巍 许艺 穆雷

“语言能力”是外语研究中的一个基本术语，在外国语言学与应用语言学界及翻译学界其所指却并不一致，翻译（口译）能力与语言能力的关系交错不清，因此，有必要厘清语言能力的定义和口译能力的定位及两者之间的关系。本文从 Chomsky 的“语言能力”概念、Bachman 的“交际语言能力”框架、及 CEFR 的“语言综合能力”等视角，梳理并辨析了国际语言学界关于“语言能力”的发展历程、术语范畴和概念关系。基于此，笔者尝试阐释翻译（尤其是口译）在综合语言发展中的定位，探讨口译能力与语言能力的关系，为中国英语能力等级量表（口译）的研发设计提供理论依据。

A Study on the CEFR-oriented Construct of Interpreting Competence

Guangdong University of Foreign Studies LIU Jianzhu

Adopting action-oriented approach, the Council of Europe set up the Common European Framework of Reference(CEFR) and proposed a concept of overall language proficiency, which covers communicative activities, communicative strategies and communicative competence. Based on the CEFR framework of overall language proficiency, the paper makes clear the relationship between linguistic competence and interpreting competence, which is divided into action competence, strategic competence and communicative competence. Considering the nature of interpreting, this paper explores the parameters of the three sub-competences so as to form the CEFR-based construct of interpreting competence.

A Preliminary Study on the Scales of Chinese Middle School Students'

Strategic Competence in Oral English Communication

Hunan Normal University ZHU Meifang

To cultivate students' communicative competence, which can be enhanced by communication strategies, is the essential aim of foreign language teaching. Communication strategies are a vital component of communicative competence, which should also be taken as a fundamental purpose of foreign language teaching. However, Chinese teachers stress much on the importance of reading and listening while pay little attention to speaking, let alone oral communication strategies. As a result, Chinese students, who even learned English for many years, still have deficiency in communication, especially in oral communication. Nowadays, many researches have been done on the construction of Chinese Common English Proficiency Scale, with the purpose to better connect teaching and assessing, which is the background of the present paper. Theoretically based on Communicative Language Ability and Cognitive Processing Ability of Discourse Information and targeted at Chinese middle school students, this paper aims to build a preliminary scales of Chinese middle school students' strategic competence in oral English communication with its focuses on the definition of strategic competence of oral communication and the dimensionalities of the scales. The results show that 1) strategic competence has rich connotations, including all methods and

techniques that employed to enhance the communicative goals; 2) the scales can be divided into two dimensionalities: expressing strategic competence, comprehending strategic competence, which can be further subdivided into four phases respectively: planning, execution, evaluation and repair; 3) the scales can be divided into three levels: primary, medium and advanced; 4) descriptors in the scale should be independent, precise, concise, clear and positive; 5) adopting qualitative and quantitative methods is the best way to ensure the reliability and validity of the scales. Building a sub-scale is a vital and critical step for the Chinese common English proficiency scale, which can provide some reference for the construction of Chinese common English proficiency scale and offer valuable guidance for English studying, teaching and assessing. This study is expected to give inspirations to similar researches.

Room 316

Classroom Assessment Practices Of Indonesian Junior High School Teachers In English As Foreign Language (EFL) Classes

University of Muhammadiyah Purwokerto, Indonesia Saefurrohman

The new language assessment policy in the Indonesia has impact on English teachers' assessment practices. Classroom assessment; as mandated in the current curriculum swifts from sources of information to the inseparable process of teaching and learning. This study describes Indonesian high school English teachers' classroom assessment practices in EFL classes.

This study was conducted using a mix method design with 24 Indonesian junior high school English teachers as respondents who completed the questionnaire on classroom assessment practices. Six respondents participated in an interview and observation to further clarify their practices on classroom assessment.

The study found that Indonesian junior high school English teachers used *assessment for learning* as the main purpose of assessment. *Multiple choice items* obtained the highest mean for assessing reading among Indonesian Junior high school English teachers. Moreover, in assessing writing, Indonesian Junior high school English teachers use *multiple-choice items to identify grammatical error(s) in sentence*. *The oral interview/dialogue* obtained the highest

mean for assessing speaking and listening for Indonesian junior high school English teachers.

The majority Indonesian junior high school English teachers used *items from published textbooks* as their primary sources for constructing assessment items. Indonesian junior high school English teachers used *written comments* as their primary method for providing feedback. *Total score test* and *a letter grade* were the highest percentage methods for providing a final report for Indonesian junior high school English teachers.

Competing Feedback From Summative Assessments And Classroom Instructions On

English Major Students' Use Of Reading Comprehension Strategies

Macau University of Science and Technology WEI Wei
Southwest of China University HU Rong

This study explored language learners' understanding of summative reading assessments and their use of reading comprehension strategies. It argues that students may misinterpret their performances in summative tests and the constructs being assessed. It speculates that their use of strategies may be affected by classroom instructions and other forms of formative assessments. Moreover, university professors' attempts of using other forms of assessments (e.g., write reading blogs and play drama) to upgrade students' reading comprehension strategies from local and literal level to a more global and interpretive level may be challenged by some summative assessment tasks.

98 junior English major students from Southwest of China University were recruited to complete a reading assessment tasks from TEM-4 and a survey. The data include the vocabulary size of each reading text at three levels (calculated by *Range*, a software program developed by Paul Nation), students' scores, their perceived difficulty level of each item, the sub-reading skills being tested by each item, and their use of reading strategies. Three lessons have been learned.

First of all, the difficulty level of each item estimated by test takers correlates negatively with the actual difficulty level as indicated by the average score at the 0.01 level ($r = -0.85$). As test takers may be influenced by several sources of feedback including the actual scores and perceived difficulty level, the mismatch between perceived and actual difficulty level of the reading assessment may cause further confusion on test takers' self-observation and self-assessment practices.

Secondly, explorative factor analysis was calculated to explore the underlying pattern

behind the self-estimated difficulty level of the 20 items. Five factors were generated and interestingly, the perceived difficulty level appears to be more associated with the four texts rather than the skills. All the items in Text A and Text D were generated as two separate factors, with the items from Text B and Text C forming another three factors. Further calculations on the vocabulary size indicated that the self-reported difficulty level is positively correlated with the vocabulary size at higher level. In other words, test takers seem to believe that the presence of a large number of difficult words makes their reading comprehension difficult. Meanwhile, the actual difficulty level (average score of each item) does not correlate to the number of difficult words at higher level, but the total word size of each text.

Last, student survey suggests that they believe the test assesses their reading comprehension at a more local and literal level, which seems to go against students' use of reading comprehension strategies. Explorative factor analysis identified four factors, with the strategies of comprehending reading text at a more global and interpretive level being reported as the most two popular strategies, followed by the strategies of learning new vocabulary. We suspect that this is a result of university professors' use of other forms of assessment tasks, which promote these strategies at higher level. Surprisingly, classroom based assessments seem to play a more significant role than summative and high stakes assessments.

Assessment For Learning In Classroom: Students' Management Of Their Own Learning Goals

Liuzhou Vocational & Technical College LI Xuelian

This paper aims to investigate the influence of managing English learning goals under the theoretical frame of Assessment for Learning. To be specific, it tried to answer the following three research questions by collecting data of students' self-reflection, questionnaire investigation, classroom observation, interview and students' learning achievement:

1. Would the learning achievement of students with self-set learning goals be higher than the control group?
2. Would there be any gender differences on learning initiative between two groups?
3. If yes, what were the reasons for gender differences?

Two classes, altogether 93 students, participated in the present study. The result showed that students' initiative of managing their own learning goals was positively related to their English achievement, and both could be enhanced by the improvement of each other. In addition, girls were significantly active than boys in setting their own learning goals and

carrying out, evaluating and reflecting on their learning activities based on the goals. The study also found that students' English proficiency, self-regulated learning strategies and attitude influenced students' initiative of managing learning goals. The fourth finding was that some students set too high learning goals which usually frustrated their continuing learning.

In view of the four findings and the second language learning environment, the present study enriched the literature of Assessment for Learning. It also raised four pedagogical implications to teachers: creating safe and encouraging learning environment with students, affording more attention and help to low attainers, teaching self-regulated learning strategies step by step, and monitoring students' learning goal setting to avoid too high goals.

Feasibility And Challenges Of Assessment For Learning In Large Classes

Guangdong University of Foreign Studies The University of Hong Kong

XU Yueting

Assessment for learning (AfL), the use of assessment to support learning, has been an attraction to many educational practitioners due to both its strong learning-oriented flavor in theoretical underpinning and the documented positive evidences from many Western countries where the 'assessment culture' prevails. For educators from Confucius-heritage cultures (CHC) setting (Biggs, 1996), however, their implementation of AfL seems to be confronted with many practical challenges (Carless, 2011) such as exam-oriented culture, big class size, and stakeholders' resistance, which constitute the actual contexts for their AfL practices. These challenges may undermine the effectiveness of AfL, and how to offset the negative effects mainly rely on the classroom teachers' practices.

Against this backdrop, this study seeks to understand how EFL instructors deal with AfL when teaching large classes. Drawing upon data from semester-long classroom observations, teacher interviews, and documents among three case teachers working in three Chinese universities, it attempts to answer the following research questions: 1) What are the assessment methods utilized by the three teachers? 2) What particular challenges have large classes brought to AfL implementation, and what are the teachers' strategies? 3) How effective are these strategies, and what are the implications for teacher assessment literacy? The findings include institutionally prescribed assessment plans, reduced individualized

instruction and scaffolding as the main challenge, as well as varied and context-responsive strategies as compromises made within the tensions between the mediating factors. Implications for AfL implementation in school setting and for teacher assessment literacy are also discussed.

Turning Assessment into Mediated Learning Experience

-- Assessing EFL Students' Independent Learning

by Learning Logs

Jiangxi Normal University TANG Xiongying

EFL learning success needs to be backed up by adequate practice outside classroom. However, lack of meditative support from teachers and peers, independent learning outside classroom could cause confusion, discontinuous effort, and giving up easily on the part of the students. Assessing this part of learning and awarding grade on students' performance could remind students of doing their job because it contains stakes for student evaluation, but learning difficulties may remain unnoticed and unattended by the assessors. As the efficiency of such learning can remain low, it is quite dubious that students may benefit from such assessment.

Based on the assumption that students' independent learning needs more mediation from teachers and peers, this study continued their investigation of curriculum-embedded assessment in 2014 and practiced assessing freshmen university students' independent learning outside classroom by asking them to keep learning logs, particularly of listening and reading. The process, however, was supplemented by classroom discussion, sharing among peers, and teachers' suggestion on the choosing of materials, time (text) length of practice, learning strategies, etc. and thus turned the assessment into a more mediated learning experience.

As the reflective summaries of their logs show, students saw the assessment as a task, but could finally turn such involuntary task into a healthy habit of learning. Much practice helped them become more dedicated and skillful readers and listeners. The average amount of learning, time length in listening and text length in reading, also met the set requirement at the beginning of the term, though variation existed among individuals. On the part of the teacher, learning logs revealed to her the width and depth of students' listening and reading activities, factors which might facilitate and mitigate the effect of their learning effort, benefits and drawbacks of learning log as an assessment, as well as the possible supporting teaching

activities provided in class. While some of such information prompted the teacher to give timely feedback, other could affect her curriculum decisions for the next round of teaching.

In sum, such experience reveals what benefit students can get from the assessment and what a teacher can do for the students. It also points to a future research need on a specified criteria reference provided to EFL learners, for example, in descriptions of cognitive levels of certain learning activities, suggested learning materials and minimum amount of practice etc. The supportive role of such criteria could be significant given that independent learning after class constitutes a major part of EFL learning and may as well contribute largely to learning success in absence of mediation of others.

高校外语课程中的任务型语言评价探索

江西师范大学外国语学院 刘春燕

随着交际法和任务型外语教学理论的发展和外语教学实践的变革，外语测试也出现了与之相应的范型转变，即从传统的分立的语言点多项选择型测试转向真实的任务型语言评价范型。本文从高校外语评价的实践出发，提出任务型语言评价取代传统的高校外语课程评价方式。任务型评价通过与真实世界情景中相关的言语行为表现来测量学习者语言产出技能。文章详细阐述了任务及任务型测试模型、任务型测试的语言指标（复杂性、准确性和流利性）、口头任务测试和写作任务测试的设计与实施，以及该评价方式面临的挑战。文章指出任务型评价模式是推动高校外语课程改革的重要环节。

Room 317

Formative assessment studies in English teaching in China: review and reflection

Sanming University YUAN Shuhou

At present, the further study of formative assessment in the field of foreign language teaching is of great practical significance for the examination evaluation reform and the quality specification of foreign language talents in China. In order to get the comprehensive, in-depth and systematic literature of formative assessment, the author read the content of the

titles, keywords and abstracts of the papers published in 18 Chinese foreign language journals from 2001 to 2015, finally get 156 papers. Then this paper gives a qualitative and quantitative analysis of the research on formative assessment in foreign language teaching in the fifteen years in China from the perspectives of research trends, subjects, contents and methodologies. The results indicate that: (1) the amount of papers keeps increasing on the whole; (2) the research perspectives are multi-disciplinary; (3) empirical studies account for a large part, while theoretical studies account for a small part; (4) the subjects are foreign language learners of different educational backgrounds and different learning skills; (5) the samples become larger and more refined; (6) dynamic real-time assessment studies develop faster than dynamic process assessment ones. The paper also discusses the limitations of the research area: the effectiveness of certain research achievements needs to be validated, some repeated studies are in foreign language writing teaching, research perspectives should be broadened, and research trends abroad and age needs should be followed closely. Finally, the paper proposes the directions and prospects for future research on foreign language formative assessment in China.

The cultural appropriateness and appropriation of formative assessment in the Chinese sociocultural context

Foreign Language School of Shanxi University CHEN Qiuxian

A borrowed policy would in most occasions find itself faced with multiple challenges arising from the context that it is situated in. The implementation process of the policy, thus, mostly sees that the policy is negotiated and appropriated by the sociocultural realities of the situated context. That is the case for formative assessment in China. Owing to the contextual differences, the practicing of formative assessment in the Chinese context is complicated rather than straightforward. The effect of formative assessment initiative has accordingly been much compromised.

This paper, adopting a sociocultural perspective, and backgrounded against the formative assessment initiative advocated in *College English Curriculum Requirements* [CECR] (Higher Education Commission, 2007), sets out to examine the appropriateness/inappropriateness of formative assessment principles, which were developed and advocated in the Anglophone context, to the Chinese education context where the traditional cultural values are deep-rooted.

Then, with a review of empirical research about the implementation of formative assessment in the context, this paper pinpoints the alignments/misalignments between the advocated formative assessment principles and the formative assessment implemented in the Chinese context. Further, it showcases how the principles of formative assessment has been appropriated in the area of College English, as influenced by the situated sociocultural factors.

The purposes to fulfil are four fold: the first, to clarify possible misconceptions of formative assessment in the context; the second, highlight issues to be taken better care of for formative assessment initiative to be more effectively implemented in this particular sociocultural context; the third, to identify implications to inform policy further change and classroom practices; the last, to call for the urgent need to theorise a formative Assessment for China.

形成性评估在非英语专业研究生英语写作教学中的应用研究

Fuzhou University FAN Meng

形成性评估 (formative assessment) 是在教学过程中为引导教学正确、完善地前进而对学生学习结果和教师教学效果采取的评估。相对于传统的总结性评估而言, 形成性评估是近些年西方教育学家们所大力倡导的一种创新型测试评估方式。其主要目的不是为了选拔少数优秀学生, 而是为了发现每个学生的潜质, 强化改进学生的学习, 并为教师提供反馈。因此, 形成性评估是一个持续、循环的改进教师教学和促进学生学习的过程。

形成性评估已在国内外教育领域受到广泛关注, 也是目前我国外语教学研究的一个热点。有学者提出非英语专业的写作教学课堂有大量的研究空间。然而, 针对形成性评估在非英语专业研究生英语写作教学中的应用的相关研究却很少。近些年, 我国高校英语教学一直在进行探索式改革。作为对英语特别是英语写作有较高需求的一类特殊群体, 非英语专业研究生仍未达到所期待的学习成果。

本课题以华东地区某 211 大学非英语专业研究生一年级的公共英语课为研究案例, 在形成性评估的理论基础上结合巴兹尔·伯恩斯坦的教学话语理论和休伯特·赫曼斯的对话自我理论, 通过教学实践与使用问卷量表、访谈和观察的方法, 使用定量和定性分析关注形成性评估在非英语专业研究生英语写作教学中的应用。目的是为师生写作课堂上口头与书面反馈的教学实践创新提出一些建议, 促进我国高校公共研究生英语教改, 并推进我国测试理论与实践研究的发展。

The impact of formative assessment on motivation and achievement in English classes in Chinese primary schools

School of Education University of Sheffield

April Jiawei ZHANG

Formative Assessment (FA) has been a hot topic in education for decades. Wiliam (2006) underscores that Formative Assessment and Summative Assessment (SA) are fundamental terms in education. While in China, from June in 2001, a new curriculum system was adopted into the educational field. According to the New English Curriculum for Chinese Primary Schools and Junior/Senior Middle Schools (New English Curriculum/NEC), the new system stresses the importance of FA in language learning. The latest edition of the NEC (2011) retains the emphasis on Formative Assessment; moreover, it adds a lot of specific assessment techniques for teachers. Many researchers studied the FA, but in higher education environment, my concern lies on the use and the impact of FA in primary schools. Most of Chinese students start to learn English as a compulsory subject from the third grade (8-9 years old). It is easy to influence them by reacting differently at the early stage. Therefore, to find an effective way to motivate the learners and help them achieve more in English is what I am pursuing in this research.

FA is an interactive and dynamic procedure which aims at improving learners' English by giving different reactions on the basis of learners' performance (Black & Wiliam, 2009). The 'reactions' include concrete strategies, such as sharing learning goals, asking questions, teachers' feedback, peer- and self-feedback. These strategies are the most popular ones and being discussed a lot in the literature (Black et al. 2002, 2003; Black & Wiliam, 1998, 2009; Wiliam, 2000, 2007a, 2007b; William and Thompson, 2007; Heritage, 2009a, 2009b, 2010; Torrance & Pryor, 1998, 2001; Wiliam & Thompson, 2007). My study is to seek the relations between these strategies and learners' motivation for learning English and their English achievement. I have 182 out of 235 students within 4 classes taking part in my study; 2 English teachers, one creates FA-rich environment and another uses a few assessment during the teaching. I will use mix methods to analyse my data which was collected from 2 questionnaires on motivation and 10 weekly achievement quizzes; and 17 semi-structured interviews with English teachers and students from different achievement levels.

In this talk, I am about to present: the theoretical frameworks of FA and motivation; the research instruments (questionnaire on motivation and interview questions on FA); the field work with its procedures, difficulties and achievements; and the preliminary data analysis.

On foreign language teachers' assessment literacy in portfolio assessment

遵义医学院珠海校区

陈超

Testing and assessing are closely associated with teaching. Proper testing and assessing have an effect on the facilitation of teaching. The rise of assessment for teaching and learning brings an emphasis on the assessment literacy of foreign language (FL) teachers. However, a lack of assessment knowledge at present hinders FL teachers from following up the development of testing theories and practice. How to enhance assessment literacy among FL teachers, implement formative assessment and make learning happen has become one of researchers' top issues. Meanwhile, the establishment of portfolios for teacher and students is a vital part of formative assessment. This paper reviews research in China and abroad on FL teachers' assessment literacy and portfolio assessment, explores ways to enhance FL teachers' assessment literacy in conducting portfolio assessment.

大学英语教师口语课堂形成性评价能力研究

----教学支架使用情况调查

江西师范大学外国语学院 蔡文艳

在我国，课堂作为学生学习英语的主要场所，教师的形成性评价能力对学生有着至关重要的作用。有效的形成性评价对教师的教学也有着重要意义。形成性评价应主要关心对学生课堂学习的诊断和反馈，致力于对学习过程的促进或“形成”。其中教学支架的使用是形成性评价中比较重要的内容。

本研究主要考察大学英语口语课堂中支架的使用和效用，具体了解教师更倾向于使用哪种支架和哪种支架易于被学生接受，从而促进对教师即时诊断和帮促学生学习的评价能力的理解。对课堂中使用的支架进行研究可以对形成性评价能力的建构提供借鉴和启发。

本研究主要采用观察和访谈法。笔者在已有研究的基础上，设计教学支架观察框架，在征得任课教师的同意但不告知研究目的的情况下，对江西师范大学的三位大学英语口语教师的口语课堂各进行四次观察，并对整节课进行录音（每堂课 40 分钟，共 480 分钟），课后对录音材料进行整理，并转写成文本语料以进行分析。同时，笔者将在课后即时对三位教师进行访谈以了解教师对形成性评价的理解和运用。

通过分析，该研究发现了大学口语课堂中教师运用到的教学支架，有利于理解有效的教师形成性评价手段，从而提高教师形成性评价能力。有效教学支架能加大学生的口

语输出量，口语兴趣也会得到提高，口语能力、自我效能感也会得到增强。基于以上调查研究结果，文章对大学教师的口语课堂形成性评价能力的建构进行了相应的讨论。

Room 318

A Study into the Effect of Task Type on Linguistic Features of CFL

Learners' Writings

Shanghai University ZHANG Xinling ZHU Lin

The present study is a testing attempt to probe into the effect of task types on the linguistics features of Chinese as a foreign language learners' writings. The results yielded from 94 subjects' writings showed that among the 13 indicators selected, 10 were significantly different between impromptu writing group and the two integrative writing groups, and both reading-to-write and reading-listening-writing successfully improved the correctness of verb, adverbs, cohesive devices and so on, resulting in better language accuracy, fluency, and complexity. However, integrative tasks require higher ability in terms of multi-literacy, on the condition of comprehensible reading and listening input is provided.

翻译专业本科成就测试口译部分题型研究

——基于《翻译专业本科国家标准》的测试设计

广东外语外贸大学 伍志伟 许艺 穆雷

本文汇报翻译专业本科成就测试口译题型的设计开发。基于《翻译专业本科国家标准》，本研究设计了英语听力、信息概要、职业道德、百科知识、对话口译、英中交替传译以及中英交替传译等七大题型。通过对 57 名翻译专业本科学生的试测，本研究比较了在不同测试条件下，英语听力、信息概要、职业道德这三个题型的难度是否存在差异。此外，本文还从命题信度、共时效度、理论效度等方面探讨了上述七个题型之间的联系，并结合《国标》及有关口译能力模型，为翻译专业本科成就测试的设计提出有益的参考。

外国留学生汉语语用能力测试

广东外语外贸大学 范香娟

本研究的目的是开发一套测试汉语学习者中介语语用能力的多项选择话语填充测试的试卷，并采用Messick（1989）的理论对其进行效度验证。中介语语用能力的操作化定义为语用理解和产出两个方面，包括对言语行为和程式语的产出，对会话含意的理解三个方面。

本研究采用Liu(2006)的试卷开发方法，开发过程包括情景调查，可能性调查、言语行为社会变量的元语用调查、书面话语填充测试、选项的征集和试测以及多项选择话语填充试测等多个步骤。从最初1269个情景中得到60个情景，组成多项选择话语填充测试的试卷。试测后的试卷由50个题目组成，包括30个言语行为，7个程式语和13个理解会话含意的题目。试卷的正式施测有208名汉语学习者参加，对其中2名受试进行了有声思维。

对测试得到的结果进行定量和定性分析，依据Messick的效度理论，从内容效度、实质效度、结构效度、概化效度、外部效度五个方面对结果进行了分析和讨论。结果显示整份试卷的内容效度和实质效度良好。题目稳定性较高，为0.97，受试稳定性，即受试区分信度虽然仅为0.75，对于这一测试来说是可以接受的。相关分析和因素分析显示了言语行为和会话含意都属于语用能力这一因素，而程式语可能属于不同的因素或构念。语言能力、文化背景、性别、年龄和教学情况都为测试提供了很好的外部效度，时间和华裔背景两个因素还需要进一步的调查。最后，讨论了试卷的实用性，本研究对试卷开发和语用教学的启示，指出了本研究的贡献和局限，并对将来的研究提出了建议。

What Teachers Should Know about Vocabulary Tests

for EFL Freshman Students

King Saud University

Reima Al-Jarf

At my college, translation students take 2 vocabulary courses in the first year of the program. The courses aim to develop freshman students' knowledge of core vocabulary. About 50 lessons with different themes are covered. The students take 2 in-terms and a final exam. The presentation will give a comprehensive guide to planning and designing vocabulary tests as follows:

(1) Identify the skills to be covered by the test: Pronunciation, spelling, identifying the part of speech, morphological structure, semantic aspects, register and so on.

(2) Outline the course content covered.

(3) Prepare a table of specifications that shows the skill, content topics and number of questions allocated to each.

(4) Prepare test instructions such as: (i) Write the silent letters in each word on the lines; (ii) Write the part of speech of each word; (iii) Circle the words in which -er is not a suffix; Underline the words that have no singular form; (iv) For each word give a synonym; (v) Give the meaning of each word in Arabic; (vi) What is the American expression for the following; (vii) Fill in the blanks with a preposition; (viii) Show the difference between each pair by giving the Arabic meaning; (ix) For each underline phrase, which idiom/collocation can be used? (x) For each situation, use an expression of 'apology'.

(5) The test should meet the following criteria: (i) Instructions should be brief and clear; (ii) questions should cover all kinds of skills, tasks and exercises covered in the classroom and textbook; (iii) they should require the students to perform tasks at the paragraph/ discourse level as well as word and sentence levels; (iv) questions should test student's ability to think, apply, infer, connect, and synthesize information, not mere recall, and should not use exact sentences and examples from the textbook; (v) tests should have as many production questions as possible; (vi) they should have adequate discriminating power; (vii) should be reliable and valid; (viii) should be a power and a speed test; (ix) should have an average difficulty level.

In addition, the presentation will show the following:

(6) What each test covers; test length (how many words, how many pages, how many words per question), when to give the tests during the semester and test duration.

(7) Test paper format (margins, font size, instructions, line spacing, pages, numbering pages and test items, simplicity, no decorative boxes, circles, flowers, test and student information).

(8) How tests are scored: marks allocated for each question type and whole test; using whole numbers not fractions; deducting points for spelling and grammatical mistakes; putting a check mark next to the correct answer, add up the check marks; where to record the total mark for each question, each page and the whole test.

(9) After scoring the test papers: return marked test papers to the students; show the marking system and go through common errors.

(10) Follow-up issues such as using test results for diagnosing weaknesses and remedial work.

Are we over-testing or double-checking?

Anadolu University

Sercan SAGLAM

In assessment of language skills, there are decisions that test administrators have to make for effective and accurate measure of student performance. One of these decisions is how language skills will be assessed. One way is to measure it directly, whereas the other option is to test it indirectly. Both direct and indirect assessment of language skills have been applied in various testing contexts and there is still no consensus on which one is better. Yet, direct testing of language skills is generally advocated by language testing experts.

In the context where this study is conducted, both direct and indirect assessment of vocabulary and grammar knowledge have been employed. At the end of an intensive language program, the students have to sit a proficiency exam to display their expertise in the language. The proficiency exam consists of three components. There is a multiple choice test focusing on reading and listening skills, as well as vocabulary and grammar knowledge. There are also speaking and writing components where students display their command of spoken and written language. These skills are graded using an analytical criterion where vocabulary and grammar knowledge are part of the grading scheme. In other words, the students are assessed both directly and indirectly with reference to vocabulary and grammar knowledge.

The aim of this study is to examine the correlation between the scores the students obtain from direct and indirect testing of vocabulary and grammar. For the data analysis, the scores from the vocabulary and grammar sections of the multiple choice test will be correlated with the scores the students obtain from the speaking and writing component of the proficiency exam. The analysis will hopefully show whether we are over-testing our students' performance or double-checking.

测试方法对语法测试的影响

广东外语外贸大学英文学院 张惠玲

The present paper reports results of a study which examined the effect that two methods used for testing grammar have on test-takers' grammatical performance in the tests. The two methods compared were the gap-filling cloze grammar test (G/F) and the multiple-choice cloze grammar test (M/C) with identical content. 164 senior two students first took the G/F test and four weeks later they took the M/C test. 12 of them were invited for interview about their test-taking processes immediately after they finished each test. After examining their test-taking processes, the verbal data were coded in terms of test-takers' grammatical knowledge and strategies used in these two tests.

Comparison of test-takers' scores and their individual answers in the two tests found that their scores were significantly different and their answers had changed between the two tests in about 45% of the overall items answered, which indicated that the two tests elicited different performance of test takers. This was further evidenced by comparing the frequency with which grammatical knowledge and strategies were used by test-takers. Results showed that test-takers' reporting use of accurate, inaccurate or no reporting use of grammatical knowledge in the two tests were significantly different ($\chi^2=6.448$, $p=0.04$). In terms of strategies used in the two tests, it was found that test-takers tended to use cognitive, metacognitive and test-wise strategies more frequently in the G/F test than in the M/C test. Such differences regarding the use of grammatical knowledge and strategies in the two tests revealed that different test-taking processes were involved. Therefore, it was concluded that test methods did affect test-takers' grammatical performance and the two test formats with identical content did not measure the same trait.

Room 422

Exploring Chinese EFL learners' cognitive processing of written corrective feedback: A case study

University of Hong Kong HAN Ye

While previous research investigated L2 learners' uptake and learning outcomes of written corrective feedback (WCF), a type of classroom assessment on the accuracy of student writing, much less attention has been paid to learners' cognitive engagement with WCF. Although some studies have touched upon learners' attention and understanding of WCF (e.g., Storch & Wigglesworth, 2010), the individual learners' mental processes involved in receiving and responding to WCF remains under-explored. To address the research gap, the

current study adopting a naturalistic, qualitative multiple-case study approach investigated (a) to what extent individual L2 learners understand WCF, (b) what mental processes are undertaken when individual L2 learners receive and respond to WCF. The study lasted for a semester of sixteen weeks, and involved eight non-English major students from two integrated skills English classes in a Chinese university.

Data was collected from multiple sources, including students' drafts of four essays with teacher feedback, semi-structured interviews, retrospective interviews, teacher-student writing conferences, classroom observation, reflective accounts, and class documents. The findings show substantial individual differences in terms of the depth of processing. Higher-achieving students, who were more motivated, better equipped with meta-linguistic knowledge and self-regulated strategies, tended to have more accurate and thorough understanding of WCF and underlying metalinguistic rules of target forms. They not only noticed WCF, but also analyzed WCF to identify and diagnose errors, explored and evaluated their revisions by using multiple meta-strategies and strategies. In contrast, struggling students, who mainly aimed to obtaining a passing score, doubted the effects of WCF, lacked meta-linguistic knowledge and self-regulated strategies, and tended to process WCF on the surface level, which led to a shallow understanding of WCF and target forms. The findings suggest that extra, individualized assistance should be provided to under-achieving L2 learners to develop their self-regulation of mental processes, so as to enhance their cognitive engagement with WCF.

A Correlation Study on Key Positioning and Test Performance with Language Test Multiple Choice Items

Shaanxi Normal University WANG Fang HE Junjie
Hunan Weichujingcai Training School WU Linqiong

This paper sets out to probe the question whether the position of Multiple Choice (MC) keys influence test performance. The experiment result suggests that different positioning of the keys correlates significantly with different performance. Most importantly, early appearance of the keys of moderate items and later appearance of the keys of difficult items both help increase the recognition rate for their keys.

Chinese EFL Learners' Writing Process in Reading-to-write and Reading-listening-writing Integrated Writing Tasks

Shanghai University ZHANG Xinling ZHOU Yan ZHANG Siyu

This paper is an attempt to investigate into Chinese EFL learners writing process of reading-to-write and reading-listening-writing tasks. Data of 120 high- and low-level Chinese

EFL learners indicated that: 1) both tasks sampled the subjects' ability of using discourse synthesis strategy, self-monitoring strategy, and test-wiseness strategy, 2) L2 proficiency, task type, and their interaction partly influenced the type and amount of their use of the above strategies, 3) the high L2 proficiency group used selecting, organization and monitoring strategy more frequently and preferred evaluation strategy before writing in the reading-listening-writing task, 4) while the low L2 proficiency group favored evaluation strategy in post-writing phase in the reading-to-write task, 5) the reading-listening-writing task was more desirable as far as plagiarism issue was concerned.

The Effect of Priming Position and Learners' Proficiency Level on Contextualized Word Association

华南师范大学 XU Yiying, ZHANG Ping

Previous studies show that during the development of L1 mental lexicon, native speakers usually underwent syntagmatic-paradigmatic shift in word association task. This process is viewed as an indicator of native speakers' development of semantic networks in mental lexicon. However, quite a few studies (Wolter 2006, Ping Zhang 2010) have found that syntagmatic relations seems to be harder for L2 learners. Nevertheless, previous studies present the cue words in word association without context, which may influence the word association results.

Drawing support from the native speakers' corpus, this study derives the clusters from the corpus and provides them as context for word association task. Clusters refer to the high-frequency word sequences containing three or more words which conveys a more complete meaning. They are chunks in the narrow sense. Comparing with individual word, chunks provide a more complete context and could be stored and retrieved as whole.

The context-dependent word association online experiment extracted 15 pairs of v-n collocations with their corresponding chunks and filler materials. This provides natural context for the cue words in our word association context. The goal of this study is (1) to compare the word association results in simple word association task and word association results under the context of chunks. (2) to show whether the change of cue words position in chunks provides different word association results. (3) to investigate how English proficiency level influences the word association results.

The experiment was a mixed design, 2(contexts types: context-derived, chunk context) * 2(stimulus positions: left word [verb of the collocations], right word [noun of the collocations]) * 2(subject groups: low-proficiency learners, high-proficiency learners). The subjects are senior high school students and college junior students. And the e-prime software is used as the experiment tool. The first experiment is an on-line free word association task. It requires the subjects to express the first word that come to their mind when presented with a certain cue word. In experiment two, the subjects are first required to finish a maze task, which are designed based on Foster (2009)'s experiment. In the maze task, words

and an asterisk appear simultaneously on the screen and the subject has to judge which is a word and press the corresponding T or F key. Then the chunks flash quickly on the screen for 1 second. Finally, a verb or noun cue word derived from the chunk appear and the subjects are required to carry out the word association test. We carry out a longitudinal study which lasts for a semester.

By providing chunks as context for Word Association task, the subjects are expected to produce more syntagmatic response. We hope to explore the role of context in the development of L2 mental lexicon and provide effective suggestions on L2 learners' vocabulary teaching.

A Verbal Protocol Analysis of Test-takers' Processes
of Taking Reading Test of Cambridge English: Advanced
Chongqing University TIAN Jie GU Xiangdong ZHENG Yujing

The present study was intended to investigate the test-taking processes of six Master of Art students taking CAE Reading test by utilizing think-aloud method in the purpose of examining the reading strategies employed by test-takers with different reading ability when they are answering different task types of multiple choice, gapped text and multiple-matching. The study is focused on the following three research questions: 1) what kinds of reading strategies are utilized by test-takers in the processes of CAE Reading test? 2) Are there any similarities and differences among the two groups with different reading proficiency when they are taking the CAE Reading test? If so, what are the differences? 3) Are there any similarities and differences about the utilization of strategies in different task types?

The major findings of the current study are summarized as follows:

Firstly, it is found that 20 out of 21 reading strategies from *coding schema for reading strategies* (Lin & Yu, 2013) are utilized by test-takers except “visualising information read” strategy, and the 20 strategies are classified into three strategy categories: metacognitive strategies, cognitive strategies and support strategies.

Secondly, Test-takers in high-level group and low-level group share the same reading strategy preference: their most frequently used strategy categories in descending order are supporting strategies, cognitive strategies and metacognitive strategies. However, the proficient test-takers are found to utilize more reading strategies than those with lower proficiency, and they employ significantly more metacognitive strategies as well. Besides, test-takers with lower reading proficiency seem to rely largely on the word-level cues while those proficient test-takers seem to concentrate more on the analysis and evaluation of the text.

Thirdly, test-takers show different reading strategy preferences in different task types. Specifically, test-takers use “translation”, “evaluating what is read” and “using context clues” more frequently in multiple choice task; “translation”, “using context clues” and “guessing

word meaning” in gapped text task; “translation”, “reread”, “guessing word meaning” and “asking questions” in multiple-matching task.

This study may provide some references for the understanding of information processing process in reading test, the design and implementation of think-aloud method as well as the training of students’ reading strategies. Limitations of the study include the incomplete reflection of test-takers’ cognition processes from the verbal protocol data, the small size of subjects and the lack of exploring strategy use mechanism.

The Application of Think Aloud Protocol in the Reading Strategy of New HSK Band 5 Reading Test

Chongqing University DENG Yanping YANG Ruijin

In order to investigate the strategy use by foreign students in the process of Chinese reading, this study invites two foreign students (at Chinese intermediate level) who are going to receive HSK V test to participate in this HSK V reading test through video recording based on the method of think-aloud protocol and records their test-taking process. Besides, immediate retrospective interviews are also conducted.

Through employing these two methods, this study makes clear of the reading strategies used in the HSK V reading test and the features of reading strategies according to different test items. Besides, based on the item types and length of the HSK V test, information such as reading time and reading speed are also considered in this study. Overall, the main conclusions of this study are as following: 1) Test items influence reading modes which in turn influence reading time as well as reading strategies. 2) Marking is also a reading strategy and is related to the strategy of searching key words in think aloud protocol. 3) Reading model, reading time, reading strategy and marking strategy reflect candidates’ different proficiency level of Chinese reading.

Moreover, this study finds that reading behavior such as marking, scenario simulation could reflect the reading process of candidates as well, thus it is suggested that the method of video observation could be used in the research of reading strategy from different aspects such as sound and behavior, and from different subjects such as psychology and foreign language teaching.

An Introspective Study on Test-takers' Process of Keywords Composition in New HSK-5

YANG I GU Xiangdong HONG Yue

The new Hànyǔ Shuǐpíng Kǎoshì (HSK) is an international standardized Chinese proficiency test for non-native speakers. Previous studies on the writing of the new HSK mainly focus on error analysis (e.g. Xin, 2001; Lu & Yu, 2012; Zhu & Feng, 2013; Song et al., 2015), and center on static output, but much less into the dynamic output processes. As an uncommonly used and rarely studied item type in language assessment, keywords composition constitutes an essential part of the writing section in the new HSK-5. It's worthy of our attention to study test-takers' writing process of this item type.

This study employs a concurrent think-aloud protocol and immediate retrospective interviews to investigate the dynamic process of keywords composition in the new HSK-5. The current study mainly concentrates on three elements involved in test-taking process, i.e. discourse organization, strategy use and the relation between test-takers' written products and their writing process. It intends to address three research questions: 1) What steps are involved in discourse organizational process and how do test-takers organize them? 2) What types of strategies do test-takers employ and how do they use these strategies? 3) What are the relations between test-takers' written products and their writing process? Six foreign MA students in Chongqing University who had passed HSK-5 and were at the same level (based on their new HSK-5 scores) participated in this study.

The major findings are as follows. Five steps are identified in discourse organizational process in sequence of recognizing keywords, identifying theme, writing with known keywords, making sentence with unknown keywords and reviewing. Writing strategy uses include cognitive strategies (regression, translation, review, self-question, guess, avoidance and paraphrase), metacognitive strategies (monitor, plan and evaluation), and affective strategy (anxiety overcome). Among them, regression, translation and review are most frequently used strategies. Test-wise strategy (resource replicating) is also found used by students in keywords composition. In most cases, the written products are consistent with verbal protocols, but unidentical cases including wrong spelling, content replacement, addition and missing are found as well. Attention distribution and language proficiency may be the main causes to such inconsistency.

This study provides valuable insights into the test-takers' process of keywords composition in new HSK-5 and merits the design of the new HSK-5 and also Chinese teaching and learning. It is hoped that this study offers theoretical and methodological references for further investigation into the new HSK writing and similar test types in other

tests.

Room 423

A Students' Perspective on Intended Consequences of a High-stakes Test in China and Its Effect on Test Performance

University of Melbourne

BAI Ying

Since the 1980s, validity theory in educational measurement and language testing has witnessed a shift from a focus on validity investigation of the test itself to score interpretation for a particular test use. When the social dimensions and consequences of assessment are included in discussions of validity, the test stakeholders' position has to be re-evaluated. An increasing number of scholars recognise the process of exploring students' perspectives on a test as a key part of test validation in order to re-evaluate the promoted standards and intended results, and to discover any unintended and unpredicted occurrences in practice (Bachman and Palmer, 2010; Davies, 2008; Hawkey, 2006; Saville, 2012). However, until now, no empirical research has explored students' perspective on intended test consequences, especially on the appropriateness of test uses. Furthermore, no research has attempted to take the dynamic social context and individual characteristics into consideration and question whether and how students' perspectives on the use of a test and the decisions made on the test-score interpretation affect their test performance.

This project intends to explore from students' perspectives whether a high-stakes test (the College English Test) in China has achieved its intended purpose, and to reveal the potential interactions between students' perspectives on a test and their test performance, "Attitude" is the terminology applied to capture students' perspectives. The potential effects of students' individual characteristics on their attitudes are also explored. This study aims to provide empirical evidence to reveal the significance of students' perspectives in the process of testing and that of the integration of value implications with validity discussions.

This project is composed of two phases, combining both qualitative and quantitative methodologies. Phase I focuses on exploring students' attitudes; Phase II is developed in two stages. Stage one, the development of a questionnaire, the data collection instrument for Phase II, based on the findings of Phase I; Stage two, investigation of the relationships of students' attitudes, test performance and their individual characteristics in a comprehensive model, using structural equation modelling (SEM) analysis.

The preliminary analysis showed that students in general believed the CET was a trustworthy test; and a majority of them held a supportive attitude towards CET test use. Their attitudes towards test preparation approach were largely influenced by the language policy at a local level: they were more supportive of their English teaching when the policy focused on improving English use competence; and less so when it focused more on test results. The SEM analysis demonstrated that students' English learning motivation and test-related learning experience positively affected their attitudes. Students' attitudes towards test use and test preparation approach were positively and negatively related to their test performance respectively; whilst students' attitudes towards test quality had no significant effects on their test performance.

The findings of this project imply that students are capable of providing meaningful and valuable information for testing; furthermore, whether students' values and beliefs have been appropriately expressed in tests, especially in test uses, may have a significant impact on how they perform in the test.

Applying the Dynamic System Theory to an analysis of test stakes

广东外语外贸大学 亓鲁霞

The stakes of tests are considered to be the origin of washback, a topic often prioritized in language testing research. For all large-scale high-stakes tests, due to their great impact on teaching and learning, the educational system, and even other sectors of society, it is imperative to carry out washback studies. Such studies are especially desirable when a new test is developed or existing tests are being reformed. In the past two decades or so, washback studies have mushroomed and considerable achievements have been made. Nevertheless, the stakes of tests have been largely neglected as evidenced by few studies reportedly examining the concept of test stakes. In studying test washback, researchers preferred Madaus'(1988) distinction between high-stakes tests and low stakes tests in designating a target test, but few attempts have been made to enquire into the construct of stakes attached to a test being studied. This is probably one of the reasons why effective measures are still lacking to check negative washback and reinforce positive washback of language tests. Such a situation calls for some illuminating theories to guide in-depth studies of test stakes. In this paper, the Dynamic System Theory (DST), which has been shown to be insightful for applied linguistics studies, is drawn on to analyze the construct of test stakes for a better understanding of the origin of washback and to inform test reform and future washback studies.

Could the consequences of test be removed?

-- Perspectives from Chinese EFL students of tertiary level

Shanxi University HAO Chenqian CHEN Qiuxian

It is now a consensus that improper use of test results in the educational context could bring about serious consequences to teaching and learning in the classrooms. A question arising is: would removal of the improper uses eliminate these consequences? If so, in what degree? If not, why? This study seeks to answer these questions with data from the area of English as a Foreign Language (EFL) and the Chinese higher educational context.

College English Test (CET4 and CET6) is a large scale standardised testing system in China. It was incepted in mid-1980s and has been used ever since for the evaluation of students' achievement in College English, a compulsory EFL program designed especially for non-English major undergraduates in Chinese higher educational institutions. The CET testing system is influential in China not only because it is administrated nationwide to students numbered up to hundreds of thousands each year, but also because of the uses that its results have been put to. Specially, most Chinese universities had prescribed passing CET4 as a requirement for bachelor degree certificate, and CET6 for master degree certification. Moreover, the CET4 and 6 certificates are a knocking stone for employment opportunities. These stakes, as research reveals, have entailed severe washback effects onto the teaching and learning practices in College English classrooms and resulted in phenomena such as “teach to the tests” and “learn for the tests”. They are also regarded as responsible, in part at least, for the “low efficiency and effectiveness” issue of College English education (literally translated from Chinese mandarin 费时低效 *fèishídíxiào*, means spending a lot of time, yet achieving little effect).

Beware of the seriousness of the problem, the Higher Education Commission (HEC) of China Ministry of Education launched a series of moves to innovate CET testing system. For instance, more items that are not of Multiple Choice Question type and an oral test, CET-SET (Spoken English Test) are included. Also, a test result is given to test-takers (regardless of pass or fail) instead of the original certificate in the case of pass only. On the other hand, in a 2005 press release, HEC denounced the connecting practices of bachelor degree with CET4. The issue of test result paper rather than a certificate is also intended to discourage the abuse and misuses of CET test results in other occasions. While the intention of removing the high stakes is clearly stated, questions remain: To what extent, if yes, will the removal of these uses eliminate the negative influences that the CET testing system has had on the College

English education?

A survey involving 600 College English students, that's, about 10% of the policy target group in an inland Chinese university was conducted to address the questions. Co-relation and regression analysis of the data indicated the strong grip of the test power regardless of the policy change. Social weightings and other sociocultural and historical factors are interacting and influencing the Chinese students' EFL learning.

Why do school policies on CELST test preparation differ ?

--- An investigation into the washback of CELST through examining school administrator attitudes and school policies

东莞理工学院

张放

The Computer-based English Listening and Speaking Test (CELST) is the latest reform in NMET Guangdong. Listening and speaking proficiency were never a compulsory item in previous NMET test papers. Also, the computer-assisted means of test administration places new requirements on test preparation. Therefore it is intriguing to investigate how middle schools prepare for this special test. Moreover, existing washback research mostly examine test effect through investigating the teaching and learning behaviors geared towards a test, yet little attention has been offered to the school administrators who make the school policies that pose important micro context for test preparation. This study explored on the washback of CELST through an investigation conducted in six different levels of schools in two areas of Guangdong Province. The special focus was the school administrators' attitudes towards CELST as well as the school-based measures issued to prepare for this test. The major instrument was in-depth individual interviews. Guided by the Expectancy-value Theory (Wigfield & Eccles, 2000), this study devised a coding scheme to analyze the qualitative data. Finally, it was discovered that the school administrators' subjective values of CELST and their expectation of success in the test led to discrepant school policies in CELST test preparations. The finding proves the hypothesis on the existence of washback variability (Alderson & Hamp-Lyons, 1996) and demonstrate the feasibility to employ the theories in other field of study to explain washback phenomenon.

剑桥商务英语考试的测试影响——以考生为视角

重庆大学

黄妍琪 肖巍 赵路 辜向东

作为鉴定商务英语水平的权威指标，剑桥商务英语证书/考试（Cambridge English: Business, BEC）在商务英语教学中扮演重要角色。BEC对考生英语学习产生的影响，是其考试效度的一个重要维度，值得关注。为此，本研究采用问卷法，于2014年调查了重庆、浙江、广东和安徽四地共837名BEC考生，拟回答以下问题：（1）考生对BEC有何认识？（2）BEC对考生学习过程有何影响？（3）BEC对考生学习结果有何影响？

基于问卷数据，本研究发现：

（1）考生对BEC的认识总体正面，只是个别方面略偏负面。考生对考试形式比较熟悉，但对证书的使用欠了解；内部动机非常强烈，求职需要等外部动机也较为强烈；认为BEC的时间分配合理，题量和难度合适，但听力和阅读略难。

（2）BEC对考生学习过程的负面影响很大，也有一定正面影响。负面影响表现为应试倾向明显，如严重依赖往年真题练习、辅导材料、背单词、背模板等。正面影响为学习强度增加，听、说、读、写全面展开。

（3）BEC对考生的学习结果，包括四项基本技能、词汇量和商务知识等，产生了非常正面的影响。此外，BEC对考生的综合素质，如自主学习能力等，亦有正面影响。

本研究的主要贡献为：理论上，尝试提出了基于商务英语考试的测试影响框架，廓清了通用测试影响理论的适用性和局限性；实证上，丰富了学界对BEC测试影响的认识；教学上，为提升BEC考试质量、改进商务英语教学提供了有益反馈。

学术素养培养为导向的语言测试反拨效应研究

河北经贸大学

高霄

学术素养(academic literacy)指学习者在语言学习中能够分析文本结构(structure)、阐明逻辑推理过程(logical development)、识别文本思想(general understanding)、总结主旨大意(main idea)、辨别文本细节(specific details)、区别观点与事实(opinions and facts)、思辨性分析(critical reading)与鉴赏学术词汇(academic vocabulary)等。学术素养培养教学范式区别于传统的语言教学范式,强调“自上而下”(top-down)的教学理念,即教学过程中遵循“篇-句-词-字”的逻辑顺序,尤其偏重篇章能力的培养。

本研究旨在回答三个主要研究问题:(1)学术素养是否具有可学性;(2)学术素养培养为导向的语言测试构念效度如何;(3)此种语言测试范式对教和学是否具有反拨效应。

本研究被试来自某省属重点骨干高校,共近300名,其中有效样本为252名;样本为方便样本;男生62名,女生190名;任课教师为两名;实验期为1个学期。量具包括每堂课学习者所写反思日志和期末考试试卷。日志包括三部分,即G(Gains)、P(Problems)和S(Suggestions);期末试卷包括三个题目:(1)总结主旨大意题(summarizing),考察学习者把握重要信息与归纳的能力;(2)篇章构建推理类问题(inferring),考查学生分析文本结构与逻辑推理能力,涉及辨识各种逻辑关系、做出推理、理解暗示意义以及判断论证的逻辑性;(3)分析评价类问题(evaluating),考查学生对文本作者的观点、论证过程、写作目的和语气的把握和对文章的评价能力。

研究表明:(1)学术素养培养具有明显的可学性,经过一个学期的教学实践,学习者的学术素养都得到显著提升,达到统计学意义;不同语言水平学习者的进步程度有所差异。(2)学术素养培养为导向的语言测试具有较高的构念效度,测试较好地测量了学习者的学术素养。(3)此语言测试理念对语言教学有较明显的反拨效应,在一定程度上影响到语言教师的教学行为和学习者的学习行为。

Tracking students' change of language ability during their preparation for NMET

广东外语外贸大学 王华

National Matriculation English Test (NMET) in China exerts strong washback effect on both high school teachers and students. NMET drives EFL teachers to equip themselves with test preparation strategies and the students, guided by their teachers' test preparation strategies, sharp their ability to get items correct through mountains of exercises when NMET is approaching. However, whether the students' endeavors significantly improve their scores in NMET is yet to be determined. Several studies have indicated that test preparation can increase test takers' scores. A close examination of those studies found that most of those studies used raw scores to indicate test takers' abilities in both the pre-test and the post-test. As the difficulty levels of items or tasks in the pre-test and post-test were not in a common frame of reference, the scores obtained in the pre-test and the post-test lacked comparability, thus the conclusions those studies drew were doubtful. This study used a two-tier item factor analysis model to estimate the item parameters of three mock NMET tests that a group of NMET candidates attended in different times when NMET was approaching. The two-tier model treated NMET candidates' abilities in the three mock tests as three correlated primary dimensions and treated testlets in the three mock tests as unrelated specific dimensions. The estimated parameters were then used to calibrate NMET candidates' language abilities in the three times. The results showed that overall NMET candidates' scores significantly improved when NMET was approaching. The analysis of students in different proficiency levels indicated that high-level students gained most, with an effect size of 0.393, while low level students benefited the least, as their scores in the three times were not significantly different. The study also investigated NMET candidates' performances on different parts of NMET mock tests and the results revealed that the students gained most in the reading comprehension part, with an effect size of 0.095, while the students gained least in the cloze items, with an effect size of 0.075. The results raise concerns regarding the direction of the washback effect

Room 523

Investigating the Practice of the CEFR outside Europe: A Case Study on English Writing Assessment in China

University of Southampton & Wuhan University
ZHENG Ying, ZHUANG Yanyan, YAN Youyang

Since its publication in 2001, the Common European Framework for Languages (CEFR) has been gradually understood and applied in various contexts by learners, teachers, assessors, and researchers alike (e.g., Mader&Urkun, 2010). However, most of the previous studies on the practice of the CEFR focused on Europe (e.g., Buckland, 2010; Figueras&Noijons, 2009), and there is much less discussion on the use of the CEFR in Asian contexts, particularly in China, one of the largest ELT markets. In order to explore the practicality of the CEFR in China, this project aims to investigate the possible applications of the CEFR in Chinese ELT classrooms at the tertiary level.

Through a survey and semi-structured interviews, the project investigated Chinese ELT teachers' current knowledge of the CEFR and then further familiarized them with the CEFR rating scales through trainings provided by invited CEFR experts from the UK. The project also explored the applications of the CEFR in China and provided the Chinese ELT teachers with hands-on practice of using the CEFR to assess university students' English writing across different levels according to the CEFR scales. Comparisons of the rating scores were undertaken between CEFR experts and Chinese ELT teachers. In addition, through a post-activity interview with the Chinese ELT teachers, the researchers carried out an evaluative examination on whether the CEFR scales are applicable to the Chinese context and whether any amendments are needed on the illustrative descriptors.

Results demonstrate that there's a potential in introducing the CEFR scale into the Chinese ELT context. With proper training of getting the teachers familiar with the CEFR scale, ELT teachers showed satisfactory rating consistency using the two scales. Taking into account of teachers' positive opinions of utilizing this scale to complement what they already have, it is safe to conclude that getting the teachers more familiar with the CEFR scale could potential bring positive impact on their teaching and evaluation practices.

The preliminary results of this study are meaningful in the sense that it explored an approach to transform CET scores to the CEFR scale. With more data collected and results confirmed, a transformation table could be created to enable CET-4 stakeholders to interpret their scores in relation to the CEFR can-do statements and relevant descriptors.

An Investigation of Chinese English Majors' English Language Competence with Reference to the CEFR

Chongqing University YAN Qiaozhen, GU Xiangdong, YU Chengyuan

The Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR) has become highly influential in language education in Europe and beyond. This study intended to investigate Chinese English majors' English language competence with reference to the CEFR. The European Language Portfolio (ELP) self-assessment corresponding to the CEFR was employed to explore English majors' language competence from seven aspects: Listening, Reading, Spoken Interaction, Spoken Production, Strategies, Language Quality and Writing. A total number of 235 English majors from four cohorts of a comprehensive university in southwest China were investigated. The language competence of each cohort and the language proficiency development among the four cohorts were first explored with reference to the CEFR and then their strengths and weaknesses in language competence were further analyzed with reference to the CEFR.

This study was supposed to make contribution to the knowledge of English majors' language competence within the current educational context by using an international language competence framework, the CEFR, and to shed light on the development of a common measurement scale of language competence in China.

Aligning Chinese EFL learners with the CEFR

Guangdong University of Foreign Studies PENG Chuan

The present study is aimed at aligning Chinese EFL learners with the levels of the Common European Framework of Reference for Languages (CEFR), given the current status of the CEFR as a common standard across different educational systems worldwide and the increasing significance of English education in China. Qualitative and quantitative methods will be adopted to answer the following questions: 1) Which CEFR sub-scales and descriptors are chosen to assess Chinese EFL learners? 2) Which Chinese EFL learners are mapped onto a certain CEFR level? To answer the two research questions, curricular objectives, syllabus objectives, course materials, major English proficiency tests are compared with the CEFR descriptors to identify the relevance of the CEFR categories and descriptors to every English

education level and estimate the English proficiency range of teaching and learning in relation to the CEFR levels. And then, the chosen CEFR sub-scales are distributed in either a web or paper questionnaires covering primary, secondary, and tertiary level English education in the form of self-assessment and teachers' assessment respectively. Their responses in the two assessment contexts are collected and analyzed to make a final decision on which English learners are located at a certain CEFR level with the help of IRT model. Implication on the test alignment with the CEFR, English teaching and learning , appropriateness of CEFR practice in China and alignment of Common Chinese Framework with the CEFR will be given.

Sentential Context Priming in Word Association

华南师范大学 张晓蕾 张萍

As a type of testing, word association has a profound history. Consensus of L1 mental lexicon's development has been reached. But the results of L2 mental lexicon researches show the lack of paradigmatic pairs (Bahns&Eldaw 1993; Howarth 1996; Granger 1998; Nesselhauf, 2005; Ping Zhang, 2010).The commonality is that primes are shown isolated in the test. Context plays an important role in L2 vocabulary learning .Related researches on context which is specifically sentence context are mainly in the field of lexical access and word recognition(Simpson, 1989; Besner, 1992; Stolz, 1995; Jiang, 2005; Dyck, 2007; Starreveld, 2014).Sentential context is more concise comparing to discourse and more complete comparing to chunk and phrase. However, few studies discuss word association in context. V-N collocations (Verb in left-position, noun in right-position) are used as prime separately trying to help subjects get more paradigmatic pair. The present study focusing on sentential context priming in word association tries to answer: 1)are there any differences between the results of decontextualized WA and sentential contextualized WA; 2)whether the stimulus position will affect the results; 3)how will different proficiency of L2 learners perform in different contexts.

A mixed design of experiment is used in the present study, 2 (context types: traditional WA and sentential context WA) * 2 (stimulus position: left position-verb priming and right position-noun priming) * 2 (proficiency) . Stimuli and sentences are chosen from British National Corpus (BNC), the Corpus of Contemporary American English (COCA), the Chinese Learner English Corpus (CLEC) and the Leading Outline of NMET of English. E-prime is used as the tool of testing. Senior high Grade Three students are chosen to be low proficiency subjects and students of GradeThree in English major who have already passed

TEM4 are in high proficiency group. Experiment One is the traditional word association. In Experiment Two, sentences are shown in a "Maze Task" (Forster, 2009). Maze Task used here is to eliminate the memory effect. After filler sentences, filler task which are totally irrelevant to the test follows. Word association follows the experimental sentences. Subjects will have interview after the test to ensure whether context is activated and is there memory effect. The present study hope to further develop the experiment to ensure context effect in priming thus to provide more evidence of L2 mental lexicon and suggestion for L2 teaching.

Effects of the Formulation of CEFR Descriptors on Scale Comprehensibility and Calibration

Shanghai Jiao Tong University ZHANG Xiaoyi ZOU Shaoyan

The way in which scale descriptors are formulated affects users' understanding of scales directly, especially for language teachers and learners who have little knowledge about language scales. As a user-oriented language scale (Alderson, 1991), the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) claims its descriptors to be "definite, clear, brief and independent" (Council of Europe, 2001). This, however, has been doubted over the years (see Alderson *et. al*, 2004; Weir, 2005a). Despite a few exceptions (e.g. Negishi, 2011; Paparougiou, 2010), empirical studies have seldom been focused on issues with the CEFR descriptors in a non-European context..

Based on the Dutch classification scheme of CEFR scale descriptors (Alderson *et. al*, 2006), the present study investigated how well the CEFR reading descriptors from B1 to C1 levels were understood in the context of tertiary-level English education in China. In order to examine the extent to which the difficulty levels of descriptors are affected by the formulation of these descriptors *per se*, 135 college English teachers were invited to align China's College English Curriculum Requirements (Ministry of Education, 2007) with the CEFR levels, and 412 college students were asked to conduct self-assessments using the band descriptors. Moreover, a workshop on the calibration of these descriptors were held among 17 experts with a professional background in language testing. Data were analyzed using SPSS 19.0 ad Winsteps 3.90.0.

Results from descriptive analyses and Rasch analyses yielded the following reasons for the mismatching or inconsistency in calibration: 1) a lack of text features or text type in the descriptors, 2) the inclusion of reading strategies or presence of conditional terms in the descriptors, and 3) a lack of elaboration of such verbs as read, understand, etc. These findings reveal the theoretical deficiency of the CEFR in cognitive validity and context validity. The present study also delivers practical value to the editing work of band descriptors in scale

development.

An Empirical research on Correlation of Language Elements and Language Level in Net Text

Beijing University of Technology HU Wenting, HE Cencheng

This empirical research introduces the correlation of language elements and language level applying in college English net text.

According to a university scale net text total sample, this research acquired a positive correlation coefficient between language elements and language level applied in net text. Therefore, it achieves an automatic evaluation of English listening, speaking, reading, writing, and translating in utilizing intelligent net text. The result proved the approach of utilizing positive correlation coefficient of language elements and language level is effective and efficient in language formative assessment of net text. Furthermore, this approach indicates the superiorities of low cost, effective and convenient to operate, which has an actual significance in promoting and facilitating university paperless net text of total language elements.

Ethics in Language Testing: Review and Implication

中南财经政法大学 甘凌

Language testing is usually high-stake, for it involves various stakeholders like test users and test takers who make or are made decisions based on the scores of the tests. Henceforth, the decisive role of language testing on individual's lives must involve issues about ethicality. This article initially explained the relationship between ethics and validity, holding that ethics is one of indispensable aspects in the test validation in which an ethical test is a valid one. Then the article presents and discusses the existing researches concerning ethics in language testing from two aspects: testers and tests themselves, from which it is concluded that researchers have not yet come to agreement on how much responsibility the testers should take for the tests and test uses. Nevertheless, what matters is the testers should make professional judgement by applying codes of ethics and codes of practice into test practices. It's also concluded that the empirical studies of applying codes of ethics and practice are still

rarely found nowadays. In view of the discussion, the article elucidates the implications for researches about ethics in Chinese language testing context: 1. to establish and implement codes of ethics and codes of practices in Chinese language testing contexts; 2. to gradually realize the professionalization of language testing; 3. to rethink the ethical problems from a critical perspective. It is expected that the review and implication in this article can draw more attention to ethics in language testing for the sake of fairness.

Room 714

Design a Rubric Aligned with Intended Learning Outcomes

to guide Business Case Study Presentation

Fujian Normal University PENG Yuanling

This paper describes how a rubric aligned with intended learning outcomes designed, tested, revised as an indicator to guide and assessing student business case study presentation (BCSP), using Biggs' "Constructive Alignment" model. BCSP is an activity which integrates language competencies and proficiencies with real problem-solving capabilities in business situations. The purpose of this study is to know what criteria and what rubric is to aligned with the ILO can effectively guide students learning and assessment in BCSP. Broad review related to rubric design and development was explored. The criteria were defined from the ILO of Business English course objectives by the guidance of "Constructive Alignment". Aspects related to the validity of the rubric developed were investigated from the instructors and undergraduate students from the department of Foreign Languages, 2011Grade, at Fujian Normal University, Fuqing Branch, China. Upon pilot testing, the rubric was found to be a valuable, objective and ultimately timesaving grading tool.

Exploring the Comparability of TEM4 Reading Tests and English-major Students' Textbooks in One of China's International Studies University

Nanyang Technological University LI Tian

In language testing, establishing the usefulness of test tasks is an important step in test validation. Construct underrepresentation and construct-irrelevant variance are considered to be two major threats to the usefulness of test tasks and the validity of tests. Test developers and researchers are primarily concerned with these two aspects of test development and test research when examining test usefulness.

The present study aims to evaluate the usefulness of TEM4 (Test for English Majors 4) reading test tasks through exploring the comparability between TEM4 reading texts and the texts in two textbooks used by sophomore English-major students in terms of difficulty level. 32 reading texts from eight TEM4 reading tests (2005-2012) and 28 texts from two textbooks (A New English Course and Contemporary College English) were selected as the primary dataset.

Both TEM4 reading texts and the texts from the two textbooks were analyzed using quantitative and qualitative methods informed by Green et al.'s (2010) framework. Following Green et al. (2010), 27 linguistic characteristics of all the data were evaluated from the following five perspectives: (1) vocabulary and grammar, (2) cohesion and explicitness of rhetorical organization, (3) genre and rhetorical task, (4) subject and cultural knowledge, and (5) text abstractness.

For quantitative analysis, three automated textual analysis programs (i.e. Coh-Metrix, Wordsmith Tools, and Web VocabProfiler) were used to evaluate 21 linguistic characteristics of all the texts. Based on the output results of the textual analysis programs, independent samples t-tests were conducted to compare the linguistic characteristics of TEM4 reading texts and those of the texts in the two textbooks respectively to determine whether there were any significant differences in terms of any linguistic measure. For qualitative analysis, human raters were employed to assess the other six linguistic characteristics of all the texts. Based on the rating results, inter-rater agreement was calculated. With Likert scale data, Mann-Whitney U tests were conducted to see if significant differences existed between TEM4 reading texts and textbook texts.

Based on the results of quantitative and qualitative analyses, significant differences were found between the texts in the two textbooks in terms of V8 (Academic Word List level). Significant differences were also revealed between TEM4 reading texts and the texts in the New textbook concerning V1 (Average characters per word), V8 (Academic Word List level), G2 (Average sentences per paragraph), G5 (Mean number of words before the main verb of

main clause in sentences), and R1 (Flesch Reading Ease score). However, no significant differences were detected between TEM4 reading texts and the texts in the Contemporary textbook for quantitative measures. Combining the quantitative and qualitative results as well as other indices provided by Coh-Metrix, TEM4 reading texts seem to be more challenging for students compared with the texts in the two textbooks. Moreover, the texts in the Contemporary textbook were found to be more comparable to TEM4 reading texts and pose greater challenge to students in comparison with those in the New textbook.

From rigidly copying to flexibly implementing: lessons learned from an institutional-level assessment change

The University of Hong Kong ZHOU Jiming

Assessment innovation has been put on the educational agenda of many countries in recent years. In Chinese higher education, the move towards assessment for learning (AfL) is being positioned to enhance English language programs. Previous studies indicate the absence of some literature-espoused AfL practices in Chinese college English classrooms. However, the empirical question remains as to how tutors and undergraduates interpret and enact assessment change in their local context. Metaphorically speaking, this study aims to slice through the layers of the innovation onion, to reveal the agentive spaces in which tutors and undergraduates interpret and translate a top-down initiated assessment change.

This study selects an information-rich case—a university in Shanghai which moved beyond the rhetorical engagement and actively engaged itself in a change towards AfL. Three tutors and eight groups of students (n=45) were interviewed at two different times over one academic year. Forty lessons were observed and audio-recorded. Data were analyzed following an inductive coding procedure adapted from the qualitative analysis protocols established by Miles and Huberman (1994).

This study finds different forms of peer assessment implementation, ranging from peer grading for summative purposes in the first semester, to the more flexible peer feedback and online peer assessment in the second semester. Students generally perceived peer grading as “nonsense and unfair”. In contrast, they believed that peer assessment was able to facilitate learning, when they were involved in giving, receiving and using feedback. The three tutors interpreted different difficulties for their classroom implementation of peer assessment. The varied interpretations were influenced by their assessment literacy, and subsequently impacted on the way they planned, framed, conducted and used peer assessment.

This study carries theoretical implications by proposing a continuum of peer assessment supported with concrete classroom examples. It shows the continuity between peer

assessment and the broader assessment for learning activities. Three dimensions are found crucial to understanding peer assessment, i.e. purpose, criteria and feedback. The dimension of purpose extends from summative purposes to formative purposes. The other two dimensions are to what extent students are shared with assessment criteria and involved in giving feedback to their peers. At the practical level, this study highlights some tensions in the promotion of western-derived principles and practices of peer assessment in Chinese higher education, and proposes some cultural-appropriate ways for translating assessment change in exam-oriented contexts.

Assessment of Reading Comprehension: the case of Chinese Language in Singapore

Nanyang Technological University Chan Kwong Tung

How reading comprehension testing are carried out to assess students' language achievement and proficiency have been a highly debatable topic in language assessment. Whether Chinese Language (CL) is positioned as a Second Language (L2) or a Mother Tongue Language (MTL), it has direct implications for teaching, learning and assessment of reading comprehension in schools. Whatever is the position, the cognitive development of students' learning should not be compromised in language teaching and learning. Using examples from the secondary CL curriculum and assessment, this paper describes and illustrates how Bloom's Taxonomy (BT) of Educational Objectives could be applied to the teaching and assessment of students' CL reading comprehension in Singapore. While BT is used in CL lessons to enhance students' cognitive processes during learning, it is also commonly used in tests and examinations to assess students' reading comprehension skills, as well as to check on their level of cognitive development. It is, therefore, also serves to check on the alignment between the learning outcomes and assessment so that the quality (or validity) of the reading comprehension testing can be ensured. Besides, BT can also provide an assessment framework that guides language teachers in going beyond basic cognitive ability (knowledge, comprehension and application) to include higher order thinking skills (such as analysis, synthesis and evaluation) in day to day teaching.

This paper starts off with a brief review on BT and the nature of reading comprehension. Then, it illustrates how BT is conceptualized in CL reading comprehension teaching and assessment in Singapore. Here, a framework showing the relationship between BT and the types of reading is suggested to guide teachers. With this simple framework, the paper continues to show how it is used in everyday teaching, as well as in setting reading comprehension test and examination papers. This paper focuses on CL to illuminate universal

and specificity in the writing-system and style aspects of reading comprehension. It seeks to describe and explain how BT is applied to the CL reading comprehension, as well as to ask how research on CL reading comprehension in Singapore can inform universal reading comprehension science.

A Comparative Study on the Influences of Different Types of Integrated Writing

Tasks on Chinese EFL Learners' Source Use

Southern Medical University BIN Ke

Guangdong University of Foreign Studies ZHOU Yan

Integrated writing tasks, having authenticity and being similar to the process of real-world academic writing, has been attracting language testers' and researchers' interest increasingly during recent years, accompanied by a concern for the role of the source material. Integrated writing tasks require participants to write an essay, argumentation, summary, etc, based on the input source material, either in the form of reading passage or in the form of listening material, or both. The present study investigate into the influences of different types of integrated writing tasks on the amount and type of EFL learners' source use by analyzing 120 participants' performance in an English writing test which is consisted of a reading-to-write task and a reading-listening-writing task. Results show that bi-modal input in reading-listening-writing task can elicit participants' verbatim source use compared to the monomodal input in reading-to-write task. High-level participants tend to comprehend and then restate information from the source material more in both integrated writing tasks while low-level participants prefer patchwriting. What needs to mention is over 60% of low-level participants copy the reading material directly in reading-listening-task. Limitations and suggestions are presented in the end.

Unguided assessors along the way---A case study on tertiary EFL teachers'
conceptions and practices of assessment in one college from China

The University of Auckland 姚雯静

Contrary to the increasing importance attached to EFL (English as a foreign language) teacher education in China, a shocking negligence of education on language assessment theory and methodology---one of the most important aspects in preparing assessment literate teachers of English still exists in China's teacher education. One of the direct consequences is that most teachers of English still adopt a dominating summative assessment method and mentality despite the vigorous efforts of China's Ministry of Education to diminish the effect of national English examination and promote assessment for learning or formative assessment. Thus this study employs the qualitative research method of interviews with eight EFL teachers in one college and analysed the interview results with a thematic inductive method to probe into tertiary EFL teachers' extant knowledge and practices of assessment and identify the challenges and possible chances for change. Based on Constant Leung's infrastructural theory of teacher assessment research, the analysis of the findings revealed that: 1) assessment education and training is cursory for these teachers; 2) the advocating of national policy takes little effect on the transformation of teachers' knowledge and practices of assessment, and 3) the assessment practice held by the teachers is formative by appearance but non-formative in essence.

A Research and Design on the MTI Interpretation Teaching

Resources Platform

重庆邮电大学 陈璇

With the rapid development of technology in computer science, network and multimedia, the traditional teaching mode has been gradually transformed to the network teaching mode. Therefore the demand for network teaching resources is becoming increasingly strong. Although there are abundant of teaching resources on the internet, most of them are overlapped, scattered disorderly and inconvenient in use. Therefore how to strengthen the organization and management of these network teaching resources has become an important issue of great concern.

Based on the current study of teaching resources database, a deep analysis and research was carried out on the knowledge structure and the required interpretation skills of the MTI interpretation students. Furthermore, interpretation teaching methods and their corresponding requirements on the teaching resources was also studied carefully. All of the above study and research contribute to the final design and implement of the interpretation teaching resources platform.

The structure and content of this thesis is organized as follows: The first chapter introduces the background and significance of the paper; the second part provides literature study and the research situation on teaching resources database at home and abroad. Chapter three includes a deep analysis and study on the knowledge structure and the required interpretation skills of the MTI interpretation students, as well as the interpretation teaching methods and their corresponding requirements on the teaching resources. Chapter four is the general the design and implementation of the interpretation teaching resources platform and chapter five focuses on the experiment and questionnaire of this platform. Chapter six draws a brief conclusion of this thesis and introduces the plans of the future study.

Room 715

Impact of Integrating an Automated Assessment Tool into English

Writing on Teachers

McGraw-Hill Education

WANG Yihong

The study investigated how the one-year experimental teaching that integrated an automated writing assessment tool into English writing impacted on the teachers and their teaching. Data were collected from open-ended questionnaires, interviews, and journals. It was revealed that the teaching focus shifted from attending to language forms to content exploration and expression, from the learners' writing product to their writing process and pedagogy. Teachers' beliefs in teaching writing, their practices, and roles also changed. The assessment tool acted like a catalyst for reform. The main force of the changes, however, was the teachers, their learning, and their effective integration of the tool into their teaching of writing.

The Impact of Automated Writing Evaluation on EFL Writing

Tianjin Normal University MA Rui, GU Gang

The system of automated writing evaluation (AWE) provides feedback of the vocabulary, syntax, semantics and discourse presented in the writing submitted online by the students. Most of the current studies focus on the construct validation of the AWE from a purely statistical perspective, which cannot cover all factors which play roles in the impact of the AWE on the EFL writing. The current study aims to connect teacher's pedagogical practices with student's perception of learning effectiveness.

Our research tool is the TRP, a system of automated writing evaluation developed by the Higher Education Press. The three college EFL classes at different levels were chosen as subjects. The data collected include the individual interviews of the three English teachers of the classes, the individual interviews of the three students from each class, and the questionnaire answered by some students from the three classes.

Our concern is how the TRP was used in their classes, how the teachers felt about the TRP, and what factors affected their perceived effectiveness of using the TRP. In addition, students' responses to the effectiveness of the use of the TRP are also analyzed.

The preliminary finds show that different ways of integrating TRP into teaching do make differences, for example, the way of using scores, guiding students' revision through the feedback provided by the TRP, presenting model essays, and assigning group tasks, etc.

Designing an automated essay scoring system

City University of Hong Kong Roxanne Wong

This work in progress will report on an attempt to design an automated essay scoring system for the inclusion of task related vocabulary and organizational discourse markers, and source use. The system was designed using Python.

Python is an open source operating system that is functional on a variety of different computers. It is accessible to the novice programmer. I have had no training in computer programming and have shied away from mathematics and the technical sciences. If I am successful in pursuing the above goals then I feel most anyone in the field of linguistics can do the same. Information will be provided to aid the novice programmer who wishes to develop their own program.

The researcher is a linguist and English as a second language instructor and is currently undergoing an entry into programming in Python. This program will be used to help students improve their English language development. Student opinions on the program will be discussed as well as teachers views on the usefulness of it. Feedback from conference participants on areas for further improvement will be greatly appreciated.

Computerized Integrated Writing: Introduction and Application

University of Electronic Science and Technology of China, Zhongshan Institute

City University of Macau WU Jinyu

Better than paper-and-pencil test, computerized adaptive testing can accurately predict candidates' ability, greatly improve test accuracy, and be suitable for Chinese-context English test. Computerized adaptive test can check the actual capacity of the candidates; item difficulty can be adjusted according to the actual capacity of the candidates; the ability of the candidates is not only based on the numbers of correct answers, but also the item difficulty and response time, greatly improving measurement accuracy, which is not easy to achieve for the traditional paper-and-pencil test.

Except for the current functions of computerized adaptive testing, this study also takes an unconventional approach, trying to focus on the computerized integrated writing based on the latest research of foreign countries, which is the first time to apply in domestic language testing field, hoping to lead to in-depth discussion. Two softwares --Bilog and Gitest3+ -- were adopted in this research for item analysis on the test data. Bilog was the latest Rasch

Model based Test Estimating System and Gitest3+ was a product developed by experts of GDUFS. Single-parameter Rasch Model was adopted for equating pretested items and maximum likelihood algorithm was applied for estimation. Following the above mentioned steps, which were technically the core of this research, the calibrated items were compiled into an integrated writing test and imported into a self-made CAT item bank. Integrated writing was a new type of test integrating latest research findings overseas and specific needs of teaching and learning and testing of college English in China. Two sets of data would be collected in this research--one on integrated writing in the form of CAT, and another on CET band 4, both from a same group of students. The research hypothesis was that the two sets of data would be proved highly correlated, which is to say, integrated writing in the form of CAT excels in rationality, precision, and scientificity and accords with the direction of development of language testing.

Confirmatory Factor Analysis of Speaker's Oral English Level Scale Based on Identity

Dalian University of Technology LIU Li, GAO Ying

This paper mainly explores whether speakers can change their oral English strategy based on their identity. It uses exploratory and confirmatory factor analyses of data elicited from speakers to explore and examine factorial structure of speaker's oral English level scale based on identity. Confirmatory factor analysis shows that five-category model generated by the exploratory factor analysis fits for the data in this study. The findings of the study suggest that five-category structure is relatively stable. The scale includes situation identity strategy, status identity strategy, institution identity strategy, social linguistic strategy and other pragmatic strategy.

跨文化交际能力测量工具研究综述

深圳职业技术学院 杨洋

高质量的测量工具是跨文化交际能力评价的重要保证。本文首先从跨文化交际能力的构念 (construct) 出发, 确定了跨文化交际能力的组成, 即知识、感情 (态度) 和行为 (能力) 三个方面, 然后从性格测试、文化敏感性测试、知识测试和文化情境测试等四个方面介绍了国外常用的跨文化交际能力工具, 并就国内的跨文化交际能力测量工具进行了梳理。文章指出, 跨文化交际能力的评价涉及语言学、传播学、心理学、工商

管理等许多领域，不同领域的研究关注的重点不同。国内的跨文化交际能力评价研究大多是在语言学习的框架中探讨对跨文化交际能力的测试与评估。研究者在关注本领域研究重点的同时，应该打破学科限制，多关注和借鉴其他领域的相关研究。国内评价跨文化交际能力的研究很多，有不少学者使用了各式各样的测量工具，有自编的测试，也有借用国外比较成熟的量表，但很少有研究关注到测量工具本身。自编的测试是否能准确测量被试的跨文化交际能力？国外的量表是否适合中国被试？这些都有待研究。“工欲善其事，必先利其器”。在使用测量工具评价跨文化交际能力的同时应该多关注测量工具本身的研发及质量控制。

Sunday Sessions

Room 301

Theoretical backing and practice reasoning of listening ability scale

浙江大学 HE Lianzhen CHEN Dajian MIN Shangchao

Drawing on the experiences of working on the “national English ability scale” project, this paper will review several core theoretical and practical issues concerning the construction of listening ability scale. On the theoretical side, this paper primarily discusses the definition and theoretical model of listening comprehension, including issues about cognitive process of listening, listening strategy and use-oriented listening ability model. On the practical side, this article mainly discusses the methods for descriptor collecting and editing, including document searching, descriptor sampling and the semantic structure of descriptors. Through intensive discussion of the inner mechanism of listening ability scale and deep analysis of the construction process, this paper aims to enhance understanding of the listening ability scale and facilitate research and practice in English teaching and assessment.

翻译能力描述语开发研究

——兼谈中国英语能力（翻译）量表子项目设计

黑龙江大学应用外语学院 严明 冯莉

中国英语能力量表项目是教育部正在组织进行的一个重大项目，事关中国未来的外语教育改革事业。翻译能力量表项目作为其中一个子课题，具体研究内容如下：

- 一、 在回顾中国翻译教学与评测现状基础上，设立中国英语能力（翻译）等级量表开发目标。
- 二、 建立翻译能力等级量表研究的理论框架、工作原则、方法与流程。梳理国内外关于翻译能力研究的相关成果，建立适用于本量表的翻译能力构念；建立三大语料库作为研究基础；研制翻译能力等级量表制定的原则，包括以交际能力语言为基础，采用基于运用的方法，采用“能做”描述法，设计翻译能力量表参数体系，确立典型翻译活动等；设计翻译能力量表开发的研究方法，包括语料库方法、调查法、定性研究法、定量研究法、直觉法以及描述语收集与修订的相关方法。
- 三、 设计翻译能力量表方案，建立描述语库。在理论研究和原则制定的基础上，

设计翻译能力等级量表整体方案；设计翻译能力“能做”描述语的语义与句法规范；运用各种方法从各种来源收集描述语，建立描述语库；调查对象包括翻译研究专家、资深翻译教师与翻译学习者等，评估其可行性；走访翻译公司主管、翻译从业人员与各院校翻译教师，通过实地抽样调查法采集一手资料，确立典型翻译活动，从多来源收集相关描述语。

四、 实施各种验证过程，使中国语境下的翻译能力等级量表化。对翻译能力描述语进行合格判断、审核、修订之后，将在课题组内外部、全国相关院校进行大规模分类与分级信度、效度验证。通过交叉验证和多方验证，统计分析数据，不断反馈与修订描述语，最终实现翻译能力描述语的量表化。

文献研究表明，国外无汉英翻译能力量表相关成果，而国内缺少以“能做”方式描述翻译能力的描述语。本项目的研究重点与难点首先就是建立以“能做”描述语方式描述翻译能力的描述语库。本文在回顾了项目研究背景之后，重点探讨中国语境下开发汉译英“能做”翻译能力描述语和建立描述语库的过程。此项工作填补了国内外研究空白，并且对今后翻译能力的教学与评测都具有重大实践推广价值。

Scaling exemplary writing activities: An SFL text typology perspective

Guangdong University of Foreign Studies PAN Mingwei

Shanghai International Studies University ZOU Shen

As anchoring labels in a can-do-statement language proficiency scale, exemplary activities, such as *writing dairy* and *listening to news program*, serve as parameters against which language learners across different proficiency levels can be judged (in)capable of certain performance. In the context of writing proficiency scale, contingent upon a range of linguistic resources as well as varied degrees of familiarization with text styles, learners are assumed to perform a good number of writing activities that seem exemplary in various real-life domains. However, whether, and if so how these activities can be stratified on a language proficiency scale remains to be explored. The Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001) lists a limited number of exemplary writing activities, such as *writing a report* and *writing a memo*, yet literature seems scant regarding how different activities can be corresponded to different language proficiency levels. As part of the contribution to the *China Common Framework for English Language Proficiency* program initiated and funded by National Education Examinations Authority under Ministry of Education, PRC, this study, informed by text typology of Systemic Functional Linguistics (SFL, Halliday & Matthiessen, 2004), rested upon the existing descriptors of English writing proficiency for a broader landscape of exemplary writing

activities.

Based on a document analysis approach, the present study first sampled approximately 1000 descriptors of various contexts and uses for English learners. A bottom-up approach was then adopted, where the descriptors that were found explicitly embedded with writing activities were extracted, e.g. *writing a thesis* and *writing a composition*. All the extracted descriptors were neatly categorized into three domains: (1) activities for personal or social purposes; (2) activities for educational purposes; and (3) activities for business or occupational purposes. Informed by the SFL text typology, the researchers further scaled the descriptors, where more sophisticated text types, e.g. *writing a recommendation letter (to recommend)*, were presumed to be exemplary for higher-level achievers; whereas less sophisticated ones, e.g. *writing a poster (to entertain)*, fell into exemplary activities for lower-level achievers. Although certain text types were found exemplary across different proficiency levels, a consideration of text quality achieved by learners might be necessary. In order to triangulate the scaled exemplary writing activities, the researchers of this study conducted semi-structured interviews with language assessment experts and teaching practitioners in China so that modifications and refinements could be made possible.

From an SFL text typology perspective, this study enriches the existing literature of exemplary writing activities; in particular, these activities can be incorporated into English Proficiency Scales for Chinese learners and also lend empirical support to in-class activities and teaching material development. Nevertheless, more cross-validations, such as alignment studies on exemplary writing and speaking activities, need to be further explored in the future.

中国大学生英语口语互动能力等级量表研究

上海立信会计学院外语学院 唐莉莉

The research, borrowing the ideas from Common European Framework of Reference for Languages and Cambridge Practice Tests for oral English, studies the formulation of interactive scale in the context of the Chinese college English learning, aiming to improve the awareness of interaction in college English education. Based on College Learners' Spoken English Corpus(COLSEC), it conducts the conversational analysis of oral interaction among candidates on different English proficiency levels, thereby building the interactive scales for Chinese English learners. The research focuses on the section of the group discussion, decodes the materials systematically with the tool of Conversation Analysis, combines both the qualitative and quantitative analysis to summarize the interactive characteristics, upon

which the interactive scale will be formulated and the relevant washback effect will be studied in College English oral classes.

A Comparative Study on English Pronunciation Descriptors

Used Abroad and at Home

南京大学 程欣 陈桦

This study compares the English pronunciation descriptors used in rating scales and course syllabi abroad and at home, in the hope of finding a way to develop the English pronunciation descriptors appropriate to be used in the Chinese setting.

The comparison between the descriptors used in rating scales abroad and at home presents the similarities in language form and differences in constructs. Both of them tend to adopt general wording, such as “accurate pronunciation and intonation” in China, and “being intelligible” in English-speaking countries. Neither provides explicit instructions to nonnative raters, to whom intelligibility is as unclear as accuracy due to inter-language intelligibility benefit. The descriptors in syllabi, however, demonstrate the great importance attached to phonological awareness and phonics on the beginner level both abroad and at home, but with more emphasis on the pragmatic functions of prosody in the English-speaking countries.

In addition, the descriptors in rating scales and in syllabi show, to some extent, the disagreement between the constructs in English teaching and testing. In English teaching, the descriptors center around segmental accuracy and the native likeness of prosody, while in English testing, intelligibility and communicative function is the core construct. This might be understandable as a native model is needed in English teaching, but not in English testing when English is used as lingua Franca today.

On the Collection of Descriptors of China Common Framework for English Language Proficiency (Interpreting)

广东外语外贸大学 赵护林 王巍巍 许艺 穆雷

The collection of descriptors comprises an important part of the development of China Common Framework for English Language Proficiency (CCFELP). The article first introduces the process of collecting CCFELP descriptors for interpreting, including the training, the extraction of descriptors and their refinement. Then it mainly focuses on the components and the source of the descriptors, and their classification and placement. The components of the descriptors contain verb, topic, interpreting mode, organization and discourse type (Descriptor=verb+topic+interpreting mode+organization+discourse type). The descriptors primarily are selected from teaching syllabus and plan, test syllabus, curriculum, textbook, existing language proficiency scale, and interpreting research literature etc. The chosen descriptors are classified and placed into specific parameters like identification and use of language A, and levels like B3 and C2. It is hoped that the descriptor pool can lay a foundation for the development of CCFELP for interpreting.

The Development of EFL Reading Strategy scale of Chinese College Students

广东外语外贸大学 周艳琼

The slowdown of the students' reading performance in 1970s made researcher to reflect the teaching problem, since then, research and theory in second language (SL) education have shifted from examining the methods of teaching to investigating the processes of learning.” (Purpura, 1997), the researcher found that teaching should shift from teacher-centered to student-centered which made the research of strategy became popular. Reading comprehension ability as part of literacy competence is vital for developing academic language proficiency. though reading skill research was the first major interest of cognitive psychologists in 1960, in the 1980s reading strategy research became popular, a great number of researcher have indulged in reading strategy research. But till now there is no empirically validated hierarchy of reading strategy. Oxford (1990) states “there is no complete agreement on exactly what strategies are, how many strategies exist; how they should be defined, demarcated, and categorized; and whether it is ---or ever will be---possible to create a real, scientifically validated hierarchy of strategies...”this study is a try to present some answers to Oxford's question which will provide L2 researchers with revealing insights into in the field of EFL reading strategies.

A language proficiency scale defines a series of ascending bands of proficiency.

According to North(2000), when developing a language proficiency scale, two issues have to be addressed, one is related to the description issue, another is related to measurement issue. In order to develop Chinese college EFL reading strategy scale, the two issues have to be addressed. This paper mainly discuss how to address the description issue, then the ways to develop the EFL reading descriptor pool are stated as well.

Room 302

A Validation Framework for the National English Proficiency Scale of China

Shanghai Jiao Tong University ZHU Zhengcai ZHANG Lidong

This study attempts to establish an appropriate framework to validate the newly developed National English Proficiency Scale of China (NEPS), which is a scale of the English language learners' proficiency in Mainland China. Given the long history and broad scope of the term "validation", the study defines in the first place the conception of validation within the context of validating a language proficiency scale other than a test. Distinctions between a validation study of a scale and that of a testing program are identified and explicitly stated. Three categories of validity evidence encompassing interrelated facets are employed to construct a theoretical framework for validating the NEPS in current study: (a) construct validity and fair validity addressing the fundamental validity issue from the perspective of science and ethics, efficiency and fairness; (b) teaching backwash validity and social impact validity explicating the relationship between the NEPS and English language teaching, and a broader societal impact the NEPS exerts in China; and (c) decision validity, a higher level validity index, dictating the ultimate value of a psychological scale as a useful measurement instrument for decision making. Along with the three categories, concurrent validity and delay validity are considered in this proposed validation framework to collect validity evidence over time, both during the development of the NEPS and after its formal launching. The study further elaborates on the specific methods used for collecting multiple sources of validity evidence pertaining to each stage of the NEPS development. In particular, validity evidence would be primarily collected from six sources: (1) validity related to construct and its descriptive parameter system; (2) validity related to descriptors; (3) validity related to band levels; (4) validity related to fairness; (5) validity related to teaching backwash and social impact; and (6) overall validity related to decision-making. This study aims to validate the NEPS adopting the proposed validation framework, which by far is pioneering in the field of

English language testing in China. Therefore, the proposed validation framework is more of a tentative version than a fully-fledged one. More efforts are needed to improve the adequacy of this validation framework by means of integrating the opinions of various stakeholders, investigating empirical data, and subjecting it to careful scrutiny.

An EFA-based Study of Simulating Construct Validity of Objective Listening Test

中国矿业大学 李廉

Construct validity of a language test gives an index to test-takers' language traits which are measured through the test. Mapping the construct validity of established language tests can provide meaningful reference for writing a new test or building a item bank. The present study attempts to explore the feasibility of writing a new test by simulating the construct validity of a given standardized test, through extracting common factors from its test-taker's scores. Taking objective listening items of a IELTS test as criterion reference, this study applies exploratory factor analysis (EFA) to 388 test-takers' results, and draws the distribution of common factors from their scores on each item as a "item-construct" two-way checklist. A new listening test, in accordance with the checklist, is written and then taken by 1200 college students. A second independence EFA on results of the new test shows that the distribution and accumulated contribution of its common factors are significantly correlated to the criterion-reference test.

Pearson product-moment correlation analysis on the same test-takers in the two tests shows that the two tests are statistically correlated in terms of reliability index. Internal equity of the new items, calculated by Cronbach's alpha, proves to be standard and consistent with that of the criterion-reference test. Furthermore, coefficient of overall difficulty of the two tests are similar to each other.

It is concluded that it is feasible to extract common factors as constructs from given language tests and apply the construct model to writing a new test by simulating established tests or items. This criterion reference method provides comparatively exact planning of construct validity, reliability and difficulty of new test.

An Investigation into the Context Validity of CEFR Writing Descriptors

Shanghai Jiaotong University ZOU Shaoyan ZHANG Xiaoyi

The Common European Framework of Reference (hereafter CEFR) has gained worldwide attention and triggered intense academic debate ever since its inception. As Weir (2005a) pointed out, the CEFR scales are premised on an incomplete and unevenly applied range of contextual variables or performance conditions. That is, the descriptor scales fail to fully recognize how contextual variation may affect performances and the difficulty levels of ‘can-do’ statements. Such insufficiency in contextual parameters can jeopardize the context validity of descriptors.

To address this concern, a contextualized study was carried out to examine how well CEFR writing descriptors (B1-C1) can describe and distinguish college students’ English writing proficiency in China. With reference to the theory of context validity (Weir, 2005), the study employed both questionnaires and semi-structured interviews to elicit information from 135 college English teachers and 412 students. SPSS 19.0 and Winsteps 3.80.0 were used to run the descriptive analyses and Rasch analyses respectively.

Results were presented mainly from three aspects: discriminant power of descriptors, appropriacy of writing activities and accuracy and transparency of descriptors. The discriminant power and reliability of descriptors at the B1 level were generally acceptable, while a blurred borderline was identified between B2 and C1 levels, especially among descriptors addressing grammatical competence. The typical writing activities proposed by the CEFR, according to our study, were less applicable than expected to the context of tertiary-level English education in China. The vagueness of terms and wording of the descriptors, on the other hand, also hampered the participants’ understanding to some extent. The current study reveals the flaws of CEFR writing descriptors (B1-C1) when applied to the context of China, and therefore calls for a locally-developed writing scale that adequately and properly takes into consideration the contextual variables of English education in China.

The comparison of content validity study on cloze tests between the NMET of Guangdong and Nation (2011-2015)

华南师范大学 李乔娜

Cloze test is a necessary part in NMET in China, and content validity usually stands as an important indicator of the quality of a good test, which makes it meaningful to explore the content validity of the NMET. Based on the framework of task characteristics proposed by Bachman and Palmer and the theory of testing point validity proposed by Li Xiaojun, according to New National Curriculum Standards for Seniors and test syllabus, a practical assessment framework for the evaluation of the content validity of cloze tests is presented, upon which the content validity analysis of cloze tests of NMET (Guangdong) and NMET (Nation) from 2011 to 2015 is undertaken.

The study indicates that:

1) In terms of text per se, NMET (Nation) boasts more kinds of text types, its TTR and average deletion space do not see significant changes; In contrast, the text types of NMET (Guangdong) remain relatively fixed, besides, its TTR and average deletion space do not change much but keep constantly less than those of NMET (Nation); both NMET (Nation) and NMET (Guangdong) are rich in text topic and steady in readability.

2) In terms of testing point validity, both NMET (Nation) and NMET(Guangdong) cloze tests contains the largest proportion of items at discourse level, while the proportion of NMET (Nation) is larger. Neither of them consists of items at phrase or word level; Concerning focus factor, both NMET (Nation) and NMET (Guangdong) pay much attention to testing meaning.

3) In terms of items, NMET(Guangdong) sets its testing points within nouns, pronouns, numbers and prepositions, while NMET(Nation) covers not only those mentioned above, it also tests related phrases; Concerning distracters, NMET(Guangdong) outperforms those of NMET(Nation), for the consistency its forms and meaning domains.

According to what discussed above, some suggestions are given in hope to improve the content validity of cloze test in NMET in the future.

The Comparative Study of the Validity between the Senior High School
Entrance English Examination and the TOFEL Junior Standard Test —— An
Evidence-based Approach

重庆师范大学 杨秀芬

The purpose of this study is to provide some positive guidance for the reform of English examination in China by comparing and analyzing the context validity, the test takers' characteristics and cognitive processes, and test scores between the Senior High School Entrance English Examination and the TOFEL Junior Standard Test according to the validation framework put forward by Cyril J. Weir (2005). In the process of study, testing, think aloud protocol and questionnaire was employed to gain data about the test takers' cognitive processes in taking the two tests, and Rasch analysis to analyze their test scores. To be specific, 60 junior three students in a Junior High School in Chongqing were recruited as the subjects of the research. Before the test administration, the physical conditions were checked to ensure the context validity. Then, the 60 subjects did the two English tests respectively, and 6 randomly sampled students recorded their cognitive processes and strategies used while finishing the same type of task in the two tests to investigate whether their cognitive process and interactional activity are authentic. Thirdly, the students' scores in the two tests were collected for Rasch analysis to verify their reliability and the correlation between the two tests. Finally, all the 60 test takers did the questionnaire to know more about their physical, psychological and experiential characteristics involved in the process of performing the two English tests.

The comparison of content validity study on reading comprehension between the
NMET of Guangdong and Nation (2011-2015)

华南师范大学 张晓丹

Reading comprehension takes a relatively large part in NMET, as for a good test, content validity is one of the most important criteria. Validating a reading comprehension and producing a reading comprehension test of good quality will ensure the success of NMET.

The author studies the reading comprehension tests of NMET (Guangdong and Nation)

from 2011 to 2015 with the related theory of content validity. To start with, the author establishes a new framework based on Bachman & Palmer's framework of language task characteristics, and then combines with the requirements laid down in National English Curriculum Standards and the syllabus of NMET; the author analyzes the content validity of the reading comprehension tests in the recent five years from the aspect of testing constructs, testing rubrics, testing input and the expected response, the conclusions are as follows:

First, the genre and topic of the reading passages are both relatively rich, and narrations are mainly used; as for NMET (nation), four kinds of genre are used each year, while in NMET of Guangdong, there is no use of practical passages, and they are tested in another testing method called information matches; Besides, the most frequent used topic is about social life and culture, meanwhile the passages about science are increasingly more;

Second, the length of passages is controlled within a reasonable range, but the passage is relatively longer in NMET of Guangdong, which requires students' vocabulary and reading amount more;

Third, in terms of passage difficulty, the readability of NMET of Guangdong is slightly lower, so it is more difficult for students to read;

Four, as to the skills examined, the items examine "understanding specific information" take the largest proportion, but there is no item that examines "understanding the discourse structure" proposed by the syllabus of NMET whether in NMET of Guangdong or NMET (nation) from 2011 to 2015.

Through analyzing, the author also finds that nearly all the items of practical passages in NMET (nation) test students' ability of "understanding specific information", which is too easy. Besides, the interfering force of some options is too weak in some items, students can easily exclude by their common sense or background knowledge. At last, the author raises some relevant recommendations for the design of reading comprehension tests as well as the teaching and learning of reading in senior high school.

A Study on the Context Effects in Word Association Task

华南师范大学 龚晨 张萍

Word association test is a feasible and economic way of probing learners' mental lexicon (Lambert, 1956; Fitzpatrick 2000,2003). A number of studies have investigate the context effects on native language word association and found that in congruent context, subjects had higher probability to produce the target words (Mckoon&Ratcliff, 1995;Zeelenberg, 2003). While In the studies of second language word association tests, in most cases, cue words were

presented alone, few has investigated into the context effects on word association. So this replication of Zeelenberg (2003) is intended to study the context effects on the word association test of different levels of ESL learners and try to answer 1) whether the EFL learners will be affected by the context biasing the interpretation of the ambiguous target words like the native English speakers; 2) whether different levels of EFL learners will produce different word association results under the context effects. The experiment follows the pattern of Zeelenberg (2003). The whole experiment is programmed by E-prime. The second grade postgraduates and second grade senior high students are chosen the subjects so as to compare the outcomes of WA of different levels of ESL learners. The results of the pilot study performed by second grade postgraduates shows great disparity in the results between the English native speakers and EFL learners (Riegel1972; Meara,1978). Compared with native English speakers, even high-level EFL learners produces more nonsemantic responses, part of which are affected by the culture differences. This result suggests the material used in Zeelenberg 2003 extracted from the Word Association Norms (Nelson& Schreiber, 1998) cannot be directly applied to the EFL learners. Another pilot study is needed to select the materials during which the frequency issues will refer to the criteria of Zeelenberg (2003). Meanwhile the experiment materials should match the subjects proficiency and culture background, which is also the difficult part of this study. Based on the results of Zeelenberg (2003) and the pilot study, the author assumes the context will exert effects on the results of word association test of EFL learners, and the effect will be more salient for high level EFL learners.

Room 311

标准设定应用于大学英语听力成就测试的可行性分析

张敏 三峡大学外国语学院

语言测试结果的解释和使用是整个测试工作的重要组成部分。有效的语言测试能够对分数进行有意义的解释、公平的使用。标准设定是测试使用的重要环节，依据它所做的决策会对学生产生不同程度的影响。一直以来，它被认为是高风险评估的一个核心问题。成就测试属于基于标准的评价。这种评价依据一系列预先设定的、描述不同阶段的评估标准来评价学生的学习，给出学生的成绩。它一般在学期末课程结束时举行，主要用于检验教学效果，它会影响学生是否能够通过这门课程的决策。因此，它的标准设定关系到测试是否能被有效的使用，能否帮助做出正确的决策。标准设定方法众多，但研究者们已意识到，没有完全客观的标准设定方法。纵使其再详尽再系统，都离不开人的主观判断，因此必须检验分数线的有效性。目前，大学英语听力测试的效度方面普遍

存在这些问题：将成就测试的标准等同于教学内容；语言能力标准设置的主观性；分数线的设定的主观性；将成就测试的使用混同于水平测试等，这些问题不同程度的影响了测试的有效使用，需要采取相应的措施，才能保证标准设定应用于大学英语听力成就测试的可行性。本文通过回顾标准设定的概念及意义、标准设定的方法及其评价，借鉴一些测试实践中标准设定方法的运用，分析是否能将其用于大学英语听力课程评估，以期进一步完善当前的课程评估体系。

美军国防语言水平测试体系建设及启示

解放军理工大学 张锦涛 吴苓

进入 21 世纪,美军在海外采取的数次重大军事行动彰显懂外语的军事人才对国家安全的极大价值,同时也从血的教训中认识到语言障碍对美军有效履行全球反恐任务所产生的巨大负面影响,为此把官兵的职业外语能力上升到“克敌制胜的关键作战能力”。为使美军能够更有效应对全球反恐、地区维稳、国际人道主义救援、联军行动等形势,维护国家利益,2005 年,美国防部制定了《国防语言转型路线图》,2011 年,美国防部又出台了《国防部语言技能、区域知识和文化能力战略规划:2011-2016》,旨在进一步强化美军具备履行岗位职责所需的职业外语能力。为推进国防语言转型建设,美军成立了国防语言办公室,构建了国防语言水平测试系统(DLPT),按熟练程度将外语水平划分为生存型、技能型、专业型、专家型、母语型五个不同等级,明确各等级听说读技能具体标准要求官兵按不同岗位达到相应标准,依托国防语言学院(DLI)对美军展开大规模职业外语技能培训,并对军人开展常态化的外语能力测评,美军目前掌握的语种高达 376 种。中国和平发展之路面临复杂、多元的安全威胁与挑战,军队要切实履行维护国家利益和世界和平的使命任务,遂行涉外多样化军事任务将日益增多,我军新型军事人才的岗位任职外语素质必须与新的任务要求相适应,必须构建以岗位任职需求为指向的外语能力培训与评测体系。本文从三个方面分析了美军国防语言水平测试体系建设的内涵、功能和特点,并从四个方面论述了构建具有我军特色的外语测评体系的必要性。

国内语言测试公平性研究的现状与对策

武汉理工大学外国语学院 肖奕 吴宜静 王凯旋 朱汉雄

自中国自80年代改革开放以来,我国对英语的重视程度日益提高,各种方式的英语测试层出不穷。作为我国高等院校招生中大规模选拔性考试的重点科目,语言考试一直是选拔人才的有效途径,而考试的公平性乃重要要务。然而在一系列相关的规章制度下,我们仍旧不能忽视隐藏在考试中存在着的 unfair 的现象。

在国内的语言测试界,关于研究测试公平性的理论探索很少,但从各方面研究语言测试公平性的文献很多,相关的实证研究是围绕考题对于不同考生群体的差异效应,如高考、大学英语四、六级、专四专八等(王后雄2011[1],涂晴2012[2],邹申2011[3])或者从测试偏差方向入手,如偏差的定义、来源、主要侦测方法等方面具体的分析(李静2009[4],孔文2008[5],张琨[6])。

本文将结合国内语言测试的研究现状,对一系列可能存在考试当中对公平性造成影响的因素及其产生的原因、存在的区域做出相关的总结和判断,如何在已经存在的这些不公平因素中,最大程度的减少不公平现象,更好的体现考试的公平合理性,使英语测试更好地发挥其作用进行一系列的讨论。

本文把这些不公平因素概括为三大类:存在于考试前期的不公平现象、考试中的不公平现象,以及存在于考试后的不公平现象。这些现象产生将分化为几大小类:考试前期,不同制度,地域,性别的学生群体对同样的题型所产生的不同反应,使结果出现了偏差所造成的不公平现象;同等学力的学生对不同类型题型的掌握情况存在差异所产生的不公平现象;考试中,由于作弊等现象的发生,以及考场监管系统不力所引起的对考试公平性的直接影响;考试后,由于批改阶段存在评价性的偏好,如在作文中,对某类话题不同查阅人可能产生的对公平性的影响。以上问题我们都将在本文中进行详细的研究并探讨相应的对策。

独立学院 ESP 测评体系探究——以金城学院为例

Nanjing University of Aeronautics and Astronautics LIU Xiurong

Recent years have witnessed a large-scale development of ESP courses, especially in independent colleges that orientate themselves toward cultivating practical talents. Nonetheless, ESP evaluation lags behind the course development. Consequently, it is essential to explore the framework of ESP evaluation. Based on Multi-intelligence Theory, the Concept of Higher Education Theory, Modern Quality Management Theory, and Constructivism Educational Theory, ESP assessment should feature diversity, external quality, and place an extra emphasis on learners' study process and autonomy. The assessment framework is formed by ESP learner assessment, ESP teacher assessment and ESP course assessment. In

project-based ESP teaching, the assessment for ESP learners combines formative assessment and summative assessment, complete with self-assessment. The former evaluates the accomplishment of project-related tasks and autonomous learning in computer-assisted learning centre. The latter takes the form of pen-and-paper tests, evaluating ESP learners' specific language knowledge, skills and learning strategies in simulated situation of workplace. ESP teacher assessment is conducted by the evaluation body, including ESP learners, peer specialist teachers, experts and ESP teachers themselves from the aspects of teaching attitudes, teaching methodology, teaching process management and teaching effect. ESP course evaluation entails answering the following four questions: whether learner needs have been fulfilled, whether the learning materials and tasks are authentic and practical, whether the course has successfully fostered learner autonomy, and whether the course has favorably promoted teacher development.

A Review of English Speaking Test Studies in China over the Past Three Decades

Chongqing University of Science and Technology

YANG Zhiqiang YANG Ruijin MENG Lei

Based on the articles published in ten CSSCI journals in foreign languages in China between 1986 and 2015, this paper reviews 126 representative studies on English speaking test by using multi-phase sampling. The analysis shows that 1) there has been a sharp increase in the number of research papers on speaking over the past three decades; 2) empirical studies increase every year and the number has exceeded that of non-empirical studies; the former mainly employs quantitative analysis; the methods of collecting and analyzing data tend to be diversified; 3) with regard to subjects, more attention has been paid to undergraduate students, of whom non-English majors are the main focus; 4) in terms of research content, much importance has been attached to speaking test tasks/methods and validation. However, there exists some problems, such as 1) the research of this field does not receive due attention, as it takes up only 4‰ of the total published papers; 2) non-empirical studies are inundate with literature review, the introduction of test or test system and comments on tests but few about theoretical exploration; 3) empirical studies are of imbalance, such as the dearth of exploration on English speaking proficiency of high school students or the matriculation English speaking test; current researches scarcely cover test-taker characteristics and rating criteria/English speaking ability scale; the research method primarily lies in analyzing test results while ignores the process of test taking and that of raters' rating; the validation studies

are partially carried out; the shortage of longitudinal studies are obviously felt; 4) there is repetition of similar research. Besides, this paper makes some predictions about the directions towards which English speaking test studies in China are likely to move forward.

关于制定国家外语考试质量标准的研究及建议

教育部考试中心 程蒙蒙

2014年9月，国务院颁布《关于深化考试招生制度改革的实施意见》，明确提出要“深化高考内容和形式改革，高考外语科目一年提供两次考试机会”，外语考试的质量和公平成为社会关注的焦点。然而作为外语考试大国，我国尚未有统一的国家外语考试质量标准。国际上已有一些成熟的语言测量标准可为我们提供借鉴。国际测试质量标准充分到了考试开发者、被测者和使用者等各方面考试利益相关者的权利和利益；既注重测试公平，又包含测试质量监控机制；提倡信息公开，维护考生权利。而我国现有的考试标准仅关注测试链某个单一环节的管理，缺乏测试公平性原则和人性化服务；考试信息不够充分、透明，易造成测试结果的误读误用。为了保证我国教育考试的科学性与公平，需要尽快由国家主导制定统一的测试质量与公平性，对外语考试设计、实施和使用的各个环节进行规范与检测。在借鉴国际成熟经验的同时，要充分考虑中国国情，进行整体方案设计，明确我国外语考试质量标准的服务群体、覆盖内容，以及如何制定和实施，集中解决“为谁做”、“做什么”和“怎么做”的三大核心问题，即1. 兼顾测试主办、研制和实施者，考生和测试使用者三类人群各方面的权利；2. 既做公平原则，又做专业行为规范；3. 由国家主导，广泛征求意见，多部门配合，定期修订。要以“以人为本”的教育理念为指导，增进人性化服务，为依法制考提供理论和技术支撑。

Room 313

The Ineffability of Language Ability and Philosophical Considerations in Language Testing

Guangdong University of Foreign Studies HUO Yongshou

With validation of language testing as its point of departure, this paper attempts to elucidate the ineffability of human language ability, the ability for a language user to use her language for different purposes, and on this basis consider language testing from a philosophical perspective.

The paper starts by making the claim that validation, especially the validation of the construct of a language test, is of absolute importance to the testing and assessment of language ability, including the design of a language ability rating scale. In this case, the validation of a language test or a language testing system has to presuppose a sufficient and complete definition and description of the language ability involved as the object of the test. But the problem is that such a definition and, especially, a description are impossible because the language ability to be tested and examined is unsayable: it cannot be exhaustively described linguistically.

From the perspective of philosophical ineffability, this essay analyzes human language ability at the metaphysical and empirical levels. The analysis finds that either from the Kantian thing-in-itself perspective or from the Zhuangzi/Quine art-and-skill perspective, human language ability as an object in the ultimate reality is unsayable or ineffable: it cannot be ultimately described, especially when human language itself is mobilized as the metalanguage for its description. Stated another way, the analysis concludes that the effability of human language ability is relative while its ineffability is absolute, a philosophical conception of the nature of human language ability.

On this basis, the paper examines language testing philosophically from the metaphysical, epistemological, socio-cultural, ethical and political perspectives. As a conclusion, the author maintains that, as implied and logically necessitated by the effability of human language ability, the functioning of language testing at the above-mentioned levels suffers from its own logical limitations and therefore different consequences although it can really provide its interpreters and users with uses and applications of different kinds. This, as emphasized by the author, is an eternal danger to be guarded against by interpreters and users of language testing.

Decoding and Reconstructing of Textual Signs

— Bakhtin's Dialogism and Discourse Significance of a Text in Standardized Tests

Nanjing Normal University ZHU Youyi

Dialogism is the ideological foundation and theoretical core of Bakhtin's theories of polyphony and metalinguistics, which combines the two closely. This article intends to analyze the discourse significance of a standardized testing text from the perspective of dialogism, and it focuses on dialogism among subjects, which a testing text is imbued with. Testing subjects' decoding and reconstructing processes on a testing text fully indicates the significance of dialogism. This realization enables us to have a deeper insight into the impact of non-standardizational factors on the validity of a standardized test, and, more importantly, to propose new approaches to the improvement of test validity.

文化资本视角下的外语测试研究

上海外国语大学 赵辑

目前关于外语测试的已有研究以测试技术研究为主, 强调以提高测试的信度与效度保证其有效性和公平性。但随着测试技术的提高, 测试的权威性得到认可, 对外语测试证书误用日益严重, 其社会权重越来越大, 造成的各种社会影响与日俱增, 而这些问题仅仅依靠测试技术提高是无法彻底解决的, 需要从社会学角度加以分析和解决。

本文试图以社会学家布迪厄提出的文化资本论作为理论基础, 通过三个步骤的分析, 探求外语测试对社会及个人外语学习产生影响的过程和原因, 进而探索解决问题的方向和方法。首先, 外语能力属于布迪厄的文化资本论中对文化资本三种形式中的具体状态, 而外语测试则属于体制化状态。外语测试证书可以对证书拥有者瞬时的外语能力赋予相对永恒的价值保障。第二, 外语测试证书的相对永恒性和外语能力的相对瞬时性, 使得两者具有了相对独立性。随着时间及各种因素的改变, 外语能力可能产生衰退, 然而证书却不会因此而失效, 因此两者便具有了相对独立性。第三, 外语测试证书与外语能力的相对独立性越明显, 两者剥离程度越高, 越使得外语学习者在对学习资源分配时, 更倾向于投资周期短、投资回报率高、并且具有永恒收益的外语测试, 而非投资周期长、收益时效短的外语能力提高。

最后, 根据问题产生的过程和原因, 本文从降低外语测试证书所具有的资本特性对解决方法进行分析, 探讨如何降低外语测试证书与外语能力的相对独立性, 降低其社会权重让外语测试回归其基本功能。

社会语言能力相关文献综述

湖南师范大学 盛艳玲

根据 Bachman 的交际语言能力 (Communicative Language Ability), 社会语言能力是指对特定语用环境中语言使用规则 (conventions of language use) 的敏感性或控制能力, 即说话者在某种特定语言环境下能够将话语、意图和语境相联系而构成意义的某种交际能力。作为交际能力的一个重要组成部分, 社会语言能力是外语学习者在交际行为中不可或缺的一项能力。在建立中国英语学习者社会语言能力等级量表中, 社会语言能力的定义将直接影响社会语言能力等级量表的维度划分, 因此, 对社会语言能力定义及内涵的相关文献研究变得及其重要。本文将综述 Chomsky 的语言能力观、Hymes 从社会文化角度的交际语言研究、Canale 和 Swain 在 1980 年所提出交际能力理论以及 Bachman 系统阐述的交际语言能力 CLA 理论。其中, 本文将讨论 Chomsky 语言观中语言能力和语言行为划分的发展和不足, Hymes 语言交际能力中语言可行性和语言得体性的研究、Canale 和 Swain 模型中对社会语言能力的定义及 Bachman 在交际语言能力模型中对社会语言能力的定义的继承与发展。其中, 本文将重点研究 Bachman 交际语言模型中, 社会语言能力的定义和所涉及的语言使用的特定环境并与环境相适应的能力, 包括对方言或语言变体 (dialect or variety) 差异的敏感性、语域 (register) 差异的敏感性、自然地道语言 (naturalness) 的敏感性, 也包括对特殊文化参引 (cultural references) 和修辞格 (figures of speech) 的理解能力, 以期为社会语言能力等级量表的维度研究提供借鉴。

从描述语角度探究跨文化交际能力的构成要素

湖南师范大学 陈灿

本文主要探究跨文化交际能力的构成要素, 并用收集到的描述语检验该框架的有效性。跨文化交际能力是指交际者在特定环境中与来自其他文化的成员进行得体、有效交际所需具备的能力。得体性和有效性是跨文化交际能力的核心所在这一观点得到许多学者的认同。但对于跨文化交际能力的构成要素, 不同的学者从不同的角度出发, 提出跨文化交际能力包含二至七个构成要素。因此, 有必要对跨文化交际能力的构成要素进行探讨, 以得到实际有效的跨文化交际能力构成框架。本文探讨以下两个问题: 1) 跨文化交际能力的构成框架中有哪些要素? 2) 该框架的效度如何? 本研究采用文献收集法、定性法和定量法探讨跨文化交际能力的构成框架, 并用搜集到的描述语检验该框架的有效性和实用性。研究表明: 1) 跨文化交际能力包含认知、情感和行为三个构成要素。认知主要指跨文化意识, 涵盖自我意识和文化意识两个方面; 情感指跨文化交际的敏感性, 即在特定的情形中交际者的个人情绪或偏好; 行为指交际者实施交际行为和完

成交际目标的技能。2) 跨文化交际能力构成框架建立后, 作者选取若干条有代表性的跨文化交际能力描述语对其进行检验。结果表明, 包含认知、情感和行为三要素的跨文化交际能力构成框架的效度较高。本研究旨在将跨文化交际能力构成要素理论与实践相结合, 通过教学与测评实际来验证理论。

An Analysis of Semantic Structure of “Can Do” Descriptors for Textual Competence

Hunan Normal University LIU Qian

As an explicit marker of textual coherence, cohesion plays an important role in constructing the internal meaning of a text. Under the background of developing an *English Competence Rating Scale for Chinese Learners*, this study, based on Halliday’s Cohesion Theory and Anderson’s Revised Language Objectives Taxonomy, tries to analyze the semantic structure of “Can do” descriptors for English textual competence. The semantic structure of “Can do” statements has a basic expression—descriptive index (can do sth.) + influencing parameters (under certain condition). Literature research is used to set the descriptive index. First, widely collect relative descriptors from curriculum standards, test syllabus and language competence scales. Then analyze the frequency statistics of their verbs. With these data and the six cognitive skills proposed in Revised Language Objective Taxonomy, cognitive verbs are finally confirmed. In the same way, cohesion is processed and classified into four dimensions in accordance with Cohesion Theory: reference, substitution, conjunction and lexical cohesion. Similarly, influencing parameters are defined with documentary reference. One of the most important factors is text types, which restrict cohesion in a certain kind of text. Integrating different viewpoints, text types of this thesis are narrative, explanatory, argumentative, indicative and applied texts. Next is context, giving description of the external environment where a text and cohesion stay. There are context of situation outside the language and context of culture in the common consciousness or speech community of interlocutors. The research shows, the semantic structure of “Can do” statements for textual competence is: can + cognitive verb + cohesion knowledge + text type + context. Descriptive index is the core meaning and influencing parameters work as a frame. Besides, this structure needs to be positive, precise, clear, brief and independent, five principles for compilation. This paper aims to offer guidance for the development of descriptor bank of English textual competence for Chinese learners.

Room 316

理工类英语课程形成性考核评价体系的构建与实践——以南方科技大学为例

南方科技大学 徐嗣群

长期以来,我国各类学校的英语教育一直都是一应应试教育,尤其是目前高校理工类英语考试,过分注重总结性评价在一定程度上弱化了考试功能的正常发挥和教育目标的实现,这不仅影响了高等教育质量,同时也影响了广大学生的学习积极性,阻碍了高素质理科人才的培养,存在诸多弊端。近年来,人们越来越清醒地认识到素质教育、创新对国家、对民族的极端重要性,因此改革当前的考试模式成为摆在高校面前急需解决的一个重大课题。

本研究将构建课程形成性考核评价体系,强调学生的学习过程考核,通过本体系去评价学生的智能水平、创新能力、进取精神和人格品质是否达到教学目标和人才培养目标的要求,实现教学目标,培养出高素质的理工专业本科学生。

本研究拟确立考核指标体系,实施单课程形成性考核;形成考核结果分析及反馈制度。任课教师在某一单项考核结束后应及时进行总结分析,并将分析结果及时向学生反馈,这不仅可帮助教师了解教学效果,改进教学方法,提高教学质量,更重要的还可以帮助学生了解自身的学习情况,改进学习方法,提高学习效率从而体现形成性考核评价的真正价值。通过本项目的探索与研究,试图总结出一适合目前我校理工类英语课程实际的形成性考核评价体系,作为一项学校制度去执行,以测验学生的基础知识,检测学生的学习效果;掌握教学情况,改进教学方法和督促理科人才培养目标的实现。

Action Research on Formative Assessment of Research Project Course

School of Foreign Languages, Southeast University YANG Maoxia SHAO Zheng

Based on Constructivism Learning Theory, Inquisitive Learning Theory and Academic English Teaching Practice, Research Project is a course designed in Southeast University to develop students' communicative skills, to cultivate their research ability, and to improve their ability of making an academic presentation and writing a research report in English. During the four weeks of this course, students are required to work in groups to complete a research task using English as their working language. Since this course is quite different from a traditional College English Course in terms of its teaching objectives, modes and procedure, the assessment system should be modified to ensure the achievement of its teaching

objectives.

The original assessment system of this course emphasizes the final oral presentation of the research. Although it recognizes the importance of formative assessment, what should be assessed in the process and how should it be assessed are not clearly stated. Therefore, after several rounds of implementation, it is found that 1) some students focus on the final oral presentation instead of the research itself during their study; 2) language proficiency, rather than communicative skills, research skills or other abilities, is the determinant element in getting a high score in this course; 3) it's hard for teachers to supervise the learning process and many students fail to manage their time properly during their study.

In response to the above-mentioned problems, the writer formulates the hypotheses that a clearly-structured comprehensive formative assessment system could be of help. To work out such an assessment system, two rounds of experiments are conducted in the year 2013 and 2015.

In the first experiment, 4 assessments are added to the course: Research Topic Evaluation, Research Proposal Evaluation, Mid-term Checking and PPT Skeleton Frame Evaluation. Both the students and the teacher grade and comment on the work of each group. In the second experiment, the 5 basic skills which the students need to complete their research are allotted to each group at the beginning of the course. With the help the learning materials provided by the teacher, students deliver lectures to explain these skills to the class. Each lecture is then evaluated by the teacher as part of the formative assessment procedure. Moreover, the efforts and contribution of each group member are also assessed by the group leader in this experiment.

With the help of the teaching diary, questionnaire survey and interview, the writer finds that this assessment system which covers almost all the stages of the research is able to 1) evaluate both the learning process and results of the students; 2) reinforce the teacher's supervision and guidance of the whole learning process; 3) help the teacher identify and solve in time problems students have; 4) urge students to learn from each other and be clear about each step of their research through self-evaluation and peer evolution.

Innovation in EFL Creative Writing through Focus on Assessment for Learning

Wuhan University LI Danli LU Xiaoxian

Assessment for learning has been gaining currency in educational policy in different parts of the world with a paradigm shift from a focus on product to one on process in language assessment. While assessment for learning emphasizes the use of assessment for improving learning and teaching, assessment of learning focuses on using assessment for administrative and reporting purposes (Lee, 2011). In L2 writing, assessment has traditionally been characterized by assessment of learning. Although strategies of assessment for learning like process pedagogy, formative feedback, peer response, writing consultations and workshops have been promoted in L2 writing; these strategies are not widely adopted outside native English educational contexts. In English as a foreign language (EFL) contexts, there is little research that investigates writing teachers' attempts to bring innovation to their assessment practices through a focus on assessment for learning. Using data from three classrooms in Creative Writing for first-year students in the English Department in a Chinese university, the study aimed to investigate how the teachers' determination to implement assessment for learning in creative writing influenced their instructional and assessment practices. It also examines how assessment practices impacted on students' attitudes and beliefs regarding writing by drawing their feedback to reflect on the outcome of the course. The findings showed that the implementation of assessment for learning resulted in a significant change in teachers' instructional and assessment practices, and students improved their motivation in writing. The paper concludes with some implications on curriculum design and pedagogical issues for EFL writing.

A Study on Dynamic Qualitative Assessment for Ecological Classroom of

Foreign Languages

Harbin Normal University GUO Ying

In recent years, traditional classroom instruction can no longer satisfy the needs of the modern class. With infiltrating and influencing to the field of educational ecology, more and more educators focus on ecological element of instruction and liveliness of people in classrooms. The key to classroom instruction is assessment. The dynamic qualitative assessment for ecological classroom of foreign languages is especially indispensable. It shows great significance to the ecological classroom instruction assessment of foreign languages. The ecological classroom needs integrated assessment, emphasizing on interaction and communication between assessors and targets. Thus it can accommodate the requirement of natural situation from qualitative instruction assessment, which is in favor of achieving vision-integration between assessors and targets. As a result, it makes implementing qualitative instruction assessment for the ecological classroom necessary and feasible.

This study combines constructing ecological classroom, history and status quo of educational assessment research, implementing theories of education, educational ecology and educational assessment to analyze and study dynamic qualitative instruction assessment for ecological classroom of foreign languages. Applying literature analyzing and comparative research, it constructs integrated concepts, implementing procedures and methods of dynamic qualitative assessment for ecological classroom of foreign languages. The processes of implementing qualitative assessment in ecological classroom of foreign languages are illustrated by three dimensions: before class, constituting assessment project; during class, collecting assessment materials; after class, analyzing assessment materials, composing assessment report, and establishing assessment portfolios and feedback system. By analyzing two typical operational methods, implementation of dynamic qualitative evaluation in ecological classroom of foreign languages is demonstrated.

Chinese Students' Perceptions Of Innovative Assessment Tasks

In An EFL Writing Curriculum

Guangdong University of Foreign Studies & University of Hong Kong ZHU Qiyun

Assessment innovation often originates from the dissatisfaction of tutors and students towards conventional assessment methods and the criticism from academics. A deeper reason underpinning innovation in assessment is the change in educational goals along with the advent of information age. Accordingly, educational assessment must align with the changing needs of society to help complete the socialization of its members.

Chinese education with a millennia-old tradition of examinations does not seem to be helpful to deep learning. It is evidenced that high-stakes language test is not an efficient agent for desirable change in teaching and learning despite their crucial role in education. An alternative assessment system can be a valuable complement to the existing unsatisfactory system. Further, the quality of Chinese higher education has raised considerable concern along with widening participation since the end of 1990s.

The above acute pressures have given rise to a multitude of innovation in assessment as teachers attempt to cope with contending pressures. To enhance education quality, reform in curriculum or classroom activities should be launched to engage students in meaningful learning. Such significant instructional activities as assessment tasks, when designed in interconnected and constructive sequence are likely to enhance student involvement in

learning.

The significance of assessed work in students' academic life lies in that, first, it demands time and efforts from students. Second, it can direct students' attention to the most important aspects of the course. Third, it allows students to engage in meaningful and effective learning activities. Assessment task also plays a pivotal role in helping students develop evaluative expertise and facilitate their engagement with feedback. But what determines the depth of student engagement is students' perceptions rather than the features of tasks. If the instructional goal is to orient learners to deeper learning and higher-order thinking, the first important step is to understand their appraisal of the assessed tasks they are asked to get involved in.

The proposed study will adopt a qualitative approach to explore Chinese students' approaches to and perceptions of assessment tasks in an EFL writing curriculum in higher education. The research is prompted by the call for a pluralistic approach to assessment since no single assessment can fulfill all assessment purposes and needs, and the finding that writing as a key assessment tool carries great importance to success not only in tertiary education but also after graduation.

Drawing on multiple data sources including document analysis, in- and out-of-class observations, teacher and student interviews, focus group, open-ended survey and reflective essays, the study is expected to reveal some characteristics of innovative assessment tasks that engage students and the impact on learning. The intended contribution of the study lies in a better understanding of the repeated call in literature for a synergy between formative and summative assessments.

PBL 课堂评估中模糊评分的应用

东南大学外国语学院 邵争 杨茂霞

课堂评估是指发生在课堂内或课堂外的，由教师和学生共同参与的，对在课堂上进行的教与学的活动过程及其效果做出的价值判断行为。课堂评估作为课程中的重要环节，需要根据课堂教学的内容和方法来建立合理的评价指标体系和科学的考核评价方法。PBL 教学模式是一种强调以学生自主探究为主的教学方式，关注的是问题的探究和解决的过程，注重学生能力的培养。在教学实践中，PBL 课堂评估多依赖任课教师的主观判断，评分具有较大的随意性，对 PBL 教学产生了负面的反拨作用。因此，本文介绍了模糊评分法在 PBL 课堂评估中的应用，即以模糊集综合评价理论为基础，建立基于模糊集的 PBL 课堂评估评分体系，在评价过程中引入量化指标，通过科学的评价方法客观反映学生的学习情况。本文还尝试在 PBL 课堂评估的模糊评分中加入同伴互评，在实际操作中取得良好的效果。

Room 317

形成性评价概念重构视角下的教师评价能力标准研究

中央财经大学外国语学院 黄剑

北京师范大学外国语言文学学院 罗少茜

The past nearly 50 years have seen great development of formative assessment in both theoretical and practical terms. However, no agreement has been reached on what is formative assessment so far. Inconsistencies and even contradictions in definitions result in the difficulties for both researchers and practitioners, especially the one with describing competencies required of teachers for performing formative assessment. This research, made up of two parts, is aimed at updating the understanding of nature of formative assessment and teacher competences from perspective of formative assessment by synthesizing existing findings and observations. The first part examines various definitions and conceptualizations of formative assessment since Scriven coined the term formative evaluation in 1967 and analyzes them for their sources, similarities, differences and implications for practices. On the basis of the examination and analysis, formative assessment is re-conceptualized as such: “Formative” is the ethical requirements of all educational assessment practices; Assessment is by nature “judgment” and all assessment practices (planned or unplanned; high stake or low stake) and each step of assessment (elicitation, interpretation and use) have formative potential for learning; Instead of simply “instrument” or “process”, formative assessment

represents a new assessment paradigm which offers a new learning-oriented assessment perspective to look at testing and instructing practices with the focus laid on learning information and success criteria; Teachers' tasks are to take testing and instructional practices as assessment and tap to the full their formative potential for learning . The second part examines many mainstream assessment competence standards and analyzes them for their contents, origins, similarities, differences, contributions and inadequacies. Based on the examinations and analysis as well as recent findings and insights about formative assessment, a three-layer formative assessment competence model is proposed: key knowledge and skill layer, basic knowledge and skill layer and learning competence layer. The findings of this study are expected to deepen the comprehension of formative assessment and offer reference for teachers in improving their formative assessment ability.

An Investigation into the Impact of Formative Assessment on College Oral English Teaching and Learning

湖南理工学院外国语言文学学院 赵中宝

This paper aims to explore the concept of formative assessment and its application in oral English teaching and learning. Its impact on oral English teaching and learning is analyzed. The results of both quantitative and qualitative data analyses indicated that students who took the oral test over a period of time and received feedback regularly made significant progress in their oral English proficiency. The analysis of questionnaire survey and interview data revealed that the provision of diagnostic feedback was acknowledged by a great majority of students, and they thought that the diagnostic feedback could reflect the strengths and weaknesses of their oral English ability. The challenge facing the application of formative assessment is discussed in the final part of the paper.

Understanding Students' Formative Self-Assessment Practices in the Chinese

EFL Writing Class: Two Case Studies

广东外语外贸大学国际商务英语学院 王伟强

While there has been a proliferation of empirical studies on formative teacher assessment in L2 research, the research on students' formative self-assessment practices remains thin. This paper reports on a longitudinal inquiry (a 36-week academic year) into two students' formative self-assessment practices in the Chinese EFL writing context. Two Chinese undergraduates of an EFL writing course wrote in-class drafts on five assigned essay topics, self-assessed the drafts using an instructional rubric, and revised the drafts based on the feedback from self-assessment. Multiple sources of data were collected, including the students' first and revised drafts, individual retrospective interviews, and reflective revision journals. Results showed that although both students shifted their foci in self-assessment from the global to local aspects of EFL writing over time, their self-revision patterns diverged with regard to content and language use, which may be explained by their individual learning goals set for the same EFL writing tasks and adoption of differing self-feedback strategies. The results are discussed in terms of the influence of students' internalization of the rubric and goal orientations towards EFL writing on their self-assessment practices, with implications drawn for training students to be better self-assessors in the Chinese EFL writing context.

A Study of Education Students' Perceptions of Formative Assessment

西南财经大学经贸外语学院 刘洋

西交利物浦大学文化、语言和区域研究学院 李英春

There is an extensive body of literature suggesting that formative assessment can lead to considerable learning gains. However, it is also widely found that formative assessment is often wrongly understood and weak in practice. The current research investigates the perceptions of formative assessment by a small group of education students in a normal university of China, who hold the strong potential for being teachers in the future. In particular, the researchers aim to identify whether the participants well understand the key differences between formative and summative assessment, what attitude they hold to formative assessment and what the relationship between their opinions and practice is.

By administering a questionnaire eliciting the opinions and making classroom observations subsequently, the researchers find that the real distinction between FA and SA is still blurring and even beyond the reach for most participants in this study although they

generally hold a positive attitude to formative assessment activities. Besides, the observation of classroom practice reveals a weak dialogic pattern, either being teacher-dominant or lacking necessary teacher guidance. Thus, information conveyed through this pattern is not qualified enough to improve learning performance effectively, leaving the essential purpose of formative assessment hardly achieved.

Language teachers' target language skills: assessing students' writing

Lomonosov Moscow State University Tatiana Ershova

Feedback is an integral part of formative assessment of writing in the EFL classroom. Yet, a lot of English language teachers in Russia are reluctant to provide commentary on their students' compositions. Not only is it time-consuming, but also reported to represent a complex, challenging task, both linguistically and cognitively.

With the increasing tendency for English being used as a medium of instruction in different EFL contexts, it is important that teachers are able to perform their professional tasks in English. At present, pre-service non-native English speaking teachers' language proficiency is commonly assessed as 'general English proficiency', or EAP (English for academic purposes), which does not directly correspond with the actual language use of future teachers (Freeman, Katz, Gomez & Burns 2015). Hence, there comes a need to design training materials that will help bilingual teachers to develop an integral professional linguistic competence across all four language skills.

In this presentation, I will focus the specific reading and writing skills that EFL teachers need when assessing their students' writing and providing constructive written feedback. Firstly, I will discuss linguistic and cognitive complexity of these professional tasks, and present a list of sub-skills required to enact them successfully. Secondly, I will share with the audience a set of fifteen reading-into-writing exercises that we have designed to prepare English language teachers to effectively respond to their students' writing. These exercises have been piloted with the students in the department of foreign language teaching at Moscow University, and may be of interest to practicing EFL teachers and teacher trainers.

Using formative assessment to facilitate self-regulated learning: assessment practices and student perceptions in an EFL writing classroom

Hang Seng Management College Maggie Ma

Formative assessment is increasingly implemented in diverse contexts, but much needs to be known about how it is implemented in Confucian Heritage Culture (CHC) settings, which may pose obstacles to its implementation, and student perceptions of it. This exploratory case study investigates one EFL writing teacher's formative assessment practices at a tertiary institution in Hong Kong and how students perceive of them, given the scant research in this type of context. It particularly focuses on the kind of practices that are likely to facilitate learner self-regulation (Nicol & Macfarlane-Dick, 2006). Data were collected through classroom observation, teacher and student interviews, student questionnaires, and documents. Five formative assessment practices likely to support self-regulated learning were identified. They seemed to tend towards the extended end of the formative assessment continuum (Carless, 2011), albeit with features of "restricted" forms. The students seemed to have a more positive opinion of formative assessment that were partially teacher-directed (e.g. face-to-face consultation) than student-directed assessment (e.g. peer- and self- assessment), with peer assessment being favoured the least. Suggestions are provided for the implementation of formative assessment in EFL writing classrooms in CHC settings like Hong Kong to represent the "spirit" of formative assessment and to further support learner self-regulation.

Implementing and evaluating the use of e-portfolios in FL instruction

School of Foreign Languages Thai Nguyen University Nguyen Thi Minh Loan

In the context of Vietnam, Vietnam is implementing National Foreign Language 2020 Project (NFL2020) to carry out the Vietnamese government's Decision 1400 to build national foreign language capacity. The mission of NFL2020 is to build Vietnam's foreign language (mainly English) teaching and learning capacity and to renovate the teaching and learning of foreign languages within the national education system. Teachers' development is the key to the success of the Project. The lack of effective English teachers has produced students at different levels who fail to communicate in English for any useful purposes. Vietnam is coping with the problem of poor qualification of English teachers that comes in the following forms: poor self-knowledge, subject knowledge, knowledge of students; low managerial skill

and poor pedagogical competence, lack of reflection and the ability of professional development. Hence, beside retraining the teachers of English at different levels, national foreign project turns to focusing on preservice teacher learning.

E-portfolios in teacher education programs are useful for documenting evidence of preservice teachers' competencies and for guiding preservice teachers' long-term professional growth and helping preservice teachers become reflective practitioners, and therefore increase their employment opportunities. Most importantly, through reflecting upon their experience of doing their e-portfolios, preservice teachers come to understand how technology can enhance learning and teaching.

The fact that portfolio assessment still remains an approach that has not been adopted in Vietnam, despite its worldwide-acknowledged benefits, opens up a wide door for research. Given the nation's existing problems in quality control and assessment of teachers, an introduction to the concept of portfolio assessment could be of great value in the reform of the Vietnamese education system. It is believed that the purpose of this study particularly aligns with the Vietnamese government's strategic plans for educational reform, and that it will have practical implications for applying portfolio assessment in Vietnam.

The researcher is a teacher trainer. She wants to implement this project in order to investigate the implementation and the effectiveness of the use of language portfolios in FL instruction in a university context. Established uses of FL portfolios will be looked into and assessed from learner and teacher perspectives with the aim of establishing best practices for FL portfolio implementation. It is particularly relevant for school principals who are considering assessment portfolios for use in their schools. In addition, staff developers and teachers may also benefit from reading this document to learn more about assessment portfolios and how they can improve student learning and assessment.

The study would give an overview of the existing research into portfolio use. Based on this background, the study would propose one framework for using portfolios in training pre-service students majoring in English Education at School of Foreign Languages at Thai Nguyen University in Vietnam. After a small-scale implementation of a portfolio in a course of English teaching methodology, the researcher will assess the effects of portfolio use on learner outcomes and learner motivation. Therefore, recommendations for best practices in FL portfolio implementation would be put forward.

Room 318

Let's Roll up Our Sleeves: Writing Our Own Rubric for Oral Assessment

Anadolu University Muhiddin Timuçin Denктаş

Since the interaction among people has improved a lot due to the developed technology all around the world, communicative approaches to language teaching have also become indispensable in ELT. In other words, teaching, and hence assessment, of productive skills has also gained a lot of importance. Therefore, Anadolu University School of Foreign Languages has been adapting its program according to the recent developments in ELT. Namely, this year a new curriculum has been launched according to GSE (The Global Scale of English by Pearson).

Aforementioned design suggests a list of outcomes and requires both teachers and students to achieve these learning outcomes in the classroom. These outcomes define linguistic knowledge (i.e. vocabulary and grammar) as well as four language skills (listening, reading, writing, and speaking.) Therefore, teaching policy has been revised as well as assessment policy of the school so that assessment has been based upon GSE's learning outcomes. Specifically, the school has felt the need of revisiting its own rubric for oral assessment. In order to produce a new one, which matches both GSE outcomes and the educational context, testing unit embarked upon writing its own rubric.

This study presents the first phase of designing an oral assessment rubric for proficiency, which aims to determine the raters' concept of a proficient learner. As the first step, eight raters, whose experience in rating and teaching ranges from 5 to 20 years, are randomly selected and asked to define what makes them think whether a student is proficient in speaking. In the second step, they are asked to watch spoken performances of 10 students to decide if each performance is proficient or not according to their beliefs, without use of any rubric, and they are expected to justify their decisions. In the last step, their first definitions and their decisions are compared to see which beliefs are used in practice.

The analyses obtained in the first phase of this work-in-progress study will form a sound basis for the next phases of writing our own rubric for speaking exams.

Taking into account topic familiarity in assessment design

University of Hong Kong QIU Xuyan

Content or topical knowledge is claimed to exert influence on L2 learners' language performance (Ellis, 2012; Skehan, 2014). The positive effects of familiar topics (i.e. background or prior knowledge) on EFL learners' reading and listening comprehension have been explored while there remains an open question that whether and how familiar topics affect their oral production. The selection of the topics for the assessment tasks also needs more discussion so as to develop further insightful understanding on the design of tasks to meet both the learning goals and the assessment purposes.

This study aims to investigate the effects of familiar and unfamiliar task topics on the oral performance of 60 Chinese undergraduate students with the lower and the higher L2 proficiency levels. Each of them performed four oral narrative tasks. Two familiar topics, which were related to their cultural background and past experience, were chosen while two unfamiliar topics were provided. Stimulated recall interviews were conducted for each participant immediately after their task performance so that their inner thoughts during speech production could be captured. The participants' oral output was measured in terms of fluency, accuracy and complexity (CAF) and the interviews were analyzed based on Levelt's speaking model (1999) describing the oral production into conceptualization, formulation, articulation and self-perception stages.

The results showed that EFL learners produced more fluent, accurate and complex speech under the familiar conditions. In addition, the positive influence of the familiar topics was more obvious for the lower-proficient learners. From the interviews, the familiar topics facilitated the participants' conceptual preparation so that they could switch their attention from the content to the forms. The unfamiliar topics imposed more cognitive load on the speakers which make the tasks more challenging. The research findings may yield implications for the design of assessment tasks in the classroom context and even in the tests. The selection of familiar topics and unfamiliar topics may affect the task complexity which can be taken into consideration during the design. For the lower-proficient learners, tasks with familiar topics may engage them into the language production process so that they are more willing to participate. For the higher-proficient learners, tasks with unfamiliar topics may be more useful in activating their language skills.

L2 Development of Speech Rhythm by Mandarin and

German Learners of English:

Implications for L2 English Pronunciation Assessment

Communication University of China Li Aike

University of Cambridge Brechtje Post

This presentation aims to demonstrate how results of laboratory research on second language (L2) acquisition of prosody, and speech rhythm in particular, can contribute to improvements in L2 English pronunciation assessment.

Assessing L2 English pronunciation proficiency is problematic, which is reflected in the vagueness of, for instance, the current descriptors of the Common European Framework of Reference (e.g. A2 level pronunciation is “generally clear [...] despite a noticeable foreign accent” (CEFR, 2001). There has been a call for more detailed and explicit pronunciation performance descriptors (e.g., Galaczi et al., 2012), and there has been a growing interest in the automated assessment of speaking skills including pronunciation (e.g., Hawkey&Milanovic, 2013).

Researchers have proposed that speech rhythm reflects a combination of speech properties, including vowel reduction, syllable structure, accentual lengthening and final lengthening (e.g. Prieto et al., 2012). In addition, several rhythm metrics, such as Varco-V and nPVI-V, have been proposed to quantify speech rhythm (e.g., Low et al. 2000), and they have been successfully applied in child speech studies and L2 studies (e.g. Payne et al. 2012).

In this study, we compared L2 English learners at two proficiency levels (B1 and C1), of two typologically different L1s (Mandarin and German), to investigate (1) To what extent do the various rhythm metrics and the speech properties which contribute to speech rhythm discriminate between L1 and L2 speech by the different learner groups? (1) Are the L2 learners’ developments of the systemic properties mentioned above reflected in their overall L2 rhythmic patterns (as measured by rhythm metrics)? The learners were asked to read English sentences in which syllable structure was systematically varied. Native speakers of the three languages participated as control groups. The data were analysed acoustically in *Praat*.

Results of the research could contribute to the development of more detailed and explicit pronunciation performance descriptors, by indicating that the speech properties of vowel reduction and accentual lengthening could be used as descriptors in pronunciation assessment. It has been found in this research that these two properties discriminate well between different L2 proficiency levels; and their developments appear to be free from L1 interference, as L2

English learners follow similar paths in the developments of these two properties irrespective of their L1 backgrounds.

In addition, the results of this research could also contribute to the development of the automated assessment of L2 English pronunciation. It has been found that the rhythm metric nPVI-V shows great discriminative power in differentiating L2 learners of different proficiency levels; L2 English learners of typologically different L1s exhibit similar rising trajectories in the development of L2 vocalic variability, as measured by nPVI-V, which can reflect the learners' productions of vowel reduction and prosodic lengthening. These results suggest that nPVI-V could be used as a measure in the automated assessment of L2 English pronunciation.

The role of speech content in the assessment of candidate oral performance

Renmin University of China LIU Li JIA Guodong

The content that test-takers attempt to convey is not always included in the construct definition/rating scales of general oral English proficiency assessments. The current study was thus designed to explore the relative contribution of speech content and linguistic dimensions of the rating scale to raters' holistic judgments of general English oral proficiency. Test of Oral Proficiency in China (TOPE) was designed and administered to undergraduates in Renmin University of China to determine what criteria raters associate with general oral proficiency.

Six raters were recruited to rate 60 students' speeches on three topics, based on their intuitive judgments of oral proficiency (referred to as *Overall communicative effectiveness* in the rating scale). They then assigned scores to the monologues using three analytic criteria: *Content elaboration/development*, *Delivery (Grammatical and lexical range and accuracy)* and *fluency*. The raters were also asked to provide open-ended oral feedback on the factors contributing to their intuitive judgments. To analyze the data, Statistical analyses of the scores including Rasch model and multiple regression were conducted. Results showed that *Content elaboration/development* made a substantive contribution to the intuitive judgments and composite score. The current study has identified an important dimension that is highly relevant to oral proficiency but is not fully delineated by existing communicative language ability models. An ability to elaborate speech content is found to form a major part of the non-linguistic components that greatly affect rater judgments of the second language oral performance.

Developing a Computer-aided Speaking, Listening and Writing English Test at a University Level: Validity and Reliability Evidence from Response Data

School of English and Education, Guangdong University of Foreign Studies LV Jiantao

A computer-aided speaking, listening and writing English test, which is one of the components of GDUFS Professional English Aptitude Test, (GPET) is developed at a university level for assessing students' skills of communication in real life situations. Multiple choice in nature, however, unlike traditional listening comprehension test, in which students are supposed to answer some questions related to test materials, the test involves test takers in several conversations and monologues, namely conversational strategy tasks, followed by a continuation task, in which test takers are required to first read the first half of a short passage and continue writing the second half of it accordingly. Validity of the conversational strategy tasks and rater reliability of the continuation task are investigated using Winsteps (Linacre, 2011) and Facets (Linacre, 2012). A descriptive data analysis of MC items is also conducted. Results show that the assumed latent trait manifest itself well in the data. Most of the MC items exhibit neither outfit nor overfit statistics, which indicate consistency with the assumed latent trait and moderate possibility of guessing. Raters' behavior scoring test takers' writing, features of test-takers' practical English ability and possible wash-back of the test, based upon students' feedback comments are discussed.

两种评分方法的评分效应比较研究

林椿, 肖云南

湖南大学外国语与国际教育学院

本文针对对湖南大学英语分级测试中随机抽取的 210 份作文试卷, 运用多面 Rasch 模型对比分析了整体评分法和分项评分法下 14 名评分员的评分效应。并以整体评分表现作为参照, 对分项评分的各个维度的评分信度进行比较。提出针对易产生偏差的评分维度及学生群体进行专门的评分员培训, 对评分量表上维度的描述进行调整, 使评分员达到对量表一致的诠释, 减少评分偏差, 以期对评分员培训和评分量表的解释及改进有实际应用价值。

Language Test Fairness: Conception and Framework

华中师范大学童扬芬

Due to the increasing important role the language tests play in today's society, test fairness has become one of the major concerns in the language testing field. Based on the review of language tests research, the paper makes careful analysis on the definitions and framework of test fairness as well as a discussion of tests fairness requirements. To promote better test fairness, it is essential to improve assessment literacy among item writers and to ensure content free of bias.

Room 422

Interactional patterns of Chinese EFL Learners in the Paired Speaking Test

Beijing Language and Culture University LIU Liandi

As conversational strategies such as the ability to initiate, respond and negotiate meaning in situated performances have become a focus of L2 oral proficiency assessment during the last decade, there has been a growing interest among language testing researchers in analyzing candidates' spoken production in paired and group orals. Although there has been a sizable body of research studies exploring the nature of test talk elicited by these two formats (e.g., Davies, 2009; Galaczi, 2004, 2010; Gan, 2008, 2010; Gan, Davidson & Hamp-Lyons, 2009; He & Dai, 2006; Lazaraton & Davies, 2008; May, 2000, 2007, 2009; Luk, 2010; Nakatsuhara, 2006, 2009, 2010, Van Moere, 2007, 2010, yet published studies on oral test-taker interactive

discourse by Chinese EFL learners are rare. The present study, informed by the notion of interactional competence (Ducasse & Brown, 2009; Hall, 1999; He & Young, 1998; Kramsch, 1986; McNamara, 1996, 1997; Young, 2000) and interactional sociolinguistics (Brown & Levinson, 1978, 1987; Goffman, 1974, 1983; Gumperz, 1982; Schiffrin, 1994; Tannen 1984/2005, 1994, 2006), investigates interactional patterns Chinese EFL learners exhibit in the Public English Testing System (Level 5) Spoken English Test (PETS-5-SET) by using conversation analysis techniques. To better understand this issue, the traits of Chinese students' interactions in the paired task are compared with those of Australian students to discern similarities and differences between the two groups.

Data were collected from 60 Chinese and 30 Australian university students by conducting a mock oral test in the form of paired discussion. The discourse performances were audio-recorded, transcribed, coded and analysed both qualitatively and quantitatively, focusing on interactional patterns. In total, seven different patterns were identified from the dataset, including six discrete ones (collaborative, parallel, dominant/passive, interviewer/interviewee, consultant/client, and expert/novice) and one blended pattern. The comparative analysis indicates both similarities and differences between the Chinese and Australian students in their adoption of interactional patterns in the dyadic conversation. Overall, 23 (78%) out of the 30 Chinese dyads were identified as collaborative, two parallel, one dominant/passive, one interviewer/interviewee, one consultant/client, one expert/novice, and one blended pattern (of dominant/passive and collaborative). As for the Australian group, the overwhelming majority of the dyads (96.7% of the total) assumed the collaborative patterns of interaction and only one pair used parallel pattern.

Based on findings of the study, it could be argued that more Australian than Chinese dyads worked cooperatively, therefore they produced symmetrical interaction, characteristic of higher interactivity and dynamic conversations. However, the conversations by the Chinese students generally reveal lower interactivity and hence sound somewhat unnatural. The disparities between the two groups with respect to the interactional patterns could be attributed to differences in four major factors: language proficiency, conversational management skills, content knowledge, and their respective culturally-valued conversational styles.

The study has significant implications for such areas as research on EFL spoken interaction, interactional competence theory, L2 teaching and testing, i.e., a better understanding of the construct of interactional competence underlying candidate-candidate paired task in the PET-5-SET, for empirically-based rating scale construction for the test, and for the examiner training.

Using the Trait–State–Occasion Model to examine changes in graduate students’ metacognitive strategic competence development

Hong Kong Polytechnic University CAI Yuyang

Metacognitive strategic competence (or metacognitive strategies) refers to the ability of language learners to plan, monitor and evaluate their language learning (Bachman & Palmer, 2010; Phakiti, 2008). Metacognitive strategic competence supports language proficiency development and make language learning and use transparent to language learners (Oxford, 2013). Assessment of strategic competence, however, has mainly focused on the temporal state, leaving understudied the longitudinal consistency of this competence (Cohen & Griffiths, 2015). The present research used the Trait–State–Occasion Model (Cole, Martin & Steiger, 2005) to evaluate the developmental consistency of metacognitive strategic competence of students for reading during their transition from undergraduate to graduate study. Analyses were based on three administrations of the Metacognitive Strategy Use scale (Cai, 2013) at the interval of two months. Participants included 346 first year graduate students, aged from 22 to 30, from a university in China. Results indicated that a completely stable trait factor and an autoregressive trait factor accounted for the majority of the variance in metacognitive strategic competence, whereas state factors only accounted for about 6% of the variance. That the stability of individual differences in metacognitive strategic competence during reading performance increased provides substantive and methodologically rigorous evidence to the notion that metacognitive strategic competence is an enduring rather than a transitory state-like individual difference construct. Implications for strategic competence training in language courses at graduate level will be discussed.

The Vocabulary Use of Chinese Candidates of Different Levels in Oral English Examinations

北京理工大学外语学院 张剑

广西柳州职业技术学院贸易与旅游管理系 罗柳萍

There are many studies (Laufer & Nation, 1995, 1999; Meara & Bell, 2001; Nation, 2001. Milton, 2008) on vocabulary assessment of written texts, however very few (Malvern and Richards, 2002) on spoken texts in oral examination settings. In the present research, five vocabulary use assessment measures (type, token, D, Guiraud and Advanced Guiraud) of three different levels of GESE (Graded Examinations in Spoken English for Learners of Other Languages) candidates were explored. The main purpose of the study is to investigate whether or not and how (if yes) the lexical measures can indicate the language level of GESE candidates.

The main research questions of the present study are 1) whether or not the 5 measures can differentiate between candidates of three different levels and 2) whether or not the measures can differentiate good performers from poor performers at the same level of test.

180 samples were collected from the 2008 Beijing Corpus GESE administrated by

Trinity College, London and Beijing Education Examinations Authority (BEEA). The subjects were Chinese children or adolescence who took GESE examinations in Beijing. 60 random samples were collected from Grade 2 (Initial Stage), Grade 5 (Elementary Stage) and Grade 7 (Intermediate Stage) respectively. Analysis of variance (one-way Anova) and Post-hoc comparisons using Tukey HSD tests indicate that candidates of three different levels use vocabulary differently. There is a trend that the higher the grade, the higher the indexes of lexical measures. The differences between Grade 2 and Grade 5, Grade 2 and Grade 7 are significant; however, the differences between Grade 5 and Grade 7 are not significant. Regarding the lexical measures of good and poor performers at the same grade, all the lexical measures can distinguish between good and poor performers in Grade 2 and Grade 5; however, only Type, D and AG can distinguish between good and poor performers at Grade 7. The results raised important questions for research on lexical richness and assessment: why some lexical measures such as D and AG are able to suggest the subtle difference of candidates' lexical usage, and second, why no lexical measures can distinguish between Grade 5 and 7 candidates, who are expected to be significantly different from each other at language proficiency level.

Multimodal Analysis of Interviewers' Pragmatic Identity in English Oral Testing

河海大学 张聪

Based on the Pragmatic Identity Theory, this paper discusses the identity construction of interviewers (including the identity classifications, its multimodal means of construction, and the motivation of different identity construction) in BEC oral tests by analysing the video data of BEC. It is found that, in order to achieve the purpose of encouraging the candidates to present their language capability and obeying the principle of fairness and objectivity, the interviewers not only construct the identities of professional examiners but also identities of servants, coordinators, monitors, managers, and motivators via a multimodal communication (verbal and non-verbal) with candidates in an institutional discourse context of English Oral Tests. This research will be helpful in the examiner training to some extent.

An Introspective Study on Test-takers' Process of Picture

Composition in the New HSK-5

Chongqing University YANG Ji GU Xiangdong LIANG Yansong

The new HSK (Hànyǔ Shuǐpíng Kǎoshì), translated as Chinese Proficiency Test or Chinese Standard Exam, is China's only standardized test of Standard Chinese language proficiency for non-native speakers. The new HSK-5 is regarded as the watershed which distinguishes intermediate and advanced Chinese proficiency. According to statistics, among

the global means of the three sections including listening, reading and writing, that of writing section is the lowest (Luo & Zhang, 2011). Obviously, there is a lot of room for the candidates to improve their writing abilities, so it is worthy of attention to investigate candidates' performance in this test.

The new HSK-5 writing test consists of three parts, sentences composition, keywords composition and picture composition. This study employs a concurrent think-aloud protocol and immediate retrospective interviews to investigate the dynamic process of picture composition in the new HSK-5. Six foreign MA students in Chongqing University who had passed the new HSK-5 and were at the same level (based on their new HSK-5 scores) participated in this study. The discourse organization process, strategy use and the relation between test-takers' written products and their writing process were investigated and analyzed in the current study. The research addressed three questions: 1) What steps are involved in discourse organizational process and how do test-takers organize them? 2) What types of strategies do test-takers employ and how do they use these strategies? 3) What are the relations between test-takers' written products and writing process?

The major findings are as follows. Four steps are identified in discourse organizational process in sequence: observing the picture, building the situation, describing the picture and examining the text. A large number of writing strategies were employed, including cognitive strategies (regression, translation, review, self-question, avoidance and paraphrase), and metacognitive strategies (monitor, plan and evaluation). Among them, regression, review and translation are most frequently used strategies. Besides, the candidates also use test-wiseness strategies to acquire higher scores. In most cases, the written products are consistent with verbal protocols, but unidentical cases including wrong spelling, content replacement, addition and missing are found as well. Attention distribution and language proficiency may be the main causes to such inconsistency.

This study provides valuable insights into the test-takers' process of picture composition in the new HSK-5. It is hoped that this study offers references and suggestions for further HSK writing instruction and Chinese as the second language writing tests.

A closer look at the cognitive processes of NMET reading

Guangdong University of Foreign Studies LI Xueping

As one of the high-stake tests in China, NMET deserves continuous efforts to collect various kinds of evidence to support its score interpretation and use. Among its several parts, the Reading section takes the form of multiple-choice, which has been criticized by many researchers for the negative influence on test validity. Therefore, it is essential to investigate whether the cognitive processes involved in the completion of NMET-Reading reflect current L2 reading theories, are consistent with what the test was supposed to measure, and thus can be used as a valid indicator of test-takers' reading ability.

This study mainly took a qualitative approach to the cognitive processes of NMET-Reading. Five high school seniors were invited to think-aloud their thoughts while

taking the NMET-Reading of a past paper; in addition, a retrospective semi-structured interview was conducted to confirm some of the observed test-taking behaviors on the one hand, and on the other hand to gather information about test-takers' opinions about the item difficulty, as well as their perceived reading ability measured by each item. Then their verbal reports and the interview were transcribed and coded in terms of reading-related processes, test-management-related processes and test-wiseness-related processes.

The results showed that test-takers took the NMET-Reading more as a problem-solving process than a reading process, since there were more test-management-related processes reported than those related to reading. Meanwhile, test-takers' reading-related processes were generally in consistency with what each test item aimed to tap into. In terms of test-takers' ratings of the item difficulty, no significant correlation was found between any two of the five students. Additionally, some problems were noticed, which might undermine the validity and therefore called for greater efforts during the phase of test design and development of NMET-Reading.

Room 423

EFL Listening Diagnostic Assessment Based on Cognitive Diagnostic

Approaches

西安交通大学 孟亚茹 马晓梅

Cognitive diagnostic assessment (CDA) has great potentials in large scale testing. With it, researchers in educational measurement and cognitive psychology are finally in a position to design tests targeted specifically for providing valuable information about students' cognitive strengths and weaknesses (Leighton & Gierl, 2007b).

Unlike traditional tests, which merely provide a single overall test score, CDA measures the specific knowledge states and cognitive processing skills underlying the overall test score. This allows for possible remedial guidance to address individual and group needs and help design better instruction for students' effective learning and progress.

CDA had not been adopted in second language assessment until in recent several years, but very few cases in EFL listening research have been found. The study makes a bold attempt to employ Cognitive Diagnostic Approaches (CDAs) to develop a listening diagnostic test model for fine-grained assessment and classification of the learner's knowledge state and presence or absence of certain attributes. The study also touches on the diagnostic report and the corresponding remedial guidance. The addressed questions are as follows:

What attributes count in EFL listening and how are they related hierarchically?

How can we construct and validate the EFL listening CDA model?

What are the considerations in presenting the online diagnostic reports?

How can we guide the learners based on their diagnostic reports?

Based on diagnostic approaches, the study first identifies attributes required to do EFL listening. Then the hypothetical Q Matrix model (H) is proposed and it is followed by the validation process. 3 versions of hypothetical models are compared and analyzed, i.e. the test developer's model (H1), students' Think Aloud Protocol (TAP) model (H2) and 7 domain experts' judgment model (H3). 534 test responses and the above three versions of hypothetical models are psychometrically analyzed using G-DINA model to find out their absolute and relative model-data fit statistics. The best-fit sub-model proves to be H3. After these exploratory procedures, the best-fit model is further verified with new data and its goodness of fit is confirmed.

The findings come as follows:

To answer the first question, seven attributes were identified: phonological level (sound discrimination, stress and intonation); lexical level (less frequent vocabulary and oral expressions); syntactic level (less frequent vocabulary and oral expressions); facts and details; main ideas; contextual and cultural related inference; note-taking and working memory. They are mainly independent attributes.

How to develop the CDA model? It generally follows the 4-step procedure of "attribute identification" → "Q-matrix construction and test item writing" → "Psychometric analysis and verification" → "Diagnostic reporting". But the research demonstrated that "remedial guidance" is a very appropriate and desirable destination for a diagnostic model. They five constitute a perfect system to empower learners in pursuing autonomy in EFL learning.

About the online diagnostic report, 5 considerations were proposed: 1) offering orientation information to the diagnostic score users; 2) integrating the total scores into the diagnostic report; 3) making diagnostic report files and scores interpretable, such as in the form of graphs, tables or other interactive ways; 4) offering multi-dimensional analysis; 5) including narrative interpretations for the unconventional diagnostic scores.

Teacher and learner users can take advantage of this diagnostic reporting system to achieve the aim of personalized or "tailored" learning and teaching.

Diagnosing EFL writing difficulties in Chinese secondary schools

Shanghai International Studies University ZHAO Guanfang

Fudan University FAN Jinsong

The evaluation of students' educational progress in Chinese schools still depends heavily and almost exclusively on the use of timed tests. The extensive use of such tests, however, leads to a general lack of writing ability among Chinese students. In EFL teaching and learning, this problem is especially salient (Yan, 2012) and is seen as "directly affecting [Chinese students'] research and learning ability at the postgraduate level" (Zhang, 2010, p.311). To help students develop their EFL writing ability, however, we need to first understand the nature and the sources of their writing difficulties.

As a response, this study aims to collect evidence from various channels at both secondary and post-secondary levels. The construct of EFL writing ability as operationalized in learning standards, textbooks, and high-stakes tests is examined vis-à-vis the Hayes-Flower (1980) writing model and the model of language ability (Bachman & Palmer, 1996). Such analysis can help identify components of writing ability that are addressed in the Chinese context, which in turn could reveal possible construct under-representation that may be a source of students' writing difficulties. Observations of writing instruction and interviews with teachers are then conducted to identify instructional deficits that may cause students to experience certain challenges when they write. Finally, insights into writing difficulties are also gained from students themselves. In addition to interviews with college students about their writing difficulties, a diagnostic tool is used to collect data from high school students, as they are less able to verbalize their challenges in the first place. This presentation will focus on reporting the development and application of this tool in such a particular context.

A preliminary version of the diagnostic tool was constructed based on Llosa et al.'s (2011) inventory of high-school students' writing difficulties, together with considerations of high school students' learning objectives, overall language proficiency, and the assessment practices in the Chinese context. Unlike the existing diagnostic tools that either focus on assessing discrete aspects of writing ability or rely on a detailed analytic rubric to diagnose strengths and weaknesses, this instrument examines both the cognitive writing process and the written product. It first offered scaffolding pre-writing questions that tapped into students' key cognitive ability in areas such as task interpretation, audience awareness, planning and organization; it then solicited the actual writing product based on their responses to the pre-writing questions; finally, a post-writing questionnaire was used to collect their immediate reflections on and self-assessment of the writing performance and specific writing difficulties.

This instrument was then administered to 190 10th and 11th grade EFL students at a

local high school in a pilot study of the usefulness of the instrument. Preliminary data analysis showed that the most prevalent and salient writing difficulties among our high-school students include the lack of ability to accurately interpret the writing task, effectively address audience's needs, and clearly differentiate writing and speaking genres, among other language-specific issues and problems. Implications and suggestions for EFL writing instruction and assessment, based on such results, will be shared and discussed.

Formative assessment of language teachers' ESP:
developing a scale for diagnostic assessment of teachers'
classroom language proficiency

Southampton Solent University	Carolyn Westbrook
Southampton Solent University	Yvonne Merry
Lomonosov Moscow State University	Alexey Korenev
Lomonosov Moscow State University	Tanya Ershova

Over the years, much research has focused on methodological aspects of foreign language teacher training (Freeman, 1989; Ellis, 1986; Burns & Richards, 2009; Scrivener 2014). Similarly, existing teacher training courses (TKT, CELTA, DELTA etc.) and language proficiency examinations (such as Cambridge main suite) assess a wide range of professional skills and general language proficiency. However, the language that teachers use both inside and outside the classroom, that is, the ESP of teaching, has not been scrutinised to the same extent. Research into teachers' language has mainly focused on classroom language use (Elder, 2001; Freeman, Orzulak & Morrissey, 2009; Walsh, 2011), but the assessment of this language does not play a major role in any teacher training qualification.

This presentation will outline an ongoing project which aims to develop an assessment scale built on the conceptualization of classroom interaction as a particular communicative competence of teachers and an essential part of their ESP. Using previous research in the areas of academic and classroom discourse (Mercer, 1995, 2008; Hyland, 2009) and works that look at interactions in foreign language classrooms (Bailey & Nunan 1996; Walsh, 2011), as well as transcripts of videorecorded lessons, we will present an online diagnostic tool for assessing teachers' and trainee teachers' classroom language, which will use the set of descriptors under development. The assessment cycle will be presented along with the methodology used to develop the descriptors. The criteria and the scale used for formative assessment of language teachers' target language will also be discussed. Participants will be invited to submit videorecorded lessons for formative assessment based on the criteria which will enable the

project team to validate the descriptors and we will provide feedback to teachers based on the assessment criteria. Materials will also be developed to help teachers improve the areas of weakness.

A Diagnostic and Prognostic Approach to Error Analysis of Ab(use) of Modifiers among Yoruba Bilinguals

Federal University OYEDOKUN-ALLI, WASIU ADEMOLA

In a fairly obvious sense, any native speaker of a language can be said to know the grammar of his or her native language. After all, native speakers clearly know how to form and interpret words, phrases and sentences in their native language, Radford (1997). But this, clearly, is not the case with L2 learners. In today's world, bilingualism has become an entrenched part of societal values. The pre-eminent position of the English language in global affairs has made its use widespread in international trade, international scholarship and scientific research. It is used as a second tongue to millions of users of other languages, Nigeria inclusive. However the study of psychological correlates of language has revealed that a bilingual speaker is (probably) never equally competent in both languages, Lado (1957). Therefore, this paper aims at discovering and describing the problems that the L2 learner of English will have. The theoretical frameworks adopted for the study involves a synthesis of *inter language theory model* and Quirk and Greenbaum's *Performance and Judgment test*. The study recommends that teachers and curriculum planners should employ both diagnostic and prognostic methods in addressing problems encountered by the L2 learners of English and that language learning tasks should be made to accommodate a variety of language activities since languages, generally, are ever dynamic.

Diagnostic assessment and tracking of English academic writing of Hong Kong university students

Hong Kong Institute of Education XIE Qin

To date, only a handful of studies in language testing and assessment utilized Cognitive Diagnostic Modeling (CDM) for diagnostic assessment of L2 academic writing. Even fewer studies tracked the development of L2 academic writing over time. This study adopted CDM to profile the English academic writing performance of first-year university students and tracked the changes of their writing profiles over an academic writing course.

This study involved 472 first-year undergraduate students from a university in Hong Kong. Participants were involved in a 13-week English academic writing course, for which they produced three essays for assessment at the beginning (Time 1: individual, timed essay), during (Time 2: pair co-construction, untimed essay), and the end (Time 3: individual and timed essay). At each assessment point, students' writing was assessed by nine EAP instructors against a fine-grained diagnostic checklist for academic writing. The diagnostic data were analyzed via the reduced RUM to generate for each student a multi-dimensional diagnostic profile on five writing attributes. Salient writing profiles was identified for each assessment point and their changes were tracked over the three observations.

The study identified four major writing profiles out of the 25 posterior probability of mastery profiles at each assessment point, and multiple transition trajectories from Time 1 to 3 and from Time 2 to 3. Comparing Time 1 and 3, students' writing performance improved on the two attributes of 'task fulfillment' and 'vocabulary use', but deteriorated in terms of 'grammar' and 'organization'. Performance on 'mechanics' has marginal insignificant decrease. Besides generating individualized feedback to students, CDM-based diagnostic assessment of academic writing provided useful information for course planning, instruction, and evaluation. Limitations of this study and further directions will be discussed.

Construction and Validation of Diagnostic Proficiency Scales for Chinese EFL

Readers in High School

广东外语外贸大学 范婷婷

Scales of language proficiency provide “characteristic profiles of the kinds and levels of performance which can be expected of representative learners at different stages” (Trim, 1978; see also North, 2000). Alderson (1991) summarized three functions of proficiency scales, namely, for users, assessors and constructors, with the first reporting information about typical candidates at a given level, the second guiding the rating process of learner performance, and the last selecting tasks for test developers in either formative or summative assessments. Pollitt and Murray (1993) advanced Alderson’s classification a step forward and added another term “diagnosis-oriented scales” to involve those describing different aspects at different levels and those designed to provide elaborated feedback for scale users. For current proficiency scales, most have assumed the first three functions. Little, however, are designed with the diagnostic purpose with the exception of Knoch’s (2007) diagnostic rating scale for the Diagnostic English Language Needs Assessment (DELNA).

The present study will explore the scale construction from a diagnostic perspective, on the basis of the general description of the National Proficiency Scales (reading section). The author will first examine the differed weights of the descriptors on NMET reading tests, followed by a diagnostic scaling process with data from students’ test performance employing Cognitive Diagnostic Assessment measures. Hopefully, the results will benefit high school English teaching and the preparation of tests either before or between two exams under the context of multiple-tests of the NMET at both group and individual levels.

Room 714

Evaluation Form For College English Oral Presentations

THE PLA Information Engineering University CHEN Zhuo LIN Yi

In student-centered college English classes, students are often asked to make oral presentations in English and are accordingly judged as a part of their formative assessment. However, ESL teachers mostly have no standard or scientific assessment tool to follow. They often give the score quite subjectively from their own experience with reference to the pronunciation, the content, the effect or other aspects of the students’ presentation. This kind of assessment has many drawbacks. On the contrary, a scientifically designed evaluation form can help the teachers and the students know more about their problems and thus provide more

meaningful feedback and guidance. In order to achieve this goal, this study gathered twelve oral presentation evaluation forms of overseas universities and designed an oral presentation evaluation form with high reliability and validity through scientific analysis and experimental tests.

Issues in Assessing the Speaking Skill in EFL

King Saud University Reima Al-Jarf

At my college, translation students take 4 speaking courses in the first 2 years of the translation program. The courses aim at developing students' ability to talk about a particular topic for at least 5 minutes without preparation, with fluency, correct pronunciation, stress, intonation, grammar, choice of words and well-organized cohesive ideas. In each course, the students take 2 in-term tests and a final exam. Speaking tests can be conducted in the language lab or face-to-face. For both testing modes, test instructions must specify the number of details and type of task that must be taken into consideration while talking about the topic, types and numbers of details, correct sentence structure, pronunciation, stress, intonation, fluency... etc. Topics chosen should be comparable to but not a repeat of what students have practiced in class.

If the speaking test is given in the language lab, questions are printed on paper and handed out to the students. Two versions of the test questions should be prepared (same questions, but in a different order). Students write their names on the question paper. A week in advance, the teacher asks the students to bring an MP3 player, practice using the MP3 player at home, test it at the beginning of the test session. Students read the questions, think about them for a while, then record their responses to questions in any order provided that they say the question number. To minimize noise, the students put on the headsets and record answers in a soft voice. Students should not write down answers and read them into the MP3 player. They should not be allowed to re-listen to their answers, erase and re-tape answers. While listening to the recordings and grading the students' responses, the teacher can take notes on the student's question paper while listening to the student's responses, noting strengths and weaknesses.

If the speaking test is given face to face, they are tested individually. Each student is handed out the tasks (several questions covering different themes, skills and language functions) printed on pieces of paper which the students draw from a basket. the student completes several speaking tasks in 10-15 minutes.

A speaking assessment rubric must be prepared. The rubric consisted of 3 skills: (i) idea generation or content; (ii) grammar and vocabulary; and (iii) pronunciation and

fluency; and 3 performance levels: (a) excellent, (b) average, (c) poor. For each skill category, the characteristics of the excellent, average and poor performance are given in the cells with the upper and lower limits of the marks allocated for each performance level given in the divider. The total marks allocated to each question on the speaking posttest is 50% for content or idea generation, 30% for grammar and vocabulary, and 20% for pronunciation and fluency.

Examples of speaking test questions, testing and grading procedures, advantages and disadvantages of face-to-face and language lab testing will be given in detail.

浸入式教学理念下外教口语课教学模式改革与

英语专业体验式教习模式的构建

China University of Mining and Technology ZU Daqing

浸入式（Immersion）教学是指用第二语言作为教学语言的教学模式。国内高校英语专业的外教口语课存在着一些典型性的共性问题。浸入式教学理念下体验式英语口语教学在发挥体验式教学理论优势的基础上，主张根据循序渐进的学习规律将口语教学分为语言信息输入、口语能力构建、口语策略发展、跨文化交际意识培养四个阶段。在体验式英语口语教学模式中，学生被视为具有极大潜能和认知能力的学习主体，教师引导他们通过体验真实情境形成对英语的理解和把握，从而构建出符合自身认知特点的知识体系，然后重新实践、体验，达到应用自如的境界，同时外教教学效果和形成性评价方式得以提高和完善。

交际性语言测试对高中英语教学的引导作用研究

Xuchang University LI Zhiying

With the development of economy and society, the education in senior high school is gaining its proper status in China. In senior high schools, English is one of the important subjects, and in the National College Entrance Examination (NCEE), English is also an important part. Nowadays, English teaching is centered with the development of students' ability to use English in real context in senior high schools. According to the New English Curriculum Standard, The purpose of teaching English is to improve the students' integrated ability to use English, especially the ability to use English in communication, and the teaching method has also changed from the traditional teacher-centered approach to student-centered approach. The latter one is more communicative.

Testing can affect teaching, and the effect is called backwash. Language test backwash is said to be either positive or negative. If a testing procedure can encourage teacher to give some good practice in teaching, the positive backwash is effective. Apparently, the testing in senior high schools has a less positive backwash. So, English teaching in senior high school needs more helpful testing methods to give it a more positive backwash. Then, Communicative Language Testing (CLT) is required.

The thesis is concerned with the CLT. The author talks about the theory, the features, the development and the application of CLT. Both quantitative and qualitative study is used in the research. In order to collect the information about the teachers and students. The following experiments in two groups can show that the CLT gives a positive backwash to English teaching. At the end of thesis, some suggestions of the way concerning with application of CLT better in senior high school are provided: Teachers should assign authentic, communicative test tasks to help the students improve their communicative ability. Teachers should pay more attention to testing and its backwash effect to teaching approach. Teachers in senior high school should value the CLT and use as many as CLT in the classroom diagnostic test or progress test as the guidance for the teaching and learning. Furthermore, the teachers and educators should develop the CLT and make it more suitable to apply in a large-scale test as NCEE.

Effects of Peer Assessment on EFL Writing in MOOCs: The Learners'

Perceptions

School of Foreign Languages, Shenzhen University LUO Sha

Peer assessment is a socially situated dialogic process, and regarded as having such potential benefits as encouraging improvement and facilitating collaboration. While much of the peer assessment observed in L2 writing classrooms occurs only incidentally as supplementing teachers' evaluation, such opportunities are substantive in language MOOCs with their large enrollment scales and abundant learner-to-learner interactions, as well as increased learner autonomy and cooperation. But the research on how a reliance on peer assessment may affect learner writing is still limited, especially in the context of EFL MOOCs. Drawing on the socio-cultural perspective, with data collected from questionnaires, interviews and reflective journals, this research seeks to address the following research questions: 1) Do learners' perceptions of formative and evaluative peer assessment on their EFL writing change over time while participating in MOOCs? 2) What factors may affect the learners' perceived effects of both forms of peer assessment? The results indicate that the learners' perceived effects of both formative and evaluative peer assessment improve significantly over a semester, with such affecting factors as understanding of rubrics, perceptions of learning gains, attitudes towards the peer assessment practice, and learner collaboration.

Pragmatic Competence and Lessons for University English Teaching, Language

Testing and Assessment

四川大学外国语学院 杨雅平

The ultimate goal of learning a language is to use it for successful communications, which requires pragmatic competence. However, the frequent occurrence of pragmatic failures impedes students of English as a Second Language (ESL). The lack of pragmatic competence among English learners suggests that educators should make adjustments in teaching, language testing and assessment. University English teaching should be grounded in practical use of English, with an emphasis on cross-cultural awareness and communication. Furthermore, traditional English testing and assessment should be more scientific in order to improve the classroom teaching quality of university English.

Room 715

Implementing Assessment for Learning through an Internet Forum:

Feedback from both EFL Learners and Tertiary English Teachers

School of English and Education , Guangdong University of Foreign Studies

YANG Manzhen LU Jiantao ZHU Yuanyuan ZHANG Decheng LIANG Li

Theory and principles of classroom assessment have been constructed comprehensively and systematically (e.g., Hall et al., 1997; Clarke, 1998). However, research devoted to observing and analyzing classroom assessment practice and procedures in EFL/ESL contexts is much rarely retrieved, especially in those “test culture” oriented countries like China, Korea and Japan. It is valuable not only to investigate how may English teacher formatively guide and assess EFL learners, but to explore new platforms where classroom assessment can be better implemented by English teachers and more easily accessed and participated by EFL learners. The present study examines the assessment function from several learning activities (i.e. group presentations, writing tasks and useful learning materials submission) held on a web learning forum. Differential participation rules are set up by several tertiary English teachers recruited in a project that promotes assessment for learning in their own teaching. Their students are required to obtain course credits, a unit representing their learning efforts. Both students’ and teachers’ comments and reports for these learning activities are collected. These contain problems and questions in the assessment and guidance work of the teachers, as well as issues in earning course credits raised by the students. However, all the teachers witness more frequent for learning interactions both between students and between teachers and students, possibly because the web environment provides a more comfortable way of communication and cultivates students’ sense of self-esteem. Based on the teachers’ and students’ remarks, adjustments to the process and procedures of these activities are proposed.

Recognizing and measuring self-regulated EFL learning within an online formative assessment module

School of English and Education, Guangdong University of Foreign Studies

LIANG Li YANG Manzhen LV Jiantao ZHU Yuanyuan ZHANG Decheng

Emerging since 1980s, self-regulated learning (SRL) and formative assessment have been relatively new and heated topics in the fields of learning strategy and assessment (eg. Berry, 2011; Li, 2012). Educational testing should not merely be the assessment of learning but the assessment for learning (Jin, 2000; Liu, 2013) and formative assessment is acknowledged to possess huge potential in promoting an assessment for learning. Generally speaking, in a formative assessment environment, students are self-regulated learners if they are motivationally, meta-cognitively and behaviorally active participants in their own learning process, instead of relying on teachers or other resources (Zimmerman, 1986; 1989). From this definition, it is argued that SRL is one of the central parts of formative assessment and well-worth an in-depth study. But Li (2012) observes that some essential issues of formative assessment remain rarely touched yet, such as the question of how formative assessment promotes learners' motivation, self-efficacy beliefs, and regulation of learning, and so on.

Given the above question mentioned, this paper reports a case study of SRL in an online blended formative assessment module (educational online game of credit coins) in the context of non-English-major college English instruction in GDUFS. The intent of the present research is to exemplify how SRL, an active area in educational psychology, can help to illustrate and understand the mechanism and process of SRL in the online formative assessment environment. In the present study the theoretical framework of SRL is based on Bandura's social cognitive theory (i.e. Reciprocal determinism) (Bandura, 1986; 2001). This paper proposes an analytical SRL model for recognizing and analyzing the key components of SRL in EFL learning. This SRL model can illuminate the relationships among three aspects of SRL in the online formative assessment module: students' self-report surveys, patterns of learning behavior in the online formative assessment environment, and learning achievement. A four-subscale survey questionnaire is designed and carried out together with observation and collection of students' remarks in reflection: self-efficacy beliefs, intrinsic motivation, extrinsic motivation, regulation of learning. Statistical analyses produces three main findings. First, students' scores in motivation scales in this case can account for whether and to what degree students can actively engage in EFL learning activities in the formative assessment module metacognitively, motivationally, and behaviorally. Second, the effects of students' self-reported motivation on their learning achievement will be revealed. Third, students'

actual behaviors in self-regulating their learning such as choice of task, effort, etc, are also a significant indicator of their motivation in SRL.

An Empirical Research on Relevance of Subjective and Objective Questions in English IBT:

A Case Study of IBT at Beijing University of Technology

School of Foreign Languages, Beijing University of Technology LI Song HE Cencheng

This paper introduces application of automatic grading system in English IBT for freshmen in 2015 at Beijing University of Technology. Based on statistical analysis of those samples, level of validity and reliability is investigated from the perspective of the relevance of subjective and objective parts in the test. The research results show that automatic grading on the basis of the relevance of subjective and objective questions is controllable compared with human grading, which proves the feasibility, economy and efficiency of automatic grading in English test.

Designing and Validating the Test Tasks of a Computer-based Speaking-Listening Test

South China Normal University HUANG Liyan
Guangdong University of Education CHENG Fenxiang

Taking the English listening-speaking test component of the ZHONGKAO in China as an example, the present study investigated the process and methods of designing and validating test tasks of a high-stakes computer-based test. By reference to an analysis on English listening-speaking testing in China and abroad, and the findings derived from interviews with three experts and a questionnaire survey with 2733 English teachers, the authors proposed three test tasks: Imitating & Reading, Listening for Information, Talking on Situations. All of them were validated through expert judgment, interviews with 13 ELT consultants, pilot testing involving 422 students and a comparison between student performance ranking made by the computer and by the teachers. Findings indicate that the construct of three test tasks is consistent with the requirements in terms of listening-speaking teaching objectives proposed by the Compulsory Education English Curriculum Standards. They are at the appropriate level of difficulty and differentiation, and have a positive washback on teaching practice. In addition, there is no significant difference between computer ranking and teacher ranking. Generally the proposed test tasks are valid and

applicable for a high-stakes computer-based English listening-speaking test.

Using Generalizability Theory (GT) and Many-facet Rasch Model (MFRM) to
investigate

Computerized English Listening and Speaking Test (CELST) rating

South China University of Technology XU Ying

Guangdong Teachers College of Foreign Language and Arts

Guangdong University of Foreign Studies

ZENG Yongqiang

This study aims to examine the rating of Computerized English Listening and Speaking Test of National Matriculation English Test (NMET) (Guangdong Version). Five experienced raters from different sources were invited to mark 117 senior high school students' performance at a mock test of CELST NMET (Guangdong Version). The results of GT analysis indicated that the task was the major source of total variance (46.8%). Under the current situation where raters are nested within each task and each test-taker ($R : (p \times T)$), the generalizability coefficient (0.647) was acceptable. MFRM results showed that raters differed significantly in severity, but they were internally consistent. The bias analysis for rater source and task interaction showed that college teacher raters were significantly more severe with Task 1 and more lenient with Task 3, but the middle-school teacher raters were just the opposite. The above results not only provide empirical evidence for the scoring design of CELST NMET (Guangdong Version), but also demonstrate that the two methods are complementary in the research angle.

Methods of Scaling and Equating in Constructing Item Bank of Computer-based English Placement Test

湖南大学外国语与国际教育学院 肖云南 林椿

Under the situation that the National Matriculation English Tests haven't been equated onto a common scale, comparison and contrast cannot be drawn between items in tests, and between the scores of examinees from all different provinces and areas. The various developments of English ability levels in natural classes when they obtain admission into universities and colleges remain a major hurdle to English teaching. In order to improve students English ability effectively, College of Foreign Language and International Studies of Hunan University implements a hierarchical teaching based on the students' English proficiency and develops a Computer-based College English Placement Test. The fact is that there is a large population of freshmen every year and that there is an immediate need of the placement results, so in order to meet the needs we design the placement test as a computer-based test and construct the item bank.

In the construction of the computer-based English placement test, we need to set language proficiency levels under the circumstance that there is not a common National English Language Proficiency Scale, need to set the cut-off scores of different levels, and then need to randomly match the items after estimating item parameters of the data. We adopt the Canadian Language Benchmarks which is basically fit for the case of Chinese English learners, and we make the cut-off scores based on language test results and teaching expert experience. Two different equating methods based on Item Response Theory are adopted and compared when estimating the data in Computer-based College English Placement Test of Hunan University. Data are collected using the fixed forms examination based on common-item nonequivalent groups design. The two-parameter logistic model and the three-parameter logistic model are applied to estimate the parameters of items in different test parts and different item types respectively. It is shown that the two-parameter logistic model is ideal to the dichotomous items including multiple-choice items and open-ended items in the Placement Test. Then, an anchor item bank of 123 items is developed with concurrent calibration and BILOG—MG. The bank will evolve with the help of anchor items in the following tests, providing comparative scores of examinees over test forms and over years.

The authors hope that these technologies and methods will be put forward at the meeting and hope to discuss with other researchers in this field to make further progress.

Automatic and human rating differences in CELST in Guangdong NMET:

A many-facet Rasch analysis

Guangdong University of Foreign Studies CAO Linlin

The CELST (Computerized English Listening and Speaking Test) in Guangdong NEMT mainly focuses on the measurement of communicative ability and language using behaviors and has received widespread recognition. However, as a subjective test format, rating reliability is quite easily influenced by various factors. Thus, effectively control raters' quality is an important method to insure the quality of subjective tests. However, it is not explicit how good the training effect is, where the raters' rating bias lies and which aspects of the raters should be directed to improve. In order to improve the rating reliability for this large-scale, high-stakes test, in recent years the decision-making section plans to conduct the automatic rating reform to this computerized English listening and speaking test in Guangdong NMET. Thus the reliability of automatic rating and the differences between automatic and human rating have become serious problems.

A number of researchers abroad have employed mathematical models to analyze raters' rating results, and have made some explorations. One of the widely used models is the Many-Facet Rasch Model, which derives from the Item Response Theory in psychological measurement area.

This study made use of the materials in the listening and speaking test in 2013 Guangdong NMET, tested 119 senior three students and analyzed the rating result. The students were rated by three types of raters, aiming to analyze in detail the different types of raters' rating difference. According to their background, the raters are divided into three groups—college teachers, high school teachers and automatic rater. Through Rasch model, this paper analyzed and explored differences between automatic and human raters on aspects like rater reliability, severity, central tendency, random effect and unexpected ratings. What's more, it concretely evaluated and compared each type of rater's severity and reliability, analyzed possible causes of each type of rater's bias on specific test takers during rating, and extracted abnormal scores. The results showed that all the three types of raters possessed good inter-rater reliability, though the automatic rater indicated less intra-rater reliability than the other two types of raters under the stringent infit limits. There didn't exist central tendency and random effect among the three types of raters; and the automatic rater and the college teacher rater had a few unexpected ratings to some students on certain item parameters.

I hope this research can offer concrete statistical basis for the automatic rating reform of

the English listening and speaking test in Guangdong NMET, and encourage the application of MFRM in actual score monitoring.

Author Index

Alexey Korenev	17	133				
April Jiawei Zhang	7	54				
BAI Ying	9	66				
Barry O’Sullivan	3	24				
BIN Ke	11	82				
Brechtje Post	16	122				
CAI Wenyan	7	55				
CAI Yuyang	16	127				
CAO Linlin	19	146				
Carolyn Westbrook	17	133				
Chan Kwong Tung	1	81				
CHEN Chao	7	55				
CHEN Dajian	1	89				
CHEN Qiuxian	7	9	52	68		
CHEN Yangfang	4	35				
CHEN Zhuo	18	136				
CHENG Fenxiang	19	143				
CHENG Liying	19	25				
DENG Yanping	9	64				
FAN Jinsong	4	17	32	132		
FAN Meng	7	53				
GAO Ying	11	87				
GU Gang	11	85				
GU Xiangdong	5	8	9	10	17	36 41
		63	65	74	128		
GUI Shichun	20	26				
GUO Ying	15	111				
HAN Baocheng	3	21				
HAN Ye	8	60				
HAO Chenqian	9	68				
HE Cencheng	1	18	77	143		
HE Lianzhen	3	12	21	89		
HONG Yue	9	62				
HU Rong	6	47				
HU Wenting	10	77				
HUANG Liyan	19	143				
HUANG Rui	5	41				
HUO Yongshou	14	105				
JIA Guodong	16	123				
John de Jong	19	26				
LI Aike	16	122				

LI Danli	14	111			
LI Song	18	143			
LI Tian	10	79			
LI Xuelian	6	48			
LI Xueping	17	129			
LI Zhiying	18	139			
LIANG Li	18	141	142		
LIANG Yansong	17	128			
LIN Yi	18	136			
LIU Jianda	3	21			
LIU Jianzhu	6	44			
LIU Li	11	16	87	123	
LIU Liandi	16	125			
LIU Qian	14	108			
LIU Xiurong	14	102			
LU Xiaoxian	14	111			
LUO Sha	18	140			
LV Jiantao	16	18	124	141	142
MA Rui	11	68			
Maggie Ma	15	118			
MENG Lei	5	14	41	103	
MIN Shangchao	12	89			
Muhiddin Timuçin					
Denktaş	16	120			
Nguyen Thi Minh Loan	16	118			
Nick Saville	3	22			
OYEDOKUN-ALLI	17	134			
PAN Mingwei	12	90			
PENG Chuan	10	12	74		
PENG Yuanling	10	78			
QIU Xuyan	16	121			
Reima AL-Jarf	8	18	57	137	
Roxanne Wong	11	86			
Saefurrohman	6	46			
Sercan SAGLAM	8	59			
SHAO Zheng	14	109			
TANG Xiongying	7	50			
Tanya Ershova	17	133			
Tatiana Ershova	15	117			
TIAN Jie	8	63			
WANG Fang	8	61			
WANG Yihong	11	85			
WASIU ADEMOLA	17	134			
WEI Wei	6	47			

WU Jinyu	11	86				
WU Zunmin	3	21				
XIE Qin	17	135				
XU Ying	19	144				
XU Yiyi	8	62				
XU Yueting	6	49				
YAN Qiaozhen	10	74				
YAN Youyang	9	73				
YANG Ji	9	17	66	129		
YANG Manzhen	18	142	143			
YANG Maoxia	14	110				
YANG Ruijin	5	9	14	42	64	103
YANG Zhiqiang	14	104				
YE Wei	6	44				
YU Chengyuan	10	75				
YUAN Shuhou	7	52				
Yvonne Merry	17	134				
ZENG Yongqiang	3	19	21	145		
ZHANG Decheng	18	142	143			
ZHANG Lidong	13	95				
ZHANG Ping	8	63				
ZHANG Siyu	8	63				
ZHANG Xiaoyi	13	10	77	97		
ZHANG Xinling	7	8	57	63		
ZHAO Guanfang	4	17	33	133		
ZHENG Ying	9	73				
ZHENG Yongyan	4	33				
ZHENG Yujing	8	64				
ZHONG Yu	5	42				
ZHOU Jiming	10	81				
ZHOU Yan	8	11	63	83		
ZHU Lin	7	57				
ZHU Meifang	6	46				
ZHU Qiyun	15	113				
ZHU Youyi	14	107				
ZHU Yuanyuan	18	142	143			
ZHU Zhengcai	13	95				
ZHUANG Yanyan	9	73				
ZOU Shaoyan	10	76	13	96		
ZOU Shen	3	12	21	90		
ZU Daqing	18	138				
陈灿	14	107				
陈桦	12	92				
陈文存	5	37				

陈璇	11	84						
程蒙蒙	14	104						
程欣	12	92						
范婷婷	17	136						
范香娟	8	57						
冯莉	12	89						
甘凌	10	77						
高霄	9	71						
龚晨	13	99						
黄剑	15	114						
黄玮莹	4	31						
黄妍琪	9	70						
贾国栋	4	31						
江进林	4	34						
揭薇	4	27	4	28				
金艳	3	4	21	27				
李廉	13	95						
李乔娜	13	97						
李清华	4	32						
李英春	15	116						
林椿	16	125	19	145				
刘春燕	7	51						
刘辅兰	5	37						
刘芹	4	30						
刘洋	15	116						
罗柳萍	17	127						
罗少茜	15	114						
马晓梅	17	130						
孟亚茹	17	130						
牟欣	4	34						
穆雷	3	6	7	12	21	43	44	
		56	93						
亓鲁霞	9	67						
邵争	15	114						
盛艳玲	14	107						
施雅俐	5	35						
泰中华	4	30						
唐莉莉	12	91						
陶百强	5	39						
童扬芬	16	125						
王华	9	72						
王凯旋	13	102						
王淑花	4	29						
王巍巍	6	12	43	44	96			

王伟强	15	116				
王晓彤	4	31				
吴苓	13	101				
吴宜静	13	102				
伍志伟	6	7	44	56		
肖立宏	6	43				
肖巍	9	70				
肖奕	13	102				
肖云南	16	19	125	145		
徐莎莎	5	38				
徐嗣群	14	109				
许艺	6	7	12	44	56	93
严明	3	12	21	89		
杨茂霞	15	114				
杨秀芬	13	98				
杨雅平	18	140				
杨洋	11	88				
姚雯静	11	83				
张聪	17	128				
张放	9	69				
张惠玲	8	60				
张剑	17	127				
张锦涛	13	101				
张妙	5	36				
张敏	13	100				
张萍	10	13	75	99		
张晓丹	13	98				
张晓蕾	10	75				
赵护林	12	93				
赵辑	14	106				
赵路	9	70				
赵中宝	15	115				
周艳琼	12	93				
朱汉雄	13	102				

Guangdong University of Foreign Studies

广东外语外贸大学

Guangdong University of Foreign Studies (GDUFS) is a key provincial university and a high-tier one with distinctive features of internationalization in South China in its education of internationally-oriented personnel and its research on foreign languages & culture, international trade and international strategic studies. The University serves 20,189 full-time undergraduate students, more than 2,800 graduate students and over 12,000 international and continuing-education students.

History

GDUFS was created by a merger in 1995 of Guangzhou Institute of Foreign Languages and Guangzhou Institute of Foreign Trade. The former, one of the three renowned institutes of foreign languages in China under the direct jurisdiction of the Ministry of Education (MOE), witnessed its establishment in 1965. The latter, directly under the Ministry of Foreign Trade and Economic Cooperation (now the Ministry of Commerce), was founded in 1980 as one of the first institutions in China offering an academic program in international economy and trade. In October 2008, Guangdong College of Finance and Economics also became part of the University.

Campuses

GDUFS is situated in Guangzhou, a city with a long history and rich cultural legacy and the economic hub in South China. The university boasts three campuses with a total area of 153 hectares. The North Campus, which covers an area of 63 hectares, is located at the northern foot of the Baiyun Mountain; the South Campus, with an area of 73 hectares, is situated in Guangzhou Higher Education Mega Center; and Dalang Campus, having an area of 17 hectares, lies in Dalang, Baiyun District. The university enjoys fine surroundings featuring green trees, crystal streams, scenic bridges and fresh air.

Teaching staff

Since the establishment of GDUFS, a large number of renowned scholars, such as Liang Zongdai, Gui Shichun and Li Xiaojun, have devoted to teaching and scholarship here, accumulating rich spiritual and cultural wealth for the University. It has a teaching staff of 1,131, among whom 53.85% are full or associate professors and 89.92% have masters and/or doctoral degrees. The University boasts one member of the Disciplinary Appraisal Panels under the Academic Degrees Committee of the State Council, 16 members of the teaching advisory boards of various disciplines of the Ministry of Education, 10 teachers enjoying the Government Special Allowance of the State Council, three candidates of the New Century Hundred, Thousand and Ten-Thousand Talent Project at the state level, 13 candidates of the Program for New Century Excellent Talents of the Ministry of Education, one national distinguished teacher, 6 provincial distinguished teachers, four distinguished professors of Pearl River Scholar, one national- and 44 provincial-level cultivation candidates of Guangdong Thousand, Hundred and Ten Project and 52 Yunshan Scholars. Besides, the

University employs over 86 guest professors and nearly 200 foreign experts on a long-term basis.

Disciplinary Development

The University consists of 21 faculties and one self-governing college (South China Business College) that offer 64 undergraduate programs in eight disciplines: literature, economics, management, law, engineering, science, education and art studies. There are eight national programs with distinctive features and another eight prestigious programs of the Province. With 20 foreign languages offered to students, GDUFS is the only university in South China that offers such a variety of languages instruction. The University was accorded the qualification to confer master degrees in 1981 and doctoral degrees in 1986, thereby becoming one of the earliest institutions in China to earn postgraduate accreditation since the country resumed its postgraduate education programs. The University is home to one national key academic discipline, six provincial key disciplines, one post-doctoral research station, one Category I discipline PhD program, 11 Category II discipline PhD programs, 10 Category I discipline masters programs, 36 Category II discipline masters programs, and seven professional masters programs.

Student Education

The University upholds the motto of *Pursuit of Integrity, Practice and Cross-cultural Learning* and defines as its mission the training of high-caliber global citizens through the integration of content studies and foreign language training for globally minded graduates with international vision and social responsibility. They should be well-versed in international norms, excellent in ethical cultivation and competence in their chosen fields of study, strong in cross-cultural communication and hands-on creativity, and ready to engage in global competition and cooperation. The University is one of the 19 universities worldwide that are qualified to recommend high-end translation professionals to the United Nations and the only foreign studies-oriented university in the CAMPUS Asia program. The University is home to the MOE-designated Teaching and Research Center for Asian and European Minor Languages, and the MOE's Pilot Scheme for Innovative Mode of Talent Training for cultivation of internationally competitive business professionals. According to *MyCos*, an educational consultancy firm, the graduation index and entry-level salary of this University's graduates ranked top among the universities across China from 2008-2013 including those under the 985 and 211 categories. The University enrolls applicants from over 20 provinces, autonomous regions and municipalities directly under the State Council, Hong Kong, Macao and Taiwan, all in the first-round or early admission category.

Academic Research

GDUFS is home to one of the MOE's national key research centers for humanities and social sciences, i.e. Center for Linguistics and Applied Linguistics; one of the Cultivation Base for Regional and National Studies of the Ministry of Education, i.e., Center for Canadian Studies; 4 provincial-level key research centers for humanities and social sciences in institutions of higher learning in Guangdong, i.e. Center for Foreign Language and Culture, Center for International Trade and Economics, Center for Translation Studies, and Research Institute of Cantonese Merchants; one Guangdong Soft Science Key Research Base; one Guangdong Local Legislation Research Evaluation and Consultation Service Base; one key

research center for humanities and social sciences in Guangzhou Research Institute of Guangzhou International Business and Trade Center; and key lab in Guangdong, i.e. Key Lab of Language Engineering and Computing. The University also boasts Guangdong Research Institute on International Strategies and Research Institute of International Services Outsourcing. GDUFS publishes four academic journals of national impact: *Modern Foreign Languages*, *International Economics & Trade Research*, *Journal of Guangdong University of Foreign Studies*, and *Journal of Strategy & Decision-Making*.

International Cooperation

GDUFS attaches great importance to a full range of cooperation and exchanges in international education. To date, it has established cooperative ties with more than 274 universities and academic or cultural institutions in 38 countries or regions. In addition, the University has co-established three Confucius Institutes overseas at Sapporo University, Japan, Ural State University, Russia and Universidad Catolica de Santa Maria, Peru.

Teaching Facilities

GDUFS is fully equipped with advanced teaching facilities which are taking the lead among similar universities in the whole country. It boasts advanced campus network and management platforms, two provincial-level key labs and eight provincial-level teaching demonstration centers, and a lab area totaling over 52,000 square meters. Covering a total area of 50,000 square meters, its libraries in the north and south campuses have a collection of 2.91 million hard-copy books, 2.1 million e-books, 74,345 foreign journals and 92 Chinese or overseas academic databases, providing network integrated management for book purchase, classification, cataloging, circulation and information search on library resources in 20 foreign languages. Its fixed assets stand at about 2.9 billion Yuan.

Development Prospects

The University adheres to the idea of whole-person education and pursuit of excellence and nurtures a campus culture of diversity, elegance and flexibility. It keeps pursuing substantive development, further reform and innovation and vigorously promotes the internationalization of education-- internationalization of teachers, students, research and management--in order to be a high-tier internationalized university with distinctive features and excellent quality and enjoying high social reputation.

National Key Research Center for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies

教育部人文社科重点研究基地外国语言学及应用语言学研究

The National Research Centre for Linguistics and Applied Linguistics (CLAL), which was previously called the Institute for Linguistics and Applied Linguistics by the National Education Commission in 1984, was founded under the auspices of the Ministry of Education in 2000. In 1988, CLAL's program "Linguistics and Applied Linguistics" was judged by the Ministry of Education to be the nation's leading discipline, and this was re-confirmed in 2002. CLAL currently boasts the only national key research centre for linguistics and applied linguistics in China.

Under the leadership of Prof. Gui Shichun, Prof. Li Xiaojun, Prof. He Ziran, Prof. Huang Jianhua and Prof. Wang Chuming, the faculty of linguistics and applied linguistics has developed into one with various research specialisms, each with its own characteristics. CLAL hosts three well-established research branches: foreign language teaching techniques and evaluation, second language studies, and theoretical linguistics, covering areas such as second language acquisition, psycholinguistics, language testing, pragmatics, sociolinguistics, lexicography, syntax and forensic linguistics.

A scientific research team with a balanced age structure, high academic credentials and scientific research capability has been set up in CLAL. It is currently staffed by 12 full-time and 36 part-time researchers, 39 of whom are professors. The CLAL library houses more than 13,000 books on linguistics and applied linguistics and subscribes to 60 Chinese journals and 59 English-language journals, covering nearly all the most important linguistics journals in the world. In addition to the main CLAL website, there are websites for branches of linguistics, such as pragmatics, second language acquisition, forensic linguistics, lexicography, and ESL writing. *Modern Foreign Languages*, a journal on linguistics and applied linguistics published by CLAL, ranks among the top 100 social science journals in China.

CLAL has achieved great success in second language acquisition, pragmatics, psycholinguistics, language testing, corpus linguistics and lexicography and, since its foundation, has won a considerable number of awards at both national and provincial level. In the past five years, CLAL researchers have undertaken 62 research projects, including national social science projects, key Ministry of Education projects, and social science projects for Guangdong Province. Researchers have published more than 26 books and 381 journal papers, among which 43 papers in international journals such as *Applied Linguistics*, *Brain and Language*, *Bilingualism: Language and Cognition*, *Language Learning*, *Journal*

of Neurolinguistics, Cognition, PLoS ONE, TESOL Quarterly, Second Language Research, ELT Journal, Language Testing, Language Assessment Quarterly, Language Teaching Research, Language Sciences, Language Awareness, Journal of Pragmatics, International Journal of Lexicography, Pragmatics, System.

Higher Education Press

高等教育出版社

Founded in 1954, Higher Education Press (HEP) is committed to meeting the needs of students, teachers, and professionals by providing quality educational resources in the major disciplines, which advance learning, knowledge and research. We have maintained and extended our reputation as a leading educational publisher in China for over half a century.

HEP expands its business at a wide range of textbooks, teaching materials, periodicals, audio-visual, electronic and digital publications and e-resources for learners throughout the country. We meet the needs in different phases of education from higher education, vocational education, school education through life long learning. We lead the market by producing nearly 10,000 titles, including almost 2,500 new titles, with circulation of 113 million copies annually.

Now, as a subsidiary of the China Education Publishing & Media Group Co., Ltd. (CHINA EDUCATION), HEP enhances the value of its content by innovatively offering educators learning tools, online resources, digital and highly-customized teaching and learning solutions. We are evolving from a traditional textbook publisher to an educational solution provider through realigning resources, streamlining the operation and focusing on service.

With its branded publications, HEP has gained great reputation and popularity among generations of Chinese readers. Now we are looking forward to closer cooperation with partners from both home and abroad.

International Business

HEP attaches great importance to international partnerships and has cooperated successfully with many publishers throughout the world. We have delivered sales with more than 100 publishers in over 50 countries and regions in 16 languages.

HEP is the only Chinese publishing company amongst the world's top 50 publishers, as well as the only Chinese publisher to be granted the "WIPO Innovation Award".

HEP's international business is currently focused on Teaching Chinese as a Foreign Language (TCFL), E-journals (Frontiers in China), Academic Publications, and Digital Educational Resources.

Vision

We are home to a group of talented, creative and hard-working people who aspire to pursue excellence and constantly innovate in educational publication. These people are motivated to reach their full potential as HEP grows. We are committed to serving our

customers at home and abroad by providing valuable educational resources and solutions in effective and innovative ways.

International Center for Language Research and Development

国际语言研究与发展中心

The International Center for Language Research and Development (ICLRD) is cofounded by Higher Education Press and Guangdong University of Foreign Studies. The Center's mission is to promote the development of language education through research, testing and evaluation, teacher training and academic exchange. It provides consulting services to universities and government on the reform and development of language education.

ICLRD has undertaken the National Social Science Fund Project "Experiencing foreign language teaching: theory and practice". Rated "excellent" in project evaluation, the project yielded a range of important research results, which greatly promoted the development of ELT. Since 2005, ICLRD has launched provincial "ELT reform research project" with the provincial government in more than 20 provinces to strengthen regional education reform and promote the exchange of research results.

Considering the status of foreign language education in China, the language assessment policy and development as well as the current needs of Foreign Language talents in society, ICLRD is committed to developing the assessment programs in accordance with the tendency of talents cultivation of Foreign Language education. The national wide assessment programs and business is under construction in areas of Business English, Vocational English, Foreign Language Translation and Interpretation, etc.

To facilitate ELT professional development, ICLRD has been organizing series of International, national and local teacher training programs in cooperation with the Higher Education Institutions Teacher Online Training Center under the Ministry of Education, and several National-level Demonstration Centers of Teacher Development and the National Foreign Language Teaching Advisory Board under the Ministry of Education. The training programs mainly focus on teachers' teaching ability and research ability and language assessment literacy development, which has been greatly welcomed and appreciated by EFL teachers.

With the efforts of the researchers all over China working with the Center, ICLRD has published a series of journals such as "Foreign Languages in China", "Linguistic Research", "Contemporary Discourse Studies", "Philosophy of Language" and so on, with the first two selected as Chinese Social Science Citation Index (CSSCI) source journals.

In addition, ICLRD has also organized a number of major events and activities like "Micro-course Contest of Foreign Languages in China" and Asia TEFL Conferences to promote the integration of language teaching, research and technology.

Contact Information:

Email: irdcfl@126.com



Wechat:

语言学期刊《现代外语》简介

(新网址: <http://xdwy.cbpt.cnki.net>)

《现代外语》创刊于1978年,是国家新闻出版总署批准出版,由广东外语外贸大学主办,向国内外公开发行的语言学及应用语言学学术期刊。创始人是我国著名的语言学家王宗炎教授和桂诗春教授,何自然、张达山、钱冠连、黄建华、宁春岩、王初明等教授先后负责本刊工作,现任主编为冉永平教授。

1997年,《现代外语》被确立为语言学与应用语言学的专业性学术期刊,学科范围包括理论语言学、语音学、音系学、句法学、语义学、语用学、社会语言学、心理语言学、认知语言学、法律语言学、词典学、语言习得、语言测试、计算语言学、语料库语言学、语言教学等。自2000年起,本刊由广东外语外贸大学外国语言学及应用语言学研究——教育部人文社科重点研究基地承办。自2014年起,由季刊改为双月刊。

《现代外语》是“中国人文社会科学核心期刊”,“中文社会科学引文索引(CSSCI)来源刊”,中国期刊方阵“双效期刊”,“全国百强社会科学学报”,“北大中文核心期刊”,“美国教育信息资源数据”(ERIC)收录,“中国学术期刊综合评价数据库统计源期刊”,“中国社会科学期刊精品数据库收录期刊”,“中国期刊全文数据库收录期刊”,“中国人民大学报刊复印资料《语言文字学》转载刊物”,“国家哲学社会科学学术期刊数据库”等收录。在新闻出版总署报刊司全国抽查的100种期刊中,《现代外语》是英文摘要编辑质量合格、仅有的15种学术期刊之一;根据中国人民大学人文社会科学学术成果评价研究中心和中国人民大学书报资料中心“复印报刊资料”转载论文综合评价数据和评审专家组评议,《现代外语》入选为2012版“复印报刊资料”重要转载来源期刊。2014年,再获“全国高校精品社科期刊”。还多次获得其它荣誉。

《现代外语》秉承“重质量,显特色,引领学术”的办刊宗旨,聘请国内外知名专家组成顾问委员会和编审委员会;按照国际惯例,实行全程匿名评审制度,并及时向作者反馈审稿意见;登载理论性、实证性、前沿性及综述性的研究论文,坚持以重大理论、重要现实问题等为导向,择优录用能够指导学科发展的原创性稿件。

主要栏目: 1) 语言学:涉及语言学主要分支的相关研究;2) 二语研究及其应用:涉及二语习得、外语学习、外语教学、外语教师发展等研究;3) 前沿研究:涉及栏目范围内的前沿动态、最新发展趋势及研究综述;4) 书刊评介:对近2年内的外文原版专著进行内容介绍和评价,突出新兴内容。

投稿方式:

- 1) 直接访问: <http://xdwy.cbpt.cnki.net> (中国知网CNKI采编系统)
- 2) 广东外语外贸大学主页(<http://www.gdufs.edu.cn>)—学科学术(正上)/科研单位(左下)—《现

代外语》

第一次投稿的作者完成网上注册（用户名、密码，长期有效）。注册成功后，作者可随时进行网上投稿，或查看审稿进程等。

特别提醒：

《现代外语》编辑部从未委托任何机构或个人组稿，也不收取任何费用。

通讯地址：510420 广州白云大道北2号 广东外语外贸大学《现代外语》编辑部

办公邮件：gplal@gdufs.edu.cn

办公电话：(020) 36207235