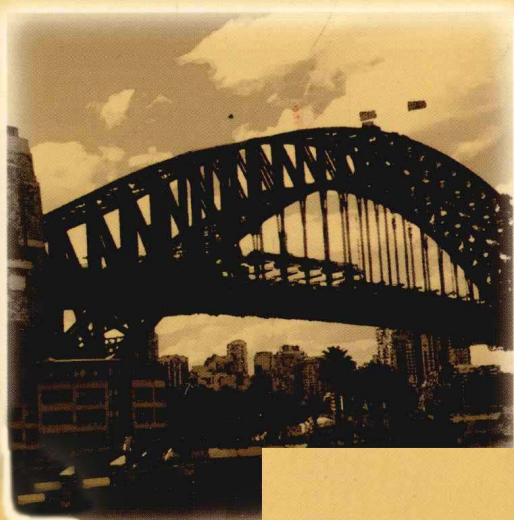


中国大学生英语口语分析性评估体系的构建与效验

Construction of and Validation Study on Analytic Rating Scales
for Oral English Proficiency of Chinese Tertiary Learners

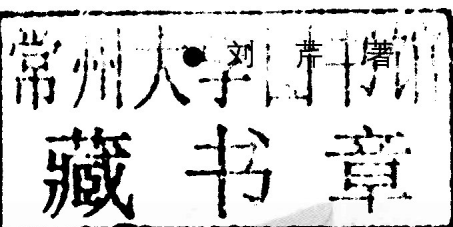
● 刘 芹 著



復旦大學出版社

中国大学生英语口语分析性评估体系的构建与效验

Construction of and Validation Study on Analytic Rating Scales
for Oral English Proficiency of Chinese Tertiary Learners



图书在版编目(CIP)数据

中国大学生英语口语分析性评估体系的构建与效验

刘芹著. —上海: 复旦大学出版社, 2010. 1

ISBN 978-7-309-07079-8

I. ①中… II. ①刘… III. ①英语-口语-教学研究
- 高等学校 IV. ①H319.9

中国版本图书馆 CIP 数据核字 (2010) 第 018147 号

中国大学生英语口语分析性评估体系的构建与效验 刘 芹 著

出版发行 复旦大学出版社 上海市国权路 579 号 邮编: 200433
86-21-65642857(门市零售)
86-21-65100562(团体订购) 86-21-65109143(外埠邮购)
fupnet@fudanpress.com http://www.fudanpress.com

责任编辑 郑梅侠

出品人 贺圣遂

印	刷	句容市排印厂
开	本	850 × 1168 1/32
印	张	10.25
字	数	252 千
版	次	2010 年 1 月第一版第一次印刷

书	号	ISBN 978-7-309-07079-8 / H · 1415
定	价	20.00 元

如有印装质量问题, 请向复旦大学出版社发行部调换。

版权所有 侵权必究

内容提要

本书作者以Bachman的交际语言能力（CLA）模式、Cohen的语用口试等级评分表、Nunn的小组讨论等级评分表及高校英语教学大纲为基础，结合大规模问卷调查结果构建了中国大学生口语评估体系。作者在将其运用于来自六类院校180名学生的口语评估后，对其进行了全面的效度研究，同时深入探讨了中国大学生英语口语的普遍特征。作者提出的口语评估体系具备科学性和可行性，对口语测试评分标准的改进和完善具有借鉴意义，对提高口语测试的总体效度有着很高的理论意义和应用价值。本书的读者对象为从事英语口语教学评估的研究人员和专业学生，也可供高校英语教师参考使用。

Preface

College teachers and students in China have always attached great importance to the cultivation of oral English proficiency. The teaching syllabuses of English majors and non-English majors, therefore, incorporate specific oral English proficiency requirements for students at each level. However, have the students met the requirements set by the syllabuses? Cai Jigang (2002), quoting from reports of CET Committee, points out that from January 1999 to May 2001, 18,550 students got a B in the six oral tests for non-English majors (CET-SET) out of a total number of 32,107, accounting for only 57.8%. He further argues that “Compared with 2 million test takers who take part in each CET written test during this period, this passing rate is extremely low, indicating poor oral English communicative ability among college students.” Meanwhile, Wen Qiufang, Wu Caixia and Lydia So (1999) conclude that “Second-year English majors have not definitely reached the requirements for oral English proficiency set by the teaching syllabus except for speaking speed” (34). Taking the results of TEM4-Oral 2000 as an example, Wen Qiufang, Zhao Xuexi and Wang Wenyu (2001) further argue that students have got such common problems as low accuracy, insufficient fluency, a lack of novelty and profundity, and unsatisfactory mastery of communicative rules while holding dialogues and discussions.

It can be concluded from the results of the above-mentioned two national oral English tests that the oral English proficiency of Chinese college students is far from satisfactory. Moreover, only students who have got excellent marks in the CET written test are eligible to take CET-SET and only a few thousand actually sit for TEM-Oral each year. A large number of students have got no experience of taking them. Thus they have got no means to evaluate their oral English proficiency, let alone comparing with students from other majors or universities in order to figure out their strengths and weaknesses. Therefore, it is a research project in urgent need to evaluate students' oral English proficiency in a complete and detailed way so that valuable information can be provided concerning oral English instruction in Chinese colleges and universities to help enhance oral English proficiency of students.

The study presented in this book aims at establishing a model of oral analytic rating scales (OARS) as an attempt to solve this problem. OARS is designed on the basis of Bachman's Communicative Language Ability (CLA) model, Cohen's Rating Scale for Pragmatic Speaking, Nunn's Rating Scales for Small Group Interaction, teaching syllabuses of English majors and non-English majors, as well as questionnaire survey results of 182 English teachers and 1,139 students. It covers two main aspects of oral proficiency — language competence and strategic competence. The former incorporates organizational competence (pronunciation, grammar and textual coherence, to be exact) and pragmatic competence (illocutionary as well as sociolinguistic competence) while the latter evaluates flexible interaction and nonverbal communication (namely body language and paralanguage). *Pronunciation* tests whether the candidate can

provide utterance with accurate pronunciation and intonation. *Grammar* assesses the candidate's capability of *producing* grammatically correct utterances (in terms of both vocabulary and syntax). *Textual coherence* assesses whether the candidate can use cohesive devices appropriately so as to form coherent utterances. *Illocutionary pragmatic competence* assesses whether the candidate can express ideas clearly, understand others' utterances and give timely and appropriate rejoinders. *Sociolinguistic pragmatic competence* examines whether the candidate's utterance is related to features of the language use context, such as appropriate uses of dialects, registers, and idiomatic expressions. *Flexible interaction* assesses whether the candidate has the ability to keep an interaction going smoothly. In doing so, he/she may have to exhibit the ability for taking turns (such as holding the floor, interrupting politely, and helping other participants to join in) and negotiating (such as taking initiatives and appreciating other participants' contributions). *Body language* includes eye contact, gestures and facial expressions while *paralanguage* refers to the vocalizations that are not words but that may convey meaning or add to the meaning of words (such as tone, volume and stress). Each of the eight traits (hereafter referred to as "Pron", "Gram", "Text", "Illoc", "Socio", "FlexI", "Body" and "Para" for convenience) is provided with a scale from 5 as the highest to 1 as the lowest, together with detailed band descriptors.

Language assessment is an endeavor closely linked to practicality. Problems do occur when a theoretically well-defined assessment model is put into practice. Thus, OARS has undergone the following two stages of validation study to prove its reliability, validity and utility.

The pilot study is mainly concerned with the feasibility of this

model, possible problems with regard to its application and its reliability and validity. It was conducted on 30 randomly selected sophomores majoring in English in University of Shanghai for Science & Technology (USST). They were randomly arranged into groups of three and each group was required to hold an 8-minute discussion. The picture recording of their performance was evaluated by two raters who did detailed ratings according to the band descriptors of OARS and six teachers who did rough ratings according to its framework. Average scores of the first kind of rating were taken as the core component of the study. The candidates' scores on TEM4 and oral English final examination, taken at roughly the same time of the experimental test, as well as their scores on TEM8 taken two years later, served as outside measures. Results were gained from tests of reliability, normality and correlation, shown as follows.

Firstly, OARS exhibits an overall reliability coefficient alpha of 0.7902, quite desirable for oral proficiency assessment. Secondly, most sub-total correlation coefficients are above 0.7 while most sub-sub correlation coefficients are below 0.6, indicating satisfactory construct validity. Thirdly, the correlation coefficient of the detailed ratings and the rough ratings reach 0.831 whereas that of the detailed ratings and the scores on the oral English final exam reach 0.653, signifying acceptable external validity. In addition, both the candidates and the teachers who conducted ratings expressed their approval of OARS for it can evaluate the oral English proficiency of Chinese college students in a detailed and comprehensive way. They also found it easy to follow the scales and band descriptors of OARS. So qualitatively and quantitatively, OARS establishes itself as a model of reliable, valid and feasible oral English rating scales.

The subsequent field study mainly focuses on the further applicability of OARS, its reliability and validity when applied in large-scale oral English assessment, and characteristics of Chinese college students with regard to their oral English proficiency. 180 candidates who have just passed TEM4 or CET6 took part in this study. They were selected from six levels of college students, namely English majors of key universities, Arts majors of key universities, Science majors of key universities, English majors of local universities, Arts majors of local universities, and Science majors of local universities. As in the pilot study, they were randomly grouped into three and required to hold an 8-minute discussion. All their performance was picture recorded and rated by two raters by means of OARS. Results display satisfactory reliability coefficient alphas (0.7900, 0.7885, 0.7693, 0.7902, 0.7853 and 0.7754 respectively) as well as construct validity (with most of the sub-total correlation coefficients reaching the theoretical demand of 0.7). It is concluded that OARS can be applied to different kinds of Chinese college students in the assessment of oral proficiency with acceptable reliability and validity.

When applied to 180 candidates as an entire sample, OARS exhibits even more satisfactory reliability coefficient and construct validity, indicating its higher reliability and validity while applied to large number of candidates. Common problems of these candidates concerning oral English proficiency are reflected in the following. The first is that candidates have not acquired advanced pronunciation skills. The trait "Pron" in OARS is meant to assess accurate pronunciation of vowels and consonants as well as appropriate pronunciation techniques and intonation. Results show that a large number of candidates have reached the basic

requirement of accurate pronunciation of phonemes, but their pronunciation and intonation have remained plain. To be specific, they have not grasped the advanced pronunciation techniques such as assimilation, liaison, stress, loss of plosion and incomplete plosion, let alone native-like pronunciation and intonation. The second is that they have not grasped communication rules. Results of the assessment from the traits of “Illoc” and “Socio” show that about two thirds of the candidates can express their ideas clearly and give timely and reasonable response to others’ utterance. In addition, their utterance is appropriate to the actual language use situation. However, the results are not so satisfactory when it comes to flexible interaction. Nearly half of the candidates lack communication means. They cannot find appropriate time to join in a discussion. Some of them interrupt others whenever they like, some keep on talking when they get a turn, and some always wait for others to give them a turn. The third is that candidates generally lack nonverbal communication awareness. A great number of linguists and experts on language teaching and assessment keep arguing that nonverbal communication is indispensable to communicative proficiency (such as Harrison 1965, Morlan & Tuttle 1976, Canale 1983, Canale & Swain 1980, Ross 1986, Nunn 2000, Taylor 2006, etc.). It is a common feature of everyday communication among native speakers. However, about half of the candidates who underwent this assessment procedure do not show this ability. While presenting ideas, they display flat tones. While discussing with others, they do not make direct eye contact. Their facial expression is dumb. They are unable to use gestures. Even though some of them use some gestures, they are not natural. The above three problems are common among Chinese college students,

which need immediate attention in oral English instruction.

Further *t* tests and MANOVAs of comparability studies show the following findings. First, key university students acquired a statistically significant higher mean score than local university students on such traits as “Gram”, “Text”, “Illoc”, “FlexI”, “Body” and “Para”. Second, English majors attained a statistically significant higher mean score than non-English majors on all traits while there was no difference between Arts majors and Science majors. Third, among key university students, English majors were only better than non-English majors on “Gram” while among local university students, English majors were better than non-English majors on such traits as “Text”, “FlexI”, “Body” and “Para”.

The present book takes the author’s doctoral dissertation as a basis. While working as a visiting scholar at the University of California, Los Angeles (UCLA), she rewrote several chapters. After returning to China, she made further modifications on the entire manuscript according to recent language assessment theories and results of empirical studies. This book, so to speak, summarizes her research on oral assessment during the past years.

During the process of writing, the author has received guidance and help from a number of professors and fellow researchers. First and foremost, her heartfelt gratitude naturally goes to Professor Zou Shen, her supervisor at Shanghai International Studies University (SISU), who has inspired her into carrying out the project and given her heuristic instructions on research design and manuscript writing, which helped lay a solid foundation for the current book. Professor Wen Qiufang at Beijing Foreign Studies University spared valuable time discussing with the author the rationale of designing the model of oral analytic rating

scales. Professor Lyle F. Bachman at University of California, Los Angeles (UCLA) helped the author modify the statistical analysis of the collected data during her stay there as a visiting scholar. Professor Jin Yan at Shanghai Jiaotong University and Professor Zeng Yongqiang at Guangdong University of Foreign Studies provided the author with valuable suggestions on the final draft. Professors and the author's fellow Ph. D candidates at SISU always lent her a hand without hesitation during her three years' doctoral study there. Fudan University Press provided enthusiastic support for the publication of this book. All the above deserve the author's sincere gratitude.

Deep appreciation must go to the several hundred teachers and more than one thousand students from SISU, Fudan University, Shanghai University of Finance & Economics, East China University of Science & Technology, Tongji University, Shanghai Normal University and University of Shanghai for Science & Technology, who endured the trouble of questionnaire survey, experimental tests and ratings.

Finally, the author is much indebted to her family for their constant support, encouragement, and endurance throughout these years.

Though the present book has undergone several modifications, it is far from perfect. Any criticism or suggestion is welcome.

Liu Qin
Shanghai
October, 2009

List of Acronyms

ACTFL	the American Council on the Teaching of Foreign Languages
BEC	Cambridge Business English Certificate
CET	College English Test
CET4	Band 4 of CET
CET6	Band 6 of CET
CET-SET	Spoken English Test of CET
CLA	Communicative Language Ability
ECUST	East China University of Science and Technology
FSI	the American Foreign Service Institute
IELTS	International English Language Testing System
IELTS Oral	Oral Test of IELTS
ISLPR	International Second Language Proficiency Ratings
O1	rating given by the first group member other than the candidate
O1T	total score of O1
O2	rating given by the second group member other than the candidate
O2T	total score of O2
OA	average of O1 and O2
OAT	total score of OA

OARS	Model of Oral Analytic Rating Scales
R1	detailed rating given by the first rater
R11A	R1 on trait 1A
R11B	R1 on trait 1B
R11C	R1 on trait 1C
R12A	R1 on trait 2A
R12B	R1 on trait 2B
R13	R1 on trait 3
R14A	R1 on trait 4A
R14B	R1 on trait 4B
R1T	total score of R1
R2	detailed rating given by the second rater
R2T	total score of R2
RA	average of R1 and R2
RA1A	RA on trait 1A
RA1B	RA on trait 1B
RA1C	RA on trait 1C
RA2A	RA on trait 2A
RA2B	RA on trait 2B
RA3	RA on trait 3
RA4A	RA on trait 4A
RA4B	RA on trait 4B
RAT	total score of RA
S	candidate's self rating
ST	total score of S
SHUFE	Shanghai University of Finance and Economics
SISU	Shanghai International Studies University
SNU	Shanghai Normal University
SPSS	Statistical Package of Social Sciences

T1	rough rating given by the first teacher
T1T	total score of T1
T2	rough rating given by the second teacher
T2T	total score of T2
T3	rough rating given by the third teacher
T3T	total score of T3
T4	rough rating given by the fourth teacher
T4T	total score of T4
T5	rough rating given by the fifth teacher
T5T	total score of T5
T6	rough rating given by the sixth teacher
T6T	total score of T6
TA	average of T1, T2, T3, T4, T5 and T6
TAT	total score of TA
TEEP	the Test in English for Educational Purposes
TEM	Test for English Majors
TEM4	Band 4 of TEM
TEM4-Oral	Oral Test of TEM4
TEM8	Band 8 of TEM
TEM8-Oral	Oral Test of TEM8
TOEFL	Test of English as a Foreign Language
TOEFL AST	Academic Speaking Test of TOEFL
TSE	Test of Spoken English
USST	University of Shanghai for Science and Technology

序

口语能力是语言能力的重要组成部分,其测评历来是语言测试研究者和语言教学工作者所关注及研究的重点之一。口语测评大多是一种基于被试者表现的考试形式(performance-based testing),即被试者根据考试要求,完成一个或数个口语任务,然后由考官根据考生表现决定其口语能力等级。在这个过程中有几个因素会影响考生表现:首先是他的口语能力,其次是口试形式、评分标准以及考官个人因素。其中一个基础且关键的环节是评分标准。因而,如何制定一个能客观、公正、全面地衡量口语能力的评分标准是口语测试人员迫切需要解决的一个课题。

本书在口试评分标准制定方面做了卓有成效的探索。作者以分析性评分模式为切入点,以多种理论模式为依据,以现行英语教学大纲为基准,在大规模调研的基础上,提出了分析性口语评估体系(OARS)。

本书具有以下特点。文献部分的综述全面、详尽,涵盖了本领域的最新研究成果。作者从不同角度阐述了相关理论模式,包括语言能力模式及口语评分理论依据。阐述条理清晰、概括性强、层次分明,显示了作者扎实的理论功底。口语评估体系的第一个特点是测量维度全面:既包括语言能力又涵盖策略能力。同时,语言能力的衡量能兼顾到口语能力的外显特点(如语音语调等)。策略能力的判断也基于口语交际的特征之上,如包括体势语、声调、音量等。因此,该评估体系能较为客观、公正地评价口语能力。口语评估体系的另一特点是兼顾英语教学要求,每个小项都有相应的等级描述。这有利于该体系的使用者(教师、学生)充分利用口语测试的信息反馈,促进教与学。另外,研究设计全面,研究方法科学,对口语评估体系的效度验证

过程严谨,结论合理,论证结果令人信服,显示了作者良好的科研素质和能力。

总而言之,本书是一项高质量的研究成果。作者提出的口语评估体系具备科学性和可行性,对口语测试评分标准的改进和完善具有借鉴意义,对提高口语测试的总体效度有着很高的理论价值和应用价值。

邹申

上海外国语大学教授、博导
教育部高等学校外语专业教学指导委员会委员
英语专业教学分指导委员会副主任委员

2009年11月