Nonverbal Delivery in Speaking Assessment

Mingwei Pan

Nonverbal Delivery in Speaking Assessment

From an Argument to a Rating Scale Formulation and Validation

Springer

Mingwei Pan
Faculty of English Language and Culture
Guangdong University of Foreign Studies
Guangzhou, Guangdong
China

ISBN 978-981-10-0169-7 ISBN 978-981-10-0170-3 (eBook)
DOI 10.1007/978-981-10-0170-3

Library of Congress Control Number: 2015955873

Springer Singapore Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer Science+Business Media Singapore Pte Ltd. is part of Springer Science+Business Media (www.springer.com)

Preface

Language, especially its spoken form, is now universally recognised as being highly complex, multidimensional and variable according to a multitude of social, contextual and affective factors (Bachman 1990). Accordingly, the taxonomy of language ability has been long arguably partitioned into a plethora of dimensions. Bachman (1990) and Bachman and Palmer (1996) propose the framework of communicative language ability (CLA), whose explanatory power and inclusiveness, as compared against the other frameworks of a similar nature concerning communicative competence, can be justified as the fittest model to which speaking assessment can be referred. In particular, deriving from strategic competence, nonverbal delivery, with its interactive and interdependent role with the accompanying verbiage, should be judged as an indispensable component in speaking assessment, especially when its promoting effect on communicativeness is considered. However, notwithstanding well-documented research might be omnipresent with regard to where learners in the Chinese EFL context are poor at, how they perform their nonverbal delivery in speaking assessment seems to be comparatively underexplored (e.g. Liu and Pan 2010a, b; Pan 2011a, b). Against this, there arises a need of examining Chinese EFL learners' oral proficiency in an all-round manner, via which a rating scale incorporating nonverbal delivery is to be formulated and validated.

With the above background as a point of departure, this research project, contextualised in formative assessment, mainly aims at (1) building an empirical argument of embedding nonverbal delivery into speaking assessment; (2) developing a rating scale, with nonverbal delivery included as a dimension, for assessing candidates' communicative competence in group discussion; and (3) cross-validating the proposed rating scale with multi-trait multi-method (MTMM) as well as multimodal discourse analysis (MDA) approaches. These three aims also constitute the three phases of research (henceforth AB phase, RSF phase and RSV phase, respectively) for the present study.

The data of this project are 150 samples of group discussion by Chinese EFL learners at the tertiary level in the context of formative assessment. For

phase-specific purposes, all the samples are accordingly video-recorded, transcribed, processed and analysed. Except for the 30 samples used in the AB phase, the other 120 samples are scored by expert raters, teacher raters and/or peer raters specific to the design of the latter two research phases.

In the AB phase, this study conducts an empirical study to explore the role of nonverbal delivery in Chinese EFL candidates' performance in group discussion, particularly how candidates across a range of proficiency levels can be discriminated with regard to their nonverbal delivery. In a sense, if nonverbal delivery can statistically discriminate the candidates of predetermined proficiency levels, an argument of incorporating nonverbal delivery into speaking assessment can be accordingly advanced. The descriptive, comparative and extrapolative statistics in this phase of study find that although there seems to be a generally low profile of employing nonverbal delivery by the observed candidates in group discussion, they can be statistically discerned vis-à-vis eye contact, gesture and head movement.

Candidates of advanced proficiency are characterised by higher frequency and longer duration of eye contact. Elementary-level candidates, though featuring a high frequency of eye contact occurrences, are inclined to shift their gaze hurriedly and not able to instantiate durable eye contact with the peer discussants. In addition, rather than enhance communication effectiveness, most occurrences of their eye contact, if not all, serve regulatory or adaptive purpose. Although intermediate-level candidates are found to present eye contact with their peers, the degree to which their eye contact can serve attentive purpose would be more impaired compared with the advanced-level counterparts. Candidates' gestures can be mainly distinguished from the perspectives of frequency, diversity and communication-conduciveness. Advanced candidates would be able to perform satisfactorily in all of the above measures, whereas candidates of elementary proficiency level are found to maintain an extremely low profile of resorting to gestures in accompanying their verbal language. Although intermediate-level candidates can be judged to perform well in gesturing frequency and diversity, a number of gesture occurrences are found to serve adaptive or performative purpose, failing to be a remarkable enhancer for intended meaning conveyance. When head movement is probed into, head nod and shake are the main manifestations. It has to be noted that, given the socio- and cultural preponderance, candidates are not significantly different in presenting lower frequency of head shake than head nod, yet whether they perform certain head movements appropriately in the given social context might be referred to as a discriminating point because candidates are found to nod even though certain negative meanings are intended.

Enlightened by the findings in the AB phase, this study draws an interim conclusion that nonverbal delivery, as reflected by eye contact, gesture and head movement, can be one of the indicators for assessing candidates' overall spoken English production and that what has been extracted to discern candidates across various proficiency levels can usefully and effectively inform how a new rating scale can be formulated consequently.

When such a rating scale is developed, two broad dimensions are perceived in the RSF phase: language competence and strategic competence. The former is

formulated by an operationalised questionnaire drawn from the related spectra of CLA model. After an exploratory factor analysis from the Chinese EFL teaching practitioners' and learners' responses to the constituents of language competence in group discussion, this study distils and organically brings forth three assessment dimensions representing language competence: Pronunciation and Intonation (D1), Grammar and Vocabulary (D2) and Discourse Management (D3). The gradable descriptors of these dimensions have been written and further fine-grained by referring to the statements in the questionnaires. Based on the review over the definitions of strategic competence and the empirical argument in the AB phase, Nonverbal Delivery (D4) is perceived as the fourth dimension on the rating scale proposed. In writing the descriptors for this dimension, what can observably and feasibly discriminate candidates regarding their nonverbal delivery in the AB phase is referred to.

A four-dimensional rating scale, therefore, is tentatively formulated, and it epitomises what would supposedly be measured in relation to communicative competence in group discussion, as guided by CLA model. Considering the fact that the expert raters' scoring reveals a high correlation between two assessment dimensions, this rating scale can be initially certified to be valid in its construct, yet it would be subject to certain modifications in wording, disambiguation and the shrinkage of bands from five to four for a higher degree of rater-friendliness.

The rating scale, afterwards, is phased into the RSV phase, where both quantitative and qualitative approaches are employed. When MTMM is deployed following Widaman's (1985) alternative model comparison method, it is found that, considering the interpretability and consistency with previous studies regarding speaking ability taxonomy, a second-order correlated trait/uncorrelated method model not only provides sound goodness-of-fit indices ($\chi^2(28) = 462.796$, $p = 0.818$; CFI = 1.000; NNFI = 1.024; SRMR = 0.015; RMSEA = 0.000; 90 % C.I. = 0.000, 0.060), but also presents divergent validity ($\Delta\chi^2(9) = 403.08$, $p < 0.001$; ΔCFI = 0.472) and discriminant validity ($\Delta\chi^2(17) = 425.68$, $p < 0.001$; ΔCFI = 0.146). The standardised parameter estimates and trait–method correlations reveal no method effect or bias concerning rating methods. Thus, this rating scale, with nonverbal delivery included as a crucial dimension, has been validated in a statistical spectrum.

The rating scale, especially its assessment dimension of Nonverbal Delivery, is further validated on a micro basis with an MDA approach, with a special reference to an integrated analytic framework drawn from Martinec's (2000a, b, 2001, 2004) taxonomy of action and Hood's (2007, 2011) works on nonverbal delivery. Three randomly selected candidates (pseudonyms: Tom, Linda and Diana) representing different proficiency levels are probed into concerning their de facto performance in nonverbal delivery. Tom, with a nonverbal delivery subscore of 1.5, is found to be rather sedentary and passive in the group discussion because only a limited number of captured nonverbal channels with ideational meanings are instantiated. A majority of his nonverbal delivery occurrences retain to be performative, or as a likely regulation to adapt himself to an assessment setting. In that sense, almost no interpersonal or textual meanings can be interpreted from his nonverbal delivery;

thus, Tom is reduced to stagnation where only the mere occurrence of nonverbal delivery employment can be captured. In stark contrast, Diana, as a representative of advanced proficiency level who is assigned a full mark in nonverbal delivery, is found to be articulate in eclectically resorting to a repertoire of nonverbal channels in accompanying her verbiage. At certain points, her nonverbal performance can also instantiate intended meanings in the absence of any synchronised verbal language. Judging from the perspective of metafunctions, she is found to be capable of realising a variety of meaning potentials via nonverbal delivery. Although she seems somewhat aggressive in group discussion, her frequent shift in instantiating different nonverbal channels with discrepant metafunctions would impress other discussants as an active and negotiable speaker as well as an attentive listener. Although Linda, whose subscore of nonverbal delivery is 3, performed quite satisfactorily in terms of formal nonverbal channels, she is found to be slightly passive and hesitant in the group discussion. In particular, when the interpersonal meaning of her gestures is looked into, she seems to be self-contained and strike a certain distancing effect on the peer discussants. The above profile of the three candidates' performance on nonverbal delivery can also be aligned with the descriptors of nonverbal delivery on the rating scale, thus lending weightier support to validate the proposed rating scale.

This research project yields significance in the sense that it organically integrates multimodal discourse analysis, a research method scarcely explored in language assessment, with rating scale validation, thus extending the literature of applying this method to more research of a similar kind. In addition, based on the research findings, how nonverbal delivery can penetrate into EFL learning and teaching is also enlightened and suggested. In particular, this thesis illuminates how EFL textbooks should be multimodally compiled for a heavier load of meaning making and how EFL teaching can be optimised with nonverbal delivery by teaching practitioners incorporated in daily instruction.

References

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

Bachman, L.F., and A.S. Palmer. 1996. *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.

Hood, S.E. 2007. *Gesture and meaning making in face-to-face teaching*. Paper presented at the Semiotic Margins Conference, University of Sydney.

Hood, S.E. 2011. Body language in face-to-face teaching: a focus on textual and interpersonal meaning. In *Semiotic margins: meanings in multimodalities*, eds. S. Dreyfus, S. Hood, and S. Stenglin, pp. 31–52. London: Continuum.

Liu, Q., and M. Pan. 2010a. A tentative study on non-verbal communication ability in Chinese college students' oral English. *Computer-assisted Foreign Language Education in China* (2):38–43.

Liu, Q., and M. Pan. 2010b. Constructing a multimodal spoken English corpus of Chinese Science and Engineering Major Learners. *Modern Educational Technology* (4):69–72.

Martinec, R. 2000a. Construction of identity in Michael Jackson's "Jam". *Social Semiotics* 10 (3):313–329.
Martinec, R. 2000b. Types of processes in action. *Semiotica* 130(3):243–268.
Martinec R. 2001. Interpersonal resources in action. *Semiotica* 135(1):117–145.
Martinec, R. 2004. Gestures that co-occur with speech as a systematic resource: the realisation of experiential meanings in indexes. *Social Semiotics* 14(2):193–213.
Pan, M. 2011a. *Reconceptualising and reexamining communicative competence: a multimodal perspective*. Unpublished Ph.D. thesis. Shanghai: Shanghai International Studies University.
Pan, M. 2011b. Incorporating nonverbal delivery into spoken English assessment: a preliminary study. *English Language Assessment 6*:29–54.
Widaman, K.F. 1985. Hierarchically tested covariance structure models for multi-trait multi-method data. *Applied Psychological Measurement* 9:1–26.

Acknowledgments

A great many people have helped in the writing of this book. In particular, I feel profoundly indebted to Professor David D. Qian from Hong Kong Polytechnic University, whose resourcefulness, insightfulness and supportiveness have removed my amateurishness in research and nurtured me towards professionalism in the academia of language assessment.

I also need to thank other scholars who so generously offered their time and voices—Professor Frederick G. Davidson from University of Illinois at Urbana-Champaign, Professor Alister Cumming from University of Toronto and Professor Zou Shen from Shanghai International Studies University—and all the others whose voices are also recorded here. Their scholarship continues to be a great source of stimulation.

I would also like to thank the Springer editorial team who have been such a pleasure to work with, in particular Ms. Rebecca Zhu and Ms. Yining Zhao.

Contents

Acronyms and Abbreviations

ACTFL	American Council of the Teaching of Foreign Languages
AERA	American Educational Research Association
AGFI	Adjusted Goodness-of-Fit Index
AOM	Agent-oriented modelling
APA	American Psychological Association
ASD	Average sample duration
ASLPR	Australian Second Language Proficiency Ratings
AUA	Assessment Use Argument
BEC	Business English Certificate
CA	Conversation analysis
CEFR	Common European Framework of Reference
CET	College English Test
CET4	College English Test Band 4
CET6	College English Test Band 6
CET-SET	College English Test Spoken English Test
CFA	Confirmatory factor analysis
CQUPT	Chongqing University of Posts and Telecommunications
CLA	Communicative language ability
CLC	Communicative language competence
CV	Convergent validity
DM	Discourse Management (rating scale dimension 3)
DV	Discriminant validity
EC	Eye contact
EC/c	Eye contact with the camera
EC/p	Eye contact with the peer(s)
EC/n	Eye contact with none
EC/r	Eye contact with the researcher
ECD	Evidence-centred design
ECUST	East China University of Science and Technology
EFA	Exploratory factor analysis
EFL	English as a Foreign Language

ETS	English Testing Service
FSI	Foreign Service Institute
GV	Grammar and Vocabulary (rating scale dimension 2)
HD	Hypothetic–deductive
HIT	Harbin Institute of Technology
IELTS	International English Language Testing System
MDA	Multimodal discourse analysis
AT-MDA	Activity theory multimodal discourse analysis
SF-MDA	Systemic functional multimodal discourse analysis
MDT	Mediated discourse theory
ME	Method effect
MTMM	Multi-trait multi-method
NAEP	National Assessment of Educational Programme
NCME	National Council on Measurement in Education
ND	Nonverbal delivery (rating scale dimension 4)
NNFI	Non-normed fit index
NUST	Nanjing University of Science and Technology
OPI	Oral Proficiency Interview
P-rating	Peer-rating
PDT	Performance decision tree
PETS-OT	Oral Test of the Public English Test System
PI	Pronunciation and intonation (rating scale dimension 1)
RMSEA	Root mean square error of approximation
SDA	Situated discourse analysis
SEM	Structural equation modelling
SFL	Systemic functional linguistics
SNU	Shanghai Normal University
SISU	Shanghai International Studies University
SRMR	Standardised root mean square residual
T-rating	Teacher-rating
TEEP	Test of English for Educational Purposes
TEM	Test for English Majors
TEM-OT	Test for English Majors Oral Test
TEM4-OT	Test for English Majors Band 4 Oral Test
TLI	Tucker–Lewis Index
TOEFL	Test of English as a Foreign Language
USST	University of Shanghai for Science and Technology
VPA	Verbal protocol analysis

List of Figures

List of Tables