PROFILING SECOND LANGUAGE SPEAKING DEVELOPMENT:

THE INTERFACE BETWEEN LANGUAGE LEARNING, TEACHING AND

ASSESSMENT

MING-FEN LO

Abstract

Despite the popularity received in the testing market, an influential test as OPI (the Oral Proficiency Interview), has experienced lots of skepticism on the validity of its band descriptors defined by ACTFL (the American Council on the Teaching of Foreign Language Proficiency Guidelines). Lantolf and Frawley (1985) questioned the validity of the Guidelines by looking into its analytic logic. Brindley (1998) argued against the claim of the Guidelines as describing language development for its lack of empirical evidence. This paper, firstly, reviews the studies on the validity of the rating scales and the argument in Brindley's (1998) article—Describing language development? Secondly, the oral observation procedure devised by Pienemann, Johnston and Brindley (1988) and the Rapid Profile program based on the findings of the ZISA research group will be presented to exemplify the interface between SLA research, language assessment and language teaching. Suggestions for further research in this area will also be discussed.

Key words: *validity, rating scales, multidimensional model, profile analyses, teachability*

Introduction

To measure language ability effectively, we have to be able to find out what the ability is and how it has developed. In the field of language assessment, the knowledge of the language and the skills possessed by the test candidates is termed test construct. Defining the construct has been the central concern of language testing research. Different views of language and language use form the theoretical basis of different tests. In the 1960s, discrete point testing which was mostly formatted in multiple choice question type focuses on learners' grammatical or vocabulary knowledge. Oller's unitary competence hypothesis brought the cloze tests into popularity in the 1970s and early 1980s. Until now, communicative language tests derived from Hyme's theory of communicative competence play a dominating role in today's test markets (McNamara 2000). Another important contribution to the development of language testing comes from the findings of second language acquisition research. From the perspective of language teaching, the ultimate purpose of language assessment is to help learners make progress. As noted by Alderson (2006), "if researchers, theoreticians and testers do not know how language proficiency develops, they can hardly claim to be able to help learners develop such an ability." Therefore, this paper, firstly, intends to examine the ACTFL (the American Council on the Teaching of Foreign Language Proficiency Guidelines) to see if this influential guidelines which form the basis for two speaking tests, OPI (Oral Proficiency Interview) and TOPT (Texas Oral Proficiency Test) correspond to the realities of language acquisition and use. This part of discussion is based on Brindley's (1998) article—Describing language development?

The second part of this paper attempts to provide a response to Clark's claim of foreign and second-language proficiency testing as an "inexact science" for the lack

of an objective testing procedure guided by a uniform theoretical structure. (Lantolf 1985, p. 338)   The oral observation procedure devised by Pienemann, Johnston and Brindley (1988) and the Rapid Profile program will be presented to exemplify the interface between SLA research, language assessment and language teaching. Suggestions for further research in this area will also be identified.

Language testing

Language acquisition   Language teaching
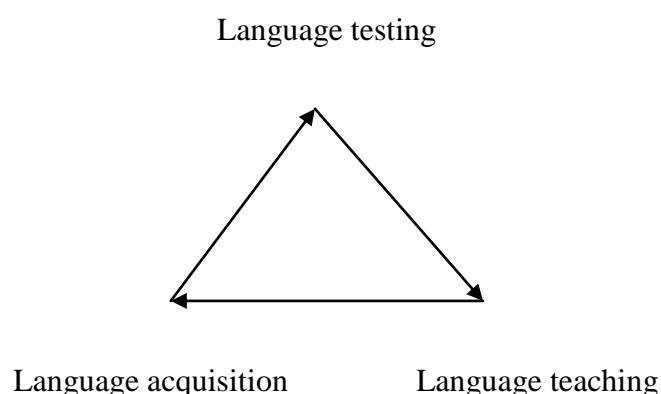
I.   SLA and the rating scales

Lantolf and Frawley's (1985) comparison of a criterion-referenced test to a driving test reminds me of my driving test experience in Taiwan.   By performing some required tasks such as backing, parallel parking, signaling and turning at the test center which simulates the real road situations, with some expected mistakes, I was granted a driver's license.   What is the meaning carried by the tiny piece of paper for which I was trained for a month at a private driving school?   Does it mean that I have mastered the skills required for passing a driving test at the virtual center, achieved the learning objectives for the training courses, acted as a qualified driver (I have acquired the driving competence.) or just simply obtained the permission to drive in my country?   The situation is somewhat like a performance test which claims to assess L2 learners' oral proficiency or communicative competence in a simulated real-life situation.   Can the oral proficiency tests in practice today measure what they claim to assess?   In other words, can the test results of learners'

performance in specific contexts infer the underlying linguistic competence they have acquired?　The following quote from Bachman and Palmer's frequently-cited paper might serve as part of an answer to this question, "measures of communicative performance must not be taken as an indication of some absolute amount of success an individual has in communicating." (Bachman and Palmer 1984)　Moreover, oral proficiency assessment is not as simple as a driving test which measures one person's performance with a pass or fail judgment.　Most performance tests, oral proficiency test such as the OPI (the Oral Proficiency Interview) for instance, adopt a behavioural rating scale for scoring.　Behavioural rating scales, in Brindley's words, "seek to define language ability globally in terms o features of 'real-life' performance and thus describe specific contexts of language use" (Brindley 1998, p. 112).

　　Despite the popularity received in the testing market, an influential test as OPI has experienced lots of skepticism on the validity of its level descriptors defined by ACTFL.　Lantolf and Frawley (1985) questioned the analytic logic of the guidelines by saying that "the reductive definitions of the Guidelines have self-defined truth: they define the truth of each level".　In other words, the Guidelines measure reality by definition because no empirical evidence for the gradation of linguistic criteria can be found.　Like most behavioural rating scales, the ACTFL scale prescribes what the test designers think learners should be able to do by using a native-speaker norm rather than describes what learners actually do.　Another objection to the use of the behavioural rating scale is its lack of diagnostic information.　One of the most important functions of language tests is to notify language teachers what strengths and weaknesses their learners bear.　Geoff Brindley (1998), from the perspective of SLA, argued against the claims made by Galloway (1987), Griffin and Nix (1991), Graham (1993) and Ingram (1984) that the level descriptions of rating scales are learner

norms-referenced and therefore describe language acquisitional or developmental processes.   In line with Lantolf and Frawley's view, Brindley argued that no studies in SLA were found to verify the above-mentioned claims.

By examining the ability levels expressed in terms of text types, task difficulty and skill hierarchy in rating scales, Brindley further questioned the relationship between the proficiency defined in the ACTFL Guidelines and what is tested in the OPI.   As Alderson (1991b cited in Brindley 1998) pointed out, "if descriptors are to be meaning characterizations of ability, then they should be able to be related to actual performance."   As mentioned before, the ACTFL Guidelines or any other holistic rating scales measure language proficiency by definition.   The method of defining the units of measurement seems to be the method we use to measure most of the things in the world.   We define the standard rule such as the metric system so that we can measure height and length.   However, mental constructs are not like physical characteristics which we can observe directly.   Is the proficiency defined in each level of the rating scales related to the performance in a test situation?   Are the intervals between the different levels equal?   Is there a hierarchy of text types and skills?   Does the performance on higher proficiency tasks subsume lower ones?   Moreover, there are problems of achieving inter-rater reliability since different raters may apply the descriptors of each level or perceive the performance differently, though this is likely to be solved by extending the time for rater training.

After reviewing the literature on the relationship between the ability levels and skills hierarchy, Brindley makes the following conclusions:

- Expert judges may perceive difficulty of the items differently from other judges and from the descriptions of task and skill hierarchy in rating scales. (Alderson 1990)

- Higher-level items failed to show better discrimination than lower-level ones which may be because they not only are dependent on language ability but also measure general knowledge or reasoning.　(Buck 1990)

- What makes the items or tasks difficult may be the wording of the questions rather than the skills being tested.　(Pollitt, et al. 1985; Lumley 1993)

By applying a Rasch approach, the FACETS programs in a scalar analysis, Kenyon and Stansfield (1996) conducted a study that compared foreign language teachers' ratings of speaking task difficulty with the level posited by the ACTFL Guidelines.　The researchers looked for empirical evidence for an underlying scale that conformed to the ACTFL Guidelines to support the validity of the Guidelines "as a description of developing competence in a second language."　The findings of their research showed a close correspondence between the teachers' scaling and the Guidelines.　Nevertheless, as Brindley argued, this agreement cannot validate the scale descriptors until further evidence showing that "the predicted order is also reflected in actual test performance…It cannot automatically be assumed that self-assessed difficulty will match task difficulty under actual test conditions."　His argument was justified by Hamp-Lyons and Mathias' study (1994) of essay prompt difficulty.　The researchers found that judges' predictions of difficulty were almost totally the reverse of the scoring patterns revealed in the test data.

II.　SLA and language assessment

> What we need to know if we want to develop good scales is not linguistic knowledge of how language is structured, what all the features of language are; we need to know how somebody acquires language, that is, what the developmental stages in language acquisition are.　(de Jong 1988 cited in Brindley 1998, p. 130)

1. The ZISA group's Multidimensional Model

The previous discussion proved from various perspectives that the descriptors of behavioural rating scales failed to describe language development as they claimed. Since it is important for language teachers and tests developers to understand how languages develop for diagnosing purposes, is there any possibility to take a developmental approach to language assessment? The answer is assertive.

Based on the empirical analysis of the developmental sequence for German L2 word-order acquisition, the first attempt to establish a preliminary profile as an assessment procedure derives from the research of Clahsen (1985).

Another significant work in this area is proposed by Pienemann, Johnston and Brindley (Pienemann et al. 1988). In this article, the authors presented an observation procedure for assessing the adult ESL (English as a second language) syntactic and morphological development. Both of the above-mentioned procedures are theoretically based on the multidimensional model of second language acquisition developed by Meisel, Clahsen and Pienemann (1981) and extended to ESL acquisition by Pienemann and Johnston (1987a). Before looking into the ESL assessment procedure, some references to the key findings of the following research group which establishes the multidimensional model should be reviewed.

The ZISA (Zweitspracherwerb italienischer und spanischer Arbeiter = second language acquisition of Italian and Spanish immigrant workers) project comprised a cross-sectional study of 45 adults, a one-year longitudinal study of three 8-year-old children and a two-year longitudinal study with 12 adult learners. This project used interview data of the naturalistic acquisition of German as an SL (GSL) by speakers of Spanish and Italian. The findings of the ZISA project that formed the basis of the assessment procedures are as follows:

➢ German word order development

It was found that both children and adult learners in the research project acquired the German word order following a five-stage developmental sequence, after the production of isolated words and formulae in an initial period.

SVO＜ADV＜SEP＜INV＜V-END

As pointed out by Pienemann and his colleagues, these learners accumulate the above-mentioned rules. In other words, "the structure of a given interlanguage can be described as the sum of all the rules the learner has acquired up to a certain point." (Pienemann et al. 1988) What makes this research group's work a significant contribution to SLA research and language assessment is that it proposes a cognitive explanation for the GSL word-order data, and this explanation has potential generalizability to other developmental sequences and to other languages such as ESL, Swedish SL and Japanese SL (Larsen-Freeman 1991).

The sequence in the acquisition of the GSL word order has been explained by the three speech-processing strategies proposed by Clahsen (1984). These strategies are Canonical Order Strategy (COS), Initialization-Finalization Strategy (IFS) and Subordinate Clause Strategy (SCS). The sequence of GSL word-order acquisition of L2 learners using various combinations of the processing strategies at different stages is illustrated by the following table.
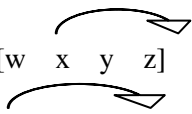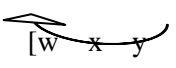
GSL Word Order Stage			Permissible permutations	Controlling Strategies
X+4	(V-END)	Sub-categorization	[w x y z]	$[-COS, -IFS, -SCS]$
X+3	(INV)	Internal movement	[w x y z]	$[-COS, -IFS, +SCS]$
X+2	(SEP)	Disruption and movement to salient position	[w x ⤻ y z]	$[-COS, +IFS, +SCS]$
X+1	(ADV)	Initialization/finalization	[w x y z]	$[+COS, +IFS, +SCS]$
X	(SVO)	Canonical order	[w x y z]	$[+COS, +SCS]$

Table 1.   Processing Strategies Underlying the GSL Word Order Stages, based on Larsen-Freeman 1991, p. 273.

> The Multidimensional Model

Another influential contribution of the ZISA group is the development of a multidimensional model.   Based on the empirical evidence provided by ZISA data, the research group discovered that there were two systematic and independent dimensions in SLA: development and variation.

Although the development of L2 acquisition follows certain invariant sequences, different learners may find his/her own path on the way to acquiring the target language.   Such differences do not conflict with the other dimension of developmental sequence.   Pienemann and his colleagues (1988) exemplified the concept of learner's orientation by demonstrating variations involving copula insertion.   Learners who are simplifying-oriented may produce equational sentences instead of the correct form (e.g. "he good") while the other type of learners labeled norm-oriented produce the sentence in the correct form immediately.   These different types of learners display the variational features systematically not only in copula insertion but also in other structures for communicative effectiveness.   (The copula

"is" in the demonstrated sentence is a grammatical device, which is redundant in the sense of communication.) Given the "incorrect" form of interlanguage structures emergent across stages of the developmental sequence, it can be certified that accuracy cannot be used as a measure of linguistic development. The L2 developmental sequence is a dynamic process. The notion that emergence of the interlanguage structure rather than mastery of the structure accounts for the evidence of development has a significant meaning in constructing the L2 assessment procedure to be introduced later.
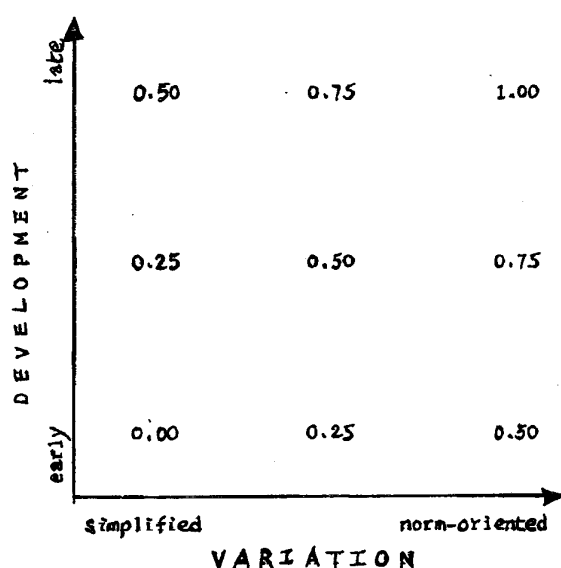

Figure 1.   A multidimensional model of second language acquisition, extracted from Pienemann, Johnston and Brindley 1988, p. 223.

2.   ESL development and profile analysis

The processing constraints in the development of GSL word-order sequence were claimed to be universal. They were claimed to be able to control the development of other interlanguages. The application of the model was done by Pienemann and Johnston (1987a) to sketch the developmental sequence of ESL

(English as a second language). The hypothesis in their study was tested against the empirical evidence provided in Johnston's (1985) cross-sectional study involving adult Vietnamese and Polish learners for the acquisition of ESL. The prediction of the order of acquisition is outlined in the following:

Stage 1: single words, formulae

Stage 2: SVO, plural marking

Stage 3: Do Fronting, Topicalization, Adverb Preposing, Neg+V

Stage 4: Pseudo-Inversion, Yes/No-Inversion

Stage 5: 3rd-Sgl-S, Aux-2nd, D-2nd.

(Pienemann, Johnston and Brindley1988, p. 228)

The eleven grammatical structures plus single words and formulae which have been identified as indicators of a developmental stage were selected to be monitored and scored in the proposed assessment procedure.

It was mentioned previously in this paper that Clahsen's (1985) preliminary procedure for assessing GSL word order adopted the methodology of profile analysis. So does the procedure devised by Pienemann and his colleagues (1988) for assessing ESL morpho-syntactic structures. Profile analysis was used in clinical linguistics where the subject's linguistic behaviour was charted and discriminated for diagnostic and remedial purposes. The most useful profiles, according to Crystal (1981), "are those which are based on an acquisitional dimension, because they can then be used simultaneously for assessment and remediation". As Clahsen remarked, linguistic profiles fulfilled the requirements he proposed as follows for L2 assessment procedures:

A.   The evaluation of oral L2 speech production should be at the core of an assessment procedure which aims at identifying the general developmental level of the learners' interlanguage.

B.   The assessment should be based on a representative sample of spontaneous speech which is gathered in a natural communicative situation.

C.   The procedure should attempt to provide a comprehensive description of the learners' interlanguage.

D.   The procedure should focus on the evaluation of syntax and morphology.

E.   The procedure should grade the linguistic structures used by the learners in terms of the order of acquisition in natural L2 development.

(Clahsen 1985, p. 286-288)

Profiling approaches have been developed and used in the assessment of language disorders, such as aphasia.   As noted by Clahsen (1985, p. 292), profiling approaches "were to provide a detailed linguistic assessment of grammatical disability and to suggest a remedial approach."   Clahsen distinguished profile analyses from formal tests by saying that "profiles are informal evaluation procedures of language behaviour which attempt to be comprehensive in the linguistic domains under investigation."   (Clahsen, ibid. p. 293)   The three characteristics of linguistic profiles described by Crystal (1982) and Clahsen (1985) corresponded to Pienemann, Johnston and Brindley's view on language assessment: "Descriptive criteria are objective; developmental criteria are psychologically plausible and interactive criteria are based on natural language use" (Pienemann et al. 1988, p. 231).

The oral observation procedure proposed by Pienemann, Johnston and Brindley (1988) has been tested for its reliability by looking at the extent of the agreement between the assessors' actual observations, their final ratings and the results of the

12

independent linguistic analysis. Although the correlations might not be satisfactory due to some errors caused by inadequate training of the assessors and wrong interpretations of the actual observations, they were significant. The linguistic observation procedure as a shorthand version of a profile analysis avoided the impracticality of profile analyses which were constrained by time, expertise, training and costs.

With the advance of information technology through the years, this procedure has been developed into a computer-assisted procedure for screening speech samples collected from language learners to assess their level of language development as compared to standard patterns in the acquisition of the target language (Mackey1991). IL data collection and elicitation for analysis is run by administering a communicative task that involves an interactive conversation between the researcher and the informant or between two informants. The naturally occurred speech sample is coded on-line using the Rapid Profile software. To analyze the data, the analyst simply clicks on the boxes which indicate the emergence or absence of targeted features on the computer interface. The information entered by the analyst is stored in the computer and then calculated by the software for locating the development level of the learner. The result of the analysis (profile) yielded by the system benefits language teachers in mapping and diagnosing the featured grammatical structures that have /not been developed by the learners.

3. Criticism of the Multidimensional Model and the developmental approach

Whenever there is a theory formed, there is challenge and criticism of it. Larsen-Freeman and Long (1991, p. 284-286) listed some problems with the Multidimensional Model, albeit with a claim that they were limitations rather than

flaws. The first and the most serious problem was that the model did not specify how language was acquired as opposed to constraints on acquisition. The second problem concerned the falsifiability of the predictive framework, such as early chunked morphology. Also, some may question the discrete point scoring method used by the profiling procedure based on the Model even though the rationale for the selection of the items is proposed. The criticism in the following by Ingram is a collection of others:

> [T]hey involve an unnatural, largely decontextualized and mechanical use of language use where meaning is much subordinated to form; they differ fundamentally from more "direct" tests because their focus is still on the piece of knowledge or skill being tested and their aim is to discriminate amongst the learners rather than specifically to assess their proficiency. (Ingram 1985, p. 236)

Despite the criticism of discrete point testing, the targeted morpho-syntactic structures for assessment are indicators of a developmental stage. Besides, discrete point testing retains value for diagnostic purposes with the fact that one thing is tested at a time. The definition of "proficiency" in the above quote may be different from how it is defined by the ZISA group.

In a nutshell, the Multidimensional Model and the ZISA group have made significant contributions to the study of SLA. This can be further supported by the following remark of Larsen-Freeman and Long (1991, p. 283):

> [T]he ZISA project was one of the first to relinquish the prevailing target-language orientation of the 1970s, thereby avoiding what Bley-Vroman has called the 'comparative fallacy in IL studies…they redefined acquisition (of a form) as the first appearance of a form in an IL, this and the subsequent evolution of form-function relationships being treated from the same learner-oriented perspective that long been taken for granted by creolists…"

As pedagogical implications are concerned, the predictive power, one of the greatest strengths of the Multidimensional Model promises the practical applications in

syllabus design, teaching methodology and language assessment (Larsen-Freeman and Long 1991, p. 287).   First, the developmental stages can be used in sequencing items in a syllabus.   Second, language teachers can utilize knowledge of developmental stages and variational features as a diagnostic resource for offering remedial help.   In the same vein, Pienemann and others identified the use and misuse of their linguistic profiling procedure by saying that "it will not predict the future learning success of the testee.   It is aimed at providing teachers with information concerning a given learner's developmental stage and hence at assisting them to make teaching learnable" (Pienemann, et al. 1988).   A third application to language assessment harmonizes the main issue discussed in this paper.   Although there are so many proficiency tests claimed as 'IL-sensitive' or 'IL-based', their validity has been found to be questionable for the lack of empirical evidence.   However, it has been proved that the oral observation procedure theoretically based on the Multidimensional Model capitalizes on research findings about developmental sequences in SLA.

III. Pedagogical implications: constructing an acquisition-based syllabus

The predictive framework of the Multidimensional Model forms the basis of the Teachability Hypothesis.   According to the Teachability Hypothesis, the course of second language development cannot be altered by factors external to the learner. (Pienemann, et al. 1988) That means the developmental stage cannot be influenced by formal instruction.   However, this does not imply that instruction cannot make any difference to learning or guarantee learning to be achieved naturally by the internal mechanisms.   As argued by Pienemann (1985), "foreign language teaching has to conform to the constraints of learnability/teachability.   Teaching can be effective

only when the learner is at the right stage for the instruction because a structure learned at one stage is a necessary prerequisite for later stages". Based on these findings in SLA development, Pienemann (1985) proposed a structural/functional syllabus, which he claimed, was "a general framework for a psychologically valid grading of teaching material". The guidelines for grading (the sequence of the material presented) are summarized as follows:

1. Do not demand a learning process which is impossible at a given stage. (i.e. order of teaching objectives be in line with stages of acquisition)

2. But do not introduce deviant forms.

3. The general input may contain structures which were not introduced for production.

General Input	Developing Interlanguage				Learning Objectives		
	structure		functions		questions	Temporal reference	etc
	Word order	Morphol.	Questions	etc			
The general input contains the learning objectives plus the 'structural consequences'	—	—	ellip. Questions Wh+x	—	ellip. Questions Wh+x	—	—
	—	—	Intonate. Marked qu.	—	Intonate. Marked qu.	connectives —	—
	verb-sep.	ge+V+t	Wh-quest. without INV	—	0	pres.perf. (+ verb-sep)	—
	INV	—		—	Wh-quest. with INV	—	—
	—	—	Wh-quest. with INV	—	—	—	—
			—			—	—

Table 3.   Communicative syllabus and interlanguage development, from Pienemann 1985, p. 67.

So far, we can almost glimpse the picture of how SLA research, language testing and language teaching weave together with the string of the developmental theory. To apply what has been discussed above to a real-life situation, a language teacher may first conduct a profiling procedure to locate the present stage of his/her learner; then constructs a learnable syllabus in accordance with the test results; finally, post-tests the learner to examine the effectiveness of the instruction and the achievement the learner has attained.

IV. Conclusion

From this discussion, we have found evidence of the interface between SLA research, language testing and language teaching. Our next step is to find out the way to increase the feasibility and applicability of the profiling procedure. It is likely to see in the near future an application of the oral procedure introduced in this paper to assessing written performance of L2 learners. Hopefully, the research in computational linguistics and corpus linguistics may be able to develop an automatic IL parsing program to replace the human assessors used in the Rapid Profile system. Moreover, it is also expectable to expand the breadth and depth of the current procedure that focuses on the development of morpho-syntactic structures domain to the development of other components of language ability, such as discourse, pragmatic competence and higher level of sentence complexity.

References

ALDERSON, J. C., 2006. *Diagnosing foreign language proficiency: the interface between assessment and learning*. London: Continuum.

BACHMAN, L. F., 1998. Language testing-SLA research interface. *In*: L. F. BACHMAN AND. A. D. COHEN, eds. *Interfaces between second language acquisition and language testing research.* Cambridge: Cambridge University Press.

BACHMAN, L. F. AND PALMER, A. S., 1996. *Language testing in practice.* Oxford: Oxford University Press.

BACHMAN, L. F. AND COHEN, A. D., 1998. Language testing-SLA research interface: an update. *In*: L. F. BACHMAN AND. A. D. COHEN, eds. *Interfaces between second language acquisition and language testing research.* Cambridge: Cambridge University Press.

BACHMAN, L. F. AND ADRIAN, P., 1984. Some comments on the terminology of language testing. *In*: R. CHARLENE, ed. *Communicative competence approaches to language proficiency assessment: research and applications.* Clevedon, England: Multilingual Matters.

BACHMAN, L. F. AND SAVIGNON, S. J., 1986. The evaluation of communicative language proficiency: a critique of the ACTFL oral interview. *Modern language journal*, 70 (4), 380-390.

BACHMAN, L. F., 1990. *Fundamental considerations in language testing.* Oxford University Press.

BRINDLEY, G., 1998. Describing language development? *In*: L. F. BACHMAN AND A. D. COHEN, eds. *Interfaces between second language acquisition and language testing research.* Cambridge: Cambridge University Press.

BROWN, A., ELDER, C., LUMLEY, T., MCNAMARA, T. AND MCQUEEN, J., 1992. Mapping abilities and skill levels using Rasch techniques. *Melbourne papers in language testing*, 1, 28-52.

BROWN, H. D., 2004. *Language assessment: principles and classroom practices.* New York: Pearson Education.

CLAHSEN, H., 1985. Profiling second language acquisition. *In*: K. HYLTENSTAMM AND M. PIENEMANN, eds. *Modeling and assessing second language development.* Clevedon: Multilingual Matters.

CRYSTAL, D., FLETCHER, P. AND GARMAN, M., 1989. *Grammatical analysis of language disability: studies in disorders of communication.* London: Whurr Publishers Ltd.

CRYSTAL, D., 1981. *Clinical Linguistics.* Wien: Springer-Verlag.

DAVIES, A., 1992. Is proficiency always achievement? *Melbourne papers in language testing*, 1, 1-16.

HAMP-LYONS, L. AND MATHIAS, S. P., 1994. Examining expert judgments of task difficulty on essay tasks. *Journal of second language writing*, 3 (1), 49-68.

HUGHES, A., 2003. *Testing for language teachers.* Cambridge: Cambridge University Press.

KENYON, D. M. AND STANFIELDS, C., 1996. Comparing the scaling of speaking tasks by language teachers and by the ACTFL Guidelines. *In*: A. CUMMING AND B. RICHARD, eds. *Validation in language testing*. Clevedon, England: Multilingual Matters.

LANTOLF, J. P. AND FRAWLEY, W., 1985. Oral-proficiency testing: a critical analysis. *The modern language journal*, 69, 337-345.

LARSEN-FREEMAN, D. AND LONG, M., 1991. *An introduction to second language acquisition research*. London: Longman.

MACKEY, A., 1991. Rapid profile: a second language screening procedure. *Working papers of the national languages institute of Australia*, 1, 61-82.

MACKEY, A., 1994. Targeting morpho-syntax in children's ESL: an empirical study of the use of interactive goal-based tasks. *Working papers in educational linguistics*, 10, 67-89.

MCNAMARA, T. F., 2000. *Language Testing*. Oxford: Oxford University Press.

PIENEMANN, M., 1985. Learnability and syllabus construction. *In*: K. HYLTENSTAMM AND M. PIENEMANN, eds. *Modeling and assessing second language development*. Clevedon: Multilingual Matters.

PIENEMANN, M., JOHNSTON, M. AND BRINDLEY, G., 1988. Constructing an acquisition-based procedure for second language assessment. *Studies in second language acquisition*, 10, 217-243.

Appendices

1. Task variation (from Pienemann, 1998, p. 280)

Task	Structure	Participants
(1) Habitual Actions	3SG S	Subject + Researcher

This task involved a set of photographs depicting "a day in the life of someone such as a librarian or a police officer". Subjects were asked questions such as "what does a librarian do every day?".

(2) Story Completion	Wh Questions	Subject + Researcher

Subjects were shown a set of pictures, which were in order, and were then given instructions to find a story behind the pictures. They were encouraged to ask for information to enable them to guess the story. One example of this task used pictures of a man who had been given poison and who needed to find an antidote.

(3) Informal Interview	General	Subject + Researcher

Subjects were interviewed informally and with sensitivity by the researcher. The situation was designed to be as close as possible to a friendly chat. The researcher asked questions of the subject and subjects were encouraged to ask questions of the researcher as well.

(4) Picture Sequencing	Questions	Subject + Subject

Subjects were each given part of a sequence of pictures. Together the parts made up one story. The pictures were lettered so that they could be identified for discussion. In order to sequence the pictures with the story, questions had to be formed and responses given which were sometimes negative. An example of this task was a story which involved a man being assaulted by three different people on his way home form work.

(5) Picture Differences	Negations/Questions	Subject + Subject

Subjects were given on picture each of the "Spot the Difference" variety. They were told that there were a number of differences. They ha to ask questions and make positive and negative responses in order to find the difference.

(6) Meet Partner	Questions	Subject + Subject

Subjects in dyads asked each other questions to find out information and then were given the opportunity to introduce each other to the researcher.

2. RP labels: L2 English Developmental Stages, extracted from:

http://groups.uni-paderborn.de/rapidprofile/docs/Stages.pdf

Stage	Phenomena	Examples
6	Indirect Question (SVO)	I wonder what he wants.
5	Neg/Aux-2nd-? Aux-$2^{nd\ -?}$ 3sg-s-	Why didn't you tell me? Why can't she come? Why did she eat that? What will you do? Peter likes bananas.
4	Copula S (x) Wh-copula S (x) V-Particle	Is she at home? Where is she? Turn it off!
3	Do-SV (O)-? Aux SV (O)-? Wh-SV- (O)-? Adverb-1st Poss (Pronoun) Object (Pronoun)	Do he live here? Can I go home? Where she went? What you want? Today he stay here. I show you my garden. This is your pencil. Mary called him.
2	S neg V (O) SVO SVO-Question -ed -ing Plural-s (Noun) Poss-s (Noun)	Me no live here./I don't live here. Me live here. You live here? John played. Jane going. I like cats. Pat's cat is fat.
1	Words Formulae	Hello, Five Dock, Central How are you? Where is X? What's your name?

About the author

Ming-Fen Lo is currently a second year IPhD student in Educational and Applied

Linguistics. Email: ming-fen.lo@ncl.ac.uk.