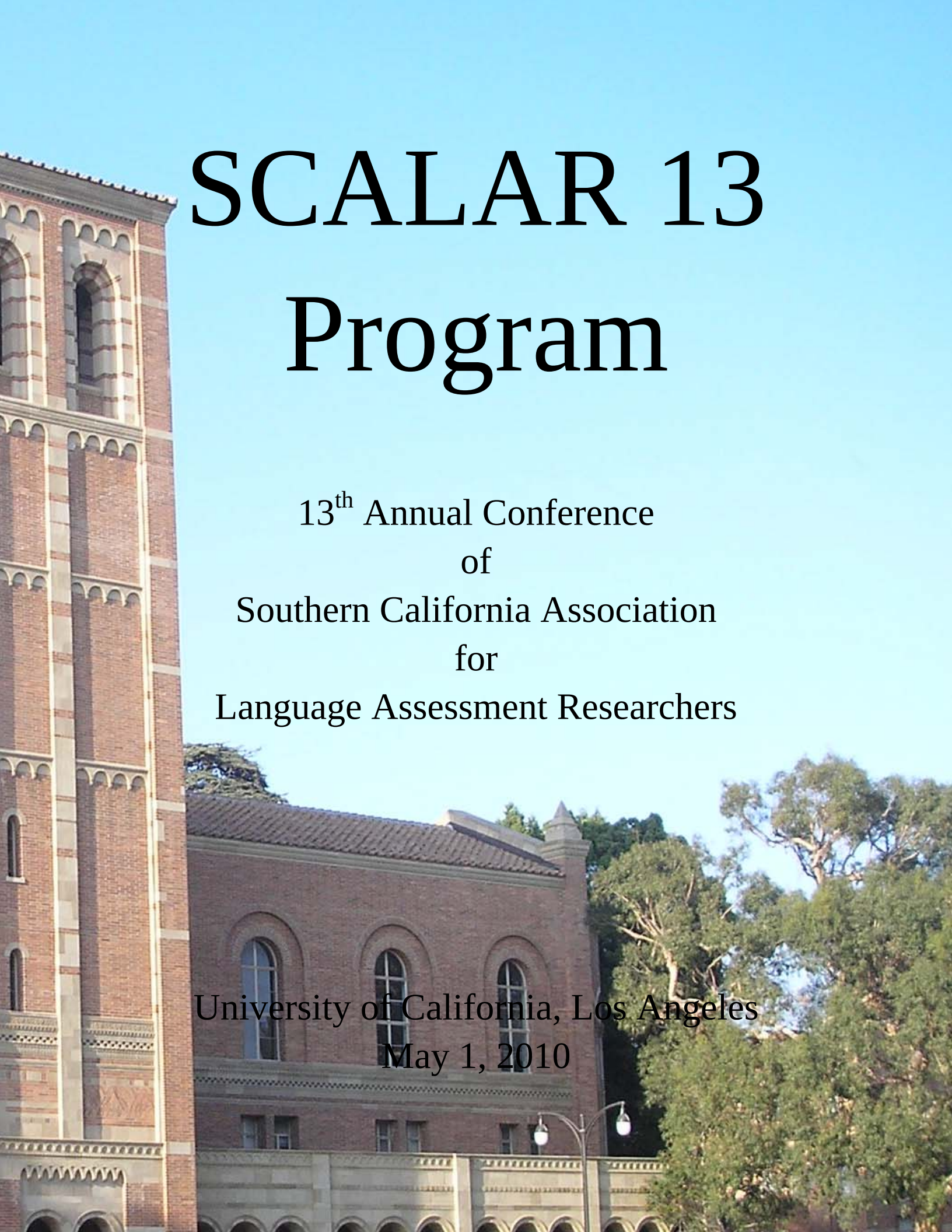
SCALAR 13 Program

13th Annual Conference
of
Southern California Association
for
Language Assessment Researchers



University of California, Los Angeles
May 1, 2010

Faculty Advisors:

Lyle Bachman	University of California, Los Angeles
Antony Kunnan	California State University, Los Angeles
Nathan Carr	California State University, Fullerton

Guest Speaker:

Priya Abeywickrama	San Francisco State University

SCALAR Committee:

Hongwen Cai	University of California, Los Angeles
Ikkyu Choi	University of California, Los Angeles
Jonathan Schmidgall	University of California, Los Angeles

Funded by

UCLA gsa GRADUATE STUDENTS ASSOCIATION

Program

Preconference workshop

Speaker:	Priyanvada Abeywickrama
Topic:	Formative Assessment
Venue:	2112 Rolfe Hall
Time:	2:00-5:00pm
Date:	April 30, 2010

Morning session

Venue:	2408 Ackerman Union
Date:	May 1, 2010
09:00-10:00	Keynote address
	Priyanvada Abeywickrama
	Classroom based Assessment: Principles and Practices
10:00-10:30	Hongwen Cai
	An AUA View on Diagnostic Language Assessment
10:30-10:40	BREAK
10:40-11:10	Henry Nguyen Pham
	Negative Feedback (or Repair) in Learner-Learner EFL Interaction
11:10-11:40	Jonathan Schmidgall
	Developing and Evaluating an Annotation Scheme for Grammatical Errors in Learner Speech
11:40-12:10	Jinghua Wang
	A Pilot Study of Classroom Assessment for EFL Writing in China
12:10-1:20	LUNCH BREAK

Afternoon session

Venue: 2408 Ackerman Union

Date: May 1, 2010

1:20-1:50	Jayne Garcia The Method Used to Test U.S. Naturalization Applicants: Is It Violative of the 14th Amendment of the U.S. Constitution?
1:50-2:20	Kyunghee Yoo Korean's Hierarchical Nationhood: Center for Two Groups: Korean Chinese and Korean American
2:20-2:50	Hsin-min Liu Building a Construct Validity Argument for the GEPT High-Intermediate Reading Test: A Confirmatory Factor Analysis Approach
2:50-3:20	Giang Hoang The National University Entrance English Test in Vietnam: At Risk of Failing the Validation Test
3:20-3:30	BREAK
3:30-4:00	Ikkyu Choi Effects of Item Content Characteristics on Item Difficulty of Multiple Choice Test Items in an EFL Listening Assessment
4:00-4:30	Amparo Pedroza, Debra Thiercof, Winnie Chen, Michelle Luster Pilot Test Report for Academic English: Intermediate Grammar and Writing
4:30-5:00	Yujie Jia Using CFA Approach to Investigate the Construct Validity of the Analytic Rating Scales in a Semi-direct Oral Test
5:00-5:30	Panel discussion Lyle Bachman, Antony Kunnan, Nathan Carr, Priya Abeywickrama

Abstracts

Classroom Based Assessment: Principles and Practices

Priyanvada Abeywickrama, San Francisco State University, abeywick@sfsu.edu

The terms classroom based assessment, teacher based assessment, school based assessment, formative assessment and alternative assessment are all used interchangeably to refer to the same practices and procedures. They highlight different aspects of the assessment process but all signify a more teacher-mediated, context based, learning based assessment practice in contrast to large scale formal assessments that are used for accountability. This presentation will discuss principles of classroom-based assessment and also examine conceptual issues and challenges in relation to classroom based assessment practices.

Formative Assessment

Priyanvada Abeywickrama, San Francisco State University, abeywick@sfsu.edu

Assessment usually conjures up images of finals and other end of unit tests. This type of assessment is known as SUMMATIVE. However, when assessments are daily opportunities to collect information about student work, to assess what/how they understand, and gives evidence of learning in process, we call them FORMATIVE. During this workshop participants will learn about the pedagogical implications of using formative assessments and will have the opportunity for incorporating formative assessments in their own courses.

An AUA View on Diagnostic Language Assessment

Hongwen Cai, University of California, Los Angeles, hwcai@ucla.edu

Diagnostic language assessment has attracted growing attention in recent years, accompanied by mixed understanding of what distinguishes it from other purposes of language assessment. The Assessment Use Argument framework (Bachman 2005, Bachman & Palmer 2010) provides a systematic approach to its defining characteristics.

Based on the AUA framework, this paper argues that relevance to decisions on adaptive and remedial language teaching and learning is the primary quality of diagnostic language assessment, while sufficiency of information for such decisions is the secondary quality. The dominant importance of these two qualities in understanding diagnostic language assessment is illustrated through the review of three diagnostic assessments on EFL listening comprehension.

(Hongwen Cai is a second year PhD student in language assessment at UCLA.)

Negative Feedback (or Repair) in Learner-Learner EFL Interaction

Henry Nguyen Pham, California State University, Los Angeles, ucsdhenry@gmail.com

Researchers and teachers agree that negative feedback (or repair) enhances second language acquisition (SLA), but language assessment studies of English learners have not focused on such a role during interaction as well as the relationship between language proficiency and choice and frequency of feedback. The present study explores learner-learner interaction in an authentic English as a foreign language (EFL) setting to assess the nature of corrective feedback that results from other- and self-repair. The investigation analyzed informal conversations between two Vietnamese women, both non-native English speakers (NNES). The paper includes 1) a qualitative analysis using the conversation analysis (CA) transcription system to examine repair, repair uptake, and subsequent modified output; and 2) tabulation of the frequency of errors and the feedback types that follow. The first major finding reveals that more than half of all error sequences were ignored or not noticed. Unnoticed errors were chiefly grammatical, followed by phonological errors, with very few lexical errors. Unnoticed errors were produced much more frequently by the less proficient speaker. Second, feedback occurred mainly for lexical errors, some grammatical errors, and very few phonological errors. The more proficient speaker noticed more errors and ultimately corrected herself and her partner nearly twice as often as did her partner, who focused primarily on self-correction and asking for help. Finally, results show that feedback, when noticed, favored self-correction and negotiation of form (respectively); the remaining types of feedback (recast, elicitation, explicit correction, asking for help) occurred at much lower frequencies. Implications are discussed for the clear preference for self-correction and negotiation of form as choice of feedback, the need to assess learner proficiency based on the choice and frequency of feedback, and whether learners should be trained to

notice and self- and other-regulate errors even when meaning is unimpaired and intelligibility is unimpeded.

(Henry Nguyen Pham is an MA TESOL student at CSULA with EFL teaching experience. His current research interests are in university entrance exams, repair, and English as an international language.)

Developing and Evaluating an Annotation Scheme for Grammatical Errors in Learner Speech

Jonathan Schmidgall, University of California, Los Angeles, JSchmidgall@ucla.edu

The process of analyzing grammatical errors in speech is fraught with complexity and difficult to implement in a systematic way. Several annotation schemes have been proposed and implemented for analyzing grammatical errors in writing, but very little has been done on a large scale for learner speech. This paper presentation will discuss the development and implementation of an annotation scheme for grammatical errors using transcripts of learner speech, and approaches to evaluate the consistency with which the scheme was applied.

The annotation system was designed based on a review of existing annotation schemes and general principles advocated by experts in the field. It was adapted in order to minimize the complexity of the system so as to facilitate a large number of annotations by a group of raters, as well as the perceived impact of errors on the comprehensibility of speech.

Speech transcripts were obtained for 1521 responses to a workplace English speaking test. All transcripts were double-annotated using a group of 24 annotators, who were trained and given a detailed annotation manual. Any discrepancies between annotations between annotator pairs were resolved by a third annotator during a subsequent adjudication phase.

In addition to presenting the annotation scheme and relevant context, methods for evaluating the consistency with which it was applied will be explored. Since annotation required (1) identifying a string of text in the annotation tool, (2) assigning an error category to the string, and (3) producing a correction, annotators could apply the scheme inconsistently in a number of ways.

In addition to quantitative approaches to evaluating consistency, results from annotator surveys will be discussed to collaborate results and critique relevant aspects of the annotation scheme. Recommendations for possible revisions to the scheme will be discussed.

(Jonathan Schmidgall is pursuing a PhD in Applied Linguistics at UCLA, with a broad interest in test validation research for performance assessments, and more specifically in validity issues and impact emerging from the use of automated scoring in language assessment.)

A Pilot Study of Classroom Assessment for EFL Writing in China

Jinghua Wang, Hebei University, wangjh2808@sina.com

The pilot study is mainly focused on applying formative assessment theory of laying particular emphasis on process-oriented learning to enhancing students' motivation, lowering their affective filter in the process of writing as well as increasing the efficiency in teachers' correction of students' compositions. By adopting the formative assessment, we convert the product-oriented writing into process-oriented writing by taking the following three steps: self-revising, peer-editing and teacher-evaluating. First, a form of self-assessment is designed for students to make their self- and peer-assessment possible. Students revise their drafts, make comments on their own writing based on the content of self-assessment form. Then, each individual student should find one of his/her classmates to help edit his/her composition, which is called peer-editing/assessing. After that, students need to hand in to their teachers their first drafts, second drafts, and final texts. Teachers not only check the weaknesses of their writing, but more important, evaluate students' efforts and progress, studying attitude according to their self-assessment and peer-assessment. About 200 first-year non-English-major undergraduates in five natural classes have been involved in this research program for a semester.

After the empirical experiment, a survey/questionnaire and a face-to-face interview are conducted, and the findings/outcomes of the investigation (will) show formative assessment can bring into play the initiative of the students, enhance their learning autonomy, foster their sense of responsibility, and cooperative spirit as well. Most

students advocate the process-oriented writing through formative assessment (self-assessment, peer-assessment and teachers' evaluation).

There are some limitations in the pilot experiment, like, the effectiveness of self-asseessing and efficiency of peer-editing, etc. We would also diversify the peer-editing, like, assessing the compositions sometimes by the student chosen by themselves, sometimes in pairs or in small groups.

(Jinghua Wang is a professor and vice dean of the Foreign Language Teaching and Research College for Non-English Major Students at Hebei University in China. Her academic interests include classroom assessment, TESOL, and faculty education. Prof. Wang is currently a Visiting Scholar in the Department of Applied Linguistics, UCLA)

The Method Used to Test U.S. Naturalization Applicants: Is It Violative of the 14th Amendment of the U.S. Constitution?

Jayne Garcia, California State University, Los Angeles, sergionjayne@yahoo.com

The Immigration and Naturalization Act of 1952 requires that applicants for naturalized citizenship have to demonstrate English language ability and knowledge of U.S. history and government. This requirement since the late 1980s is enforced through an interview test administered by an immigration officer. Many concerns have been raised regarding the Naturalization Test (both old and redesigned test in 2007) in terms of validity, reliability and fairness (see Kunnan, 2009a, 2009b, Extra and Spotti, 2009).

In this paper, I would like to argue that there could be a case for challenging the test on legal grounds as well. First, there is no legal authority that requires the administering of the language test; second, USCIS officers are given too much discretion in their pass/fail determinations; third, applicants have no means to appeal the immigration officer's decision. After examining the reasoning applied in U.S. Supreme Court cases that address English language requirements for purposes other than citizenship, as well as Appellate Court case law concerning discretion assigned to immigration officers, I conclude that the methods used by the U.S. Citizenship and Immigration Services to administer the Naturalization Test are violative of the equal rights protection of the 14th Amendment of the U.S. Constitution.

(Jayne Garcia graduated magna cum laude from the University of Nebraska, Omaha with a BA in Spanish. She is currently in the MA TESOL program at CalStateLA.)

Korean's Hierarchical Nationhood: Center for Two Groups: Korean Chinese and Korean American

Kyunghee Yoo, California State University, Los Angeles, honeybunchstella@gmail.com

This paper discusses critically Korean's hierarchical nationhood about two Korean ethnic return migration groups, Korean Chinese (called by Joseonjok) and Korean American (called by Gyopo) by analyzing the South Korean policy and social attitudes regarding two ethnic migrants. By the legal states, the Korean government defines Korean Chinese as foreigners, though, allowing them preferring legal status over other foreigners. Therefore, they are excluded from many benefits which Korean American have as the return ethnic Korean. This hierarchical legal dimension has been managed by Korean's economic and geopolitical interests not by the ethnic identity of culture, language, and the same origin. Social dimension of hierarchical nationhood is shown by public opinion data and Korean media towards Korean Chinese. The data on reported Korean Chinese considered as 'other', in adversely, Korean American is 'our' Korean. The Korean media have more promoted their negative image by portraying them as 'other', second class immigrants.

(Kyunghee Yoo is studying the TESOL master course at California State University, LA. She has interests in language testing, citizenship test and course development for EFL context.)

Building a Construct Validity Argument for the GEPT High-Intermediate Reading Test: A Confirmatory Factor Analysis Approach

Hsinmin Liu, University of California, Los Angeles, hmliu@ucla.edu

This study built and supported a theory-based construct validity argument for the General English Proficiency Test (GEPT) high-intermediate reading test based on Bachman and Palmer's (in press) 'Assessment Use Argument' validation framework. It examined the randomly sampled data from the target population using confirmatory factor analysis with item-level responses. In order to collect all relevant evidence to support the warrant that score interpretations were meaningful with respect to a general

theory of reading ability, two factor models, the higher-order factor model and the bi-factor model, were selected to represent the underlying factor structure of the test. This was due to the constraints of the test design that either model was not sufficient enough to describe the phenomena as a whole. While some possible item and passage effects were detected, the findings showed that to a large extent the underlying factor structure identified was in accordance with the test construct outlined in the test specification. Thus, the findings of the study weakened the claim that the test lacked theory specification for the construct to be measured (Pan and Roever, 2008).

(Hsinmin Liu is a PhD student in Applied Linguistics at UCLA. Her research interests include test validation and alignment issues related to real world domains.)

The National University Entrance English Test in Vietnam: At Risk of Failing the Validation Test

Giang Hoang, California State University, Los Angeles, lgiang8380@gmail.com

In keeping with recent reforms in the Vietnamese educational system, a more communicative approach to instruction can be seen through the new English textbooks used in secondary level and university EFL classrooms. Amidst these changes, the National University Entrance English Test (NUEET) designed for high-stakes nationwide matriculation is at risk of failing decision-making validation. This study is an investigation into NUEET decision-making validity through an in-depth content analysis of the four NUEET versions over the last four years (2006 to 2009). The main focus is on how linguistic knowledge and skills are treated and proportionally weighted in the test across the four versions as a way of collecting initial evidence about NUEET decision-making validity. The findings show that despite NUEET efforts to depart from the traditional testing focus on grammatical and lexical memorization, its indirect test items of productive skills and inflexible treatment of linguistic knowledge through discrete-point recognition items may not provide adequate information for error-free decision-making. One area is the lack of authenticity and relevance to real life target language use. In other words, this test neither sufficiently reflects the current high-school EFL classroom practice in Vietnam nor fulfils its predictive role of providing evidence for selecting the most suitable candidates for tertiary education. This suggests that there is an urgent need for improvement to the test so that it can synchronize with other reforms in the educational system and consequently becomes a trigger mechanism for other changes on the way.

(Giang Hoang is doing an MA TESOL program at CSULA. She is interested in doing research on assessing EFL writing through the use of alternative assessment tools.)

Effects of Item Content Characteristics on Item Difficulty of Multiple Choice Test Items in an EFL Listening Assessment

Ikkyu Choi, University of California, Los Angeles, ikchoi@ucla.edu

This study aimed to provide a systematic way of examining the difficulty of multiple choice English listening comprehension (ELC) test items. In particular, this study investigated (1) content characteristics of a set of multiple choice ELC test items and (2) relationship between the ELC item content characteristics and their difficulty. In order to address the two issues, 120 multiple choice ELC items from preparation examinations for the Korean College Scholastic Ability Test (CSAT) were selected as objects of analysis. Tests consisting of the selected ELC items were administered to a total of 1,280 secondary school students to estimate difficulty of each ELC item. Furthermore, two content raters evaluated each of the items in terms of 27 item content characteristic variables. Covariance structure analysis was then conducted to evaluate a hypothesized relationship between item content characteristics and item difficulty. The item contents analysis revealed that the ELC items investigated in this study contained few difficult words, seldom varied in terms of their stems and options, and demonstrated two different indicators for overlap between recorded stimulus and its options, namely, surface overlap counting and a judged rating of the overlap. Subsequent covariance structure analysis identified a model that accounted for item difficulty with two latent variables, stimulus complexity and item/stimulus overlap. The results of this study had a couple of major implications. First, the use of difficult words is recommended in controlling the difficulty of the CSAT-type ELC items. In addition, when an overlap between stimulus and its options is employed to manipulate difficulty of the ELC items, it is essential to measure judged degree of the overlap made by test takers during the test administration.

(Ikkyu Choi is a doctoral student in language assessment at University of California, Los Angeles.)

Pilot Test Report for Academic English: Intermediate Grammar and Writing

Amparo Pedroza (apedroza1@csu.fullerton.edu), Debra Thiercof (thiercof@cox.net), Winnie Chen, Michelle Luster, California State University, Fullerton

English for Academic Purposes (EAP) programs are faced with the challenge of academically preparing students from a large variety of educational and cultural backgrounds for admission into American colleges and universities. It is in this context that we undertook a pilot test project t to assess students in intermediate composition at mid-semester in the American Language Program at California State University, Fullerton. This was a criterion-referenced test as it assessed individual language ability in absolute terms, and it tested both achievement and progress at mid-semester in two constructs: grammar and composition. The target language domain consisted of academic environments in the ESL class and includes academic writing in a variety of rhetorical modes and related grammar forms. The test was administered to 31 students whose L1 backgrounds include Chinese, Korean, Japanese, and Arabic. The test was organized into two sections. For grammatical forms, students were tested on their knowledge and usage of conjunctive adverbials, sentence combining using coordinating conjunctions, transitions in a paragraph, subject-verb agreement, capitalization, and pronouns. The writing section tested students on organization, cohesion, grammar/mechanics, and content/coherence in paragraph writing. While the overall testing process was generally successful, there are several recommendations for improvement that we would consider for designing future tests.

(Winnie Chen is a recent graduate from the TESOL program at California State University, Fullerton. She is currently teaching English as a Foreign Language in Taiwan.

Michelle Luster is currently a student in the TESOL program at California State University, Fullerton. She is currently teaching at the American Language Program at Cal State Fullerton.

Amparo Pedroza is currently a student in the TESOL program at California State University, Fullerton. Her future goals include teaching in an Intensive English Program.

Debra Thiercof is a recent graduate from the TESOL program at California State University, Fullerton. She is currently working with LEP learners at the high school level.)

Using CFA Approach to Investigate the Construct Validity of the Analytic Rating Scales in a Semi-direct Oral Test

Yujie Jia, University of California, Los Angeles, yujiejia@ucla.edu

This study mainly addressed the convergent/discriminant validity of analytic rating scales used in a semi-direct English oral test. Confirmatory factor analysis (CFA) was adopted to analyze the responses of 551 students in a Hong Kong university to five different speaking tasks in the oral test. Each CFA model included five factors associated with the five analytic rating scales (Task fulfillment and relevance, Clarity of presentation, Grammar and vocabulary, Pronunciation, and Confidence and fluency) and five factors related to the five speaking tasks focusing on different dimensions. A series of models that depicted different relationships among these latent factors were tested to primarily examine whether the underlying multicomponential trait factor structure assumed in this English oral test could be supported. Although the satisfactory fit of the Correlated Trait Factor Model partially supported the convergent/discriminant validity of the rating scales, the Higher-order Trait Factor model was more interpretable because it does not only explain the relationships among the five rating scales but also the overall construct underlying the five dimensions and represented by the composite score: English speaking ability. It is hoped that this study can provide some insights on the validation issues surrounding speaking performance assessments especially the discriminant/convergent validity of analytic rating scales.

(Yujie Jia is a first-year PhD student in the Department of Applied Linguistics at UCLA. Her research interests include performance assessments, reading tests, test takers' strategy use, and classroom assessments.)