



Title of Project:

Test-takers' Views of the College English Test in China and their Effect on Test Performance: A Mixed Methods Study

Researcher:

Bai Ying
University of Melbourne
by@baiying.ch



Bai Ying

Research Supervisor:

Prof. Tim McNamara
University of Melbourne

Final Report

The first factor that inspired this research is the development of validity theories in language testing. Since the 1980s, validity theory in educational measurement and language testing has witnessed a shift from a focus on validity investigation of the test itself to score interpretation for a particular test use. Messick (1989), whose unified validity theory represents the apex of this development, included “value implications” and “social consequences” underlying test interpretation and use within the scope of test validity. This extension in the scope of language test quality investigation highlights the rights of stakeholders in the process, as set out in statements in relation to “Test Fairness” and “Ethical Codes of Language Testing” (Davies, 2008; Kunnan, 2000, 2003). Proponents of “Critical Language Testing” and argument-based validity theories (Bachman, 2005; Bachman & Palmer, 2010; Kane, 2004, 2006; Shohamy, 2001) have called for contributions to test development and validation from a broader test stakeholder audience rather than being limited to the ideas of “experts”.

The second factor that has inspired this research to particularly focus on students’ perspectives is that students’ perspectives may interact with their test performance. The relationship of students’ perspectives, i.e., their attitudes towards a second and/or a foreign language and language achievement has been, to a certain extent, established in a socio-psychological approach to second language acquisition (SLA) (e.g., Dörnyei, Csizér, & Németh, 2006; Gardner & Clement, 1990; McKenzie, 2008); whilst in the field of language testing, whether and how students’ attitudes interact with their test performance is still under investigation. In addition to language communicative ability, Bachman (1990) suggested that students’ personal attributes, including gender, ethnic background, cognitive style, cultural background and attitude (Kunnan, 1995), may be factors that affect their test performance. To date only a limited number of studies have explored the potential interactions between students’ attitudes and their test performance, among which students’ attitude variables have focused primarily on the psychometric quality of the test (e.g., Rasti, 2009). There is a distinct lack of research that explores how students’ attitudes towards the consequences of a test interact with test performance.



The third factor lies in the observed complexity of the interplay of central language policy at a local level in China, a test-oriented social context from historical times. As was discussed earlier, this complexity includes the disparities between the effective English teaching and learning intended to be achieved through the implementation of the test and the common practice of teaching to the test and learning for the test in reality; and the multiple decisions made on the basis of test results which are not intended by the test design and thus claimed as misuses by the NCETC.

In light of the discussion above, this project explores two main research questions (RQs) in two phases. Qualitative phase (Phase I) explores main RQ1:

1. What are students' attitudes towards the CET?

Students' perspectives focus on three aspects of the test which have been highlighted in both Bachman and Palmer's assessment use argument (Bachman & Palmer, 2010) and the CET syllabus (NCETC, 2006). The sub-questions under RQ1 include:

- a) What are students' attitudes towards the test quality of the CET?
- b) What are students' attitudes towards the test use of the CET?
- c) What are students' attitudes towards the approach to test preparation for the CET?

Students' attitudes focus on the CET4, given that in the whole of the CET testing system, CET4 is a test that all non-English majors take in Chinese tertiary institutions, while other tests (e.g., CET6 and CET-SET) are optional.

Quantitative phase (Phase II) explores main RQ2:

2. How do students' individual characteristics and attitudes affect their test performance?

In addition to the potential interactions between students' attitudes and their test performance, students' individual characteristics such as gender, English learning motivation and learning experience are also included as potential differentiating factors to their attitudes.

Students' individual characteristics such as English learning motivation and learning experience are also explored in the qualitative phase (Phase I) to capture their contextual characteristics, including:

- d) What is the nature of students' English learning motivation in this particular research context?
- e) What is the nature of students' past English learning experience in this particular research context?

This project adopts an exploratory sequential mixed methods research design, combining both qualitative and quantitative data in two phases (Creswell, 2014). The qualitative phase (Phase I) explores students' attitudes towards the CET and the nature of their English learning experience and learning motivation. The findings of Phase I also provide fundamental information for the



development of the data collection instrument, a questionnaire for Phase II, the quantitative phase. In other words, the findings in Phase I are substantially integrated into Phase II to elicit the appropriate data, thus maximising the inclusion of the characteristics unique to this research context. Phase II investigates the hypothesised interactions with students' attitudes, using structural equation modeling analysis.

The findings of this study also demonstrated that students'/test takers' attitudes towards a test and its test consequences may significantly affect their test performance. In addition, students' individual differences may determine and differentiate their attitudes towards a test, and mediated through attitude factors, they may have an indirect effect on test performance. A major concern in the design and development of language tests is to minimize the effects of factors that are not part of examined language ability. Likewise, the interpretations and uses of a test score should clarify and mitigate the extent to which a test score reflects factors other than the measured language abilities. Identifying the factors that either systematically or randomly affect test performance, therefore, eventually contributes to the validity of the interpretation of test scores and justification of the validity of a test (Bachman, 1990, 2004). The significant relationships identified in this study provided empirical evidence that test takers' attitudes towards a test could be one of those factors that deserve appropriate attention in test validation.

The significant role of students/test takers that has been revealed in this study is not unique to this particular research context. The findings of this study support the argument that students' perspectives deserve appropriate attention in the process of testing. First, students are in a unique position in revealing the complex interplay of central language policy in local settings, thus providing meaningful and valuable information for test development and validation. Second, students' attitudes towards a test may have a significant impact on test performance, and thus could be identified as factors that may undermine the validity of inferences derived from test results.



References

- Adamson, B., & Morris, P. (1997). The English curriculum in the People's Republic of China. *Comparative Education Review*, 41(1), 3-26.
- Ajzen, I. (2005). *Attitudes, personality and behaviour* (2nd ed.). Berkshire, UK: Open University Press.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317-332.
- Akiyama, T. (2004). *Introducing speaking tests into Japanese Senior High School entrance examination* (Unpublished doctoral dissertation). The University of Melbourne, Australia.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing*, 9(2), 101-119.
- Allport, G. W. (1971). Attitudes. In K. Thomas (Ed.), *Attitudes and behaviour*. Harmondsworth, England: Penguin.
- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33-40.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Ames, C. (1992). Classroom: Goals, structures, and students motivation. *Journal of Educational Psychology*, 84(3), 261-271.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43(4), 695-716.
- Bachman, L. F. (1986). The test of English as a foreign language as a measure of communicative competence. In C. W. Stansfield (Ed.), *Toward communicative competence testing: Proceedings of the second TOEFL invitational conference* (pp. 69-88). Princeton, NJ: Educational Testing Service.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.



- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34.
- Bachman, L. F. (2006). Generalisability: A journey into the nature of empirical research in applied linguistics. In M. Chalhoub-Deville, C. A. Chapelle, & P. Duff (Eds.), *Inference and generalisability in applied linguistics: Multiple perspectives* (pp. 165-207). Dordrecht, the Netherlands: John Benjamins.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *The Modern Language Journal*, 70(4), 380-390.
- Bai, Y. (2010, Oct.). *Investigating supply and demand in a college English classroom*. Paper presented at The Second Teaching Forum of China Agricultural University, Yantai, Shandong, China.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Bailey, K. M. (1999). *Washback in language testing*. Princeton, NJ: Educational Testing Service.
- Baker, C. (1992). *Attitudes and language*. Clevedon, UK: Multilingual Matters.
- Barbour, R. S. (2014). Analysing focus groups. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 313-326). London, UK: SAGE.
- Bentler, P. M. (2005). *EQS 6 Structural equations program manual*. Encino, CA: Multivariate Software.
- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modelling. *Sociological Methods and Research*, 16(1), 78-117.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons, Inc.
- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). New York, NY: Greenwood.



- Bozdogan, H. (1987). Model selection and Akaike's information criteria (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(2), 345-370.
- Bradshaw, J. (1990). Test-takers' reactions to a placement test. *Language Testing*, 7(1), 13-30.
- Bray, M., & Steward, L. (1998). *Examination systems in small states: Comparative perspectives on policies, models and operations*. London, UK: Commonwealth Secretariat.
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54(1), 106-148.
- Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, 10(2), 161-180.
- Brown, A. (1993). The role of test-taker feedback in the test development process: test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing*, 10(3), 277-303.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge, UK: Cambridge University Press.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1), 62-83.
- Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioural Research*, 24(4), 445-455.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: SAGE.
- Bryman, A. (2004). *Social research methods* (2nd ed.). Oxford, UK: Oxford University Press.
- Burrows, C. (2001). Searching for washback: The impact of assessment in the Certificate in Spoken and Written English. In G. Brindley & C. Burrows (Eds.), *Studies in immigrant English language assessment*. Sydney, Australia: National Center for English Language Teaching and Research.
- Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Y. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Byrne, B. M. (2001). *Structural equation modelling with AMOS: Basic concepts, applications and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.



- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Psychology Press.
- Cai, J. (2006). *ELT at tertiary level in China: Review, reflection and research*. Shanghai: Fudan University Press.
- Chan, D., Schmitt, N., DeShon, R., Clause, C., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, 82(2), 300-309.
- Chapelle, C. A. (2012). Conceptions of validity. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 66-98). New York, NY: Routledge.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the test of English as a foreign language*. New York, NY: Routledge.
- Charmaz, K. (2014). *Constructing grounded theory*. Los Angeles, CA: SAGE.
- Chen, J. F., Warden, C. A., & Chang, H. (2005). Motivators that do not motivate: The case of Chinese EFL learners and the influence of culture on motivation. *TESOL Quarterly*, 39(4), 609-633.
- Cheng, L. (1997). How does washback influence teaching: Implications for Hong Kong. *Language and Education*, 11(1), 38-54.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge, UK: Cambridge University Press.
- Cheng, L. (2008). The key to success: English language. *Language Testing*, 25(1), 15-37.
- Cheng, L. (2010). The history of examinations: Why, how what and whom to select? In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 13-25). New York, NY: Routledge.
- Cheng, L., & Qi, L. (2006). Description and examination of the National Matriculation English Test in China. *Language Assessment Quarterly: An International Journal*, 3(1), 53-70.
- Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modelling. In R. H. Hoyle (Ed.), *Structural equation modelling: Issues and applications* (pp. 37-55). Newbury Park, CA: SAGE.
- Chu, H. (1990). *Learning to be a sage: Selections from the conversations of Master Chu*. Berkeley, CA: University of California Press.



- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Coleman, D., Starfield, S., & Hagan, A. (2003). The attitudes of IELTS stakeholders: Student and staff perceptions of IELTS in Australian, UK and Chinese tertiary institutions. *IELTS Australia Research*, 5, 20-34.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: SAGE.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: SAGE.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: SAGE.
- Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York, NY: Harper and Row.
- Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, theory and public policy* (pp. 147-171). Urbana, IL: University of Illinois Press.
- Csizér, K., & Kormos, J. (2009). Learning experience, selves and motivated learning behaviour: A comparative analysis of structural models for Hungarian secondary and university learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 98-119). Bristol, UK: Multilingual Matters.
- Davies, A. (1977). The construction of language tests. In J. P. B. Allen & A. Davies (Eds.), *Testing and experimental methods. The Edinburgh course in applied linguistics* (Vol. 4, pp. 38-194). Oxford, UK: Oxford University Press.
- Davies, A. (2008). Ethics, professionalism, rights and codes. In S. May & N. H. Hornberger (Eds.), *Encyclopaedia of Language and Education* (Vol. 7, pp. 429-444): Boston, MA: Springer.
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292-307.
- Donahue, R. T. (1998). *Japanese culture and communication*. Lanham, MD: University Press of America.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.



- Dörnyei, Z. (2009). The L2 motivational self system. In E. Ushioda & Z. Dörnyei (Eds.), *Motivation, language identity and the L2 self* (pp. 9-43). Bristol, UK: Multilingual Matters.
- Dörnyei, Z., & Csizér, K. (2002). Some dynamics of language attitudes and motivation: Results of a longitudinal nationwide survey. *Applied Linguistics*, 23(4), 421-462.
- Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: A Hungarian Perspective*. Clevedon, UK: Multilingual Matters.
- Dörnyei, Z., MacIntyre, P. D., & Henry, A. (Eds.). (2015). *Motivational dynamics in language learning*. Bristol, UK: Multilingual Matters.
- Eagly, A., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FA: Harcourt Bruce.
- Eckstein, M. A., & Noah, H. J. (1992). *Examinations: Comparative and international studies*. Oxford, UK: Pergamon Press.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford, UK: Oxford University Press.
- Fan, J., & Jin, Y. (2010). Standards for language testing: Review, reflection and implications. *Foreign Language World*, 1, 82-91.
- Fan, J., & Jin, Y. (2011). *The way towards a code of practice: A survey of EFL testing in China*. Paper presented at the 33rd Language Testing Research Colloquium, Ann Arbor, MI.
- Flick, U. (Ed.). (2014). *The SAGE handbook of qualitative data analysis*. London, UK: SAGE.
- Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. London, UK: The Falmer Press.
- Fullan, M. (1999). *Change forces: The sequel*. London, UK: The Falmer Press.
- Fullwinder, R. K. (1996). Professional codes and moral understanding. In M. Coady & S. Block (Eds.), *Codes of ethics and the professions* (pp. 72-87). Melbourne, Australia: Melbourne University Press.
- Gao, S., Mokhtarian, P. L., & Johnston, R. A. (2008). Nonnormality of data in structural equation models. *Transportation Research Record: Journal of the Transportation Research Board*, 2082(1), 116-124.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London, UK: Edward Arnold.



- Gardner, R. C. (2001). Integrative motivation and second language acquisition. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 1-20). Honolulu, HI: University of Hawaii Press.
- Gardner, R. C. (2002). Social psychological perspective on second language acquisition. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 160-169). Oxford, UK: Oxford University Press.
- Gardner, R. C., & Clement, R. (1990). Social psychological perspectives on second language acquisition. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 495-517). New York, NY: John Wiley.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.
- Gardner, R. C., Masgoret, A. M., & Trembley, P. (1999). Home background characteristics and second language learning. *Journal of Language and Social Psychology, 18*(4), 419-437.
- Garner, M., & Huang, D. (2014). *Testing a nation: The social and educational impact of the College English Test in China*. Bern, Switzerland: Peter Lang.
- Gu, W., & Liu, J. (2005). Test analysis of college students' communicative competence in English. *Asian EFL Journal, 7*(2), 118-133.
- Gu, X. (2005). *Positive or negative? An empirical study of CET washback on College English teaching and learning in China* (Unpublished doctoral dissertation). Shanghai Jiaotong Studies University, China.
- Gu, X. (2007). An empirical study of CET washback on college English teaching and learning in China. *Journal of Xi'an International Studies University, 15*(4), 39-45.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Thousand Oaks, CA: SAGE.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Education.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing, 14*(3), 295-303.
- Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System, 28*(4), 579-591.
- Harlow, L. L. (1985). *Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study*. (Unpublished doctoral dissertation). University of California Los Angeles, CA.



- Harris, D. P. (1969). *Testing English as a second language*. New York, NY: McGraw Hill.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge, UK: Cambridge University Press.
- He, Q. (2001). English language education in China. In S. J. Baker (Ed.), *Language policy: Lessons from global models* (pp. 225-231). Monterey, CA: Monterey Institute of International Studies.
- Heaton, J. B. (1975). *Writing language tests*. London, UK: Longman.
- Henry, A. (2015). The dynamics of possible selves. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning*. Bristol, UK: Multilingual Matters.
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94(3), 319-340. doi: 10.1037//0033-295x.94.3.319
- Higgins, L., & Sun, C. H. (2002). The development of psychological testing in China. *International Journal of Psychology*, 37(4), 246-254.
- Hoare, R. (1999). *Attitudes towards language varieties in Brittany* (Unpublished doctoral dissertation). University of the West of England, Bristol, UK.
- Hu, C. (1984). The historical background: Examinations and controls in pre-modern China. *Comparative Education*, 20(1), 7-26.
- Hu, G. (2002). Recent important development in secondary English-language teaching in the People's Republic of China. *Language, Culture and Curriculum*, 15(1), 30-49.
- Hu, G. (2003). English language teaching in China: Regional differences and contributing factors. *Journal of Multilingual and Multicultural Development*, 24(4), 290-318.
- Hu, G. (2005). Contextual influence on instructional practices: A Chinese case for an ecological approach to ELT. *TESOL Quarterly*, 39(4), 635-660.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues and applications* (pp. 76-99). Thousand Oaks, CA: SAGE.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modelling: Sensitivity to underparameterised model misspecification. *Psychological Methods*, 3(4), 424-453.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling*, 6(1), 1-55.



- Huang, D. (2002). *A preliminary investigation of CET4 washback*. Paper presented at the International Conference on Language Testing and Language Teaching, Shanghai: Shanghai Jiaotong Studies University.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading, UK.
- Huhta, A., Kalaja, P., & Pitkänen-Huhta, A. (2006). Discursive construction of a high-stakes test: The many faces of a test-taker. *Language Testing*, 23(3), 326-350.
- Ichikawa, M., & Konishi, S. (1995). Application of the bootstrap methods in factor analysis. *Psychometrika*, 60(1), 77-93.
- Ingram, D. E. (1977). Basic concepts in testing. In J. P. B. Allen & A. Davies (Eds.), *Testing and experimental methods. The Edinburgh course in applied linguistics* (Vol. 4, pp. 11-37). Oxford, UK: Oxford University Press.
- Islam, M., Lamb, M., & Chambers, G. (2013). The L2 motivational self system and national interest: A Pakistani perspective. *System*, 41(2), 231-244. doi: 10.1016/j.system.2013.01.025
- Jackson, D. L., Gillaspay Jr., J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6-23.
- Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes* (2nd ed.). Boston, MA: Houghton Mifflin.
- Jin, Y. (2008). Powerful tests, powerless test designers: Challenges facing the College English Test. *CELEA Journal (Bimonthly)*, 31(5), 3-11.
- Jin, Y. (2011). Fundamental concerns in high-stakes language testing: The case of the College English Test. *Journal of Pan-Pacific Association of Applied Linguistics*, 15(2), 71-83.
- Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7 user's reference guide*. Chicago, US: Scientific Software.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modelling with the SIMPLS command language*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kane, M. T. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112(3), 527-535.
- Kane, M. T. (2002). Validating high-stakes testing programs. *Educational measurement: Issues and practice*, 21(1), 31-41.



- Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research and Perspectives*, 2(3), 135-170.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (pp. 17-64). Westport, CN: Praeger.
- Kane, M. T., Crooks, T. J., & Cohen, A. S. (1999). Validating measures of performance. *Educational measurement: Issues and practice*, 18(2), 5-17.
- Kirk, J., & Miller, M. (1986). *Reliability and validity in qualitative research*. London, UK: SAGE.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.
- Kunnan, A. J. (1995). *Test taker characteristics and test performance: A structural modeling approach* (Vol. 2). Cambridge, UK: Cambridge University Press.
- Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium* (pp. 1-14). Cambridge, UK: Cambridge University Press.
- Kunnan, A. J. (2003). Test fairness. In M. Milanovic & C. Weir (Eds.), *Selected papers from the European Year of Languages Conference, Barcelona* (pp. 27-48). Cambridge, UK: Cambridge University Press.
- Lai, C. (1970). *A scholar in imperial China*. Hong Kong: Kelly & Walsh.
- Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, 62(4), 997-1023. doi: 10.1111/j.1467-9922.2012.00719.x
- Lambert, W. E. (1963). Psychological approaches to the study of language: Part I: On learning, thinking and human abilities. *Modern Language Journal*, 14(2), 51-62.
- Lambert, W. E. (1967). A social psychology of bilingualism. *Journal of Social Issues*, 23(2), 91-109.
- Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of non-normality in structural equation modelling. *Structural Equation Modelling: A Multidisciplinary Journal*, 12(1), 1-27.



- Leung, A. K. C. (1994). Elementary education in the lower Yangtze region in the seventeenth and eighteenth centuries. In B. A. Elman & A. Woodside (Eds.), *Education and society in late imperial China 1600-1900* (pp. 381-416). Berkeley, CA: University of California Press.
- Li, Q., & Wang, X. (2003). The implication of middle school English education reform to College English teaching. *Foreign Language Teaching and Research*, 5(2).
- Liu, H., Tian, J., Zhang, Y., & Zheng, R. (2002). *The history of examination in China*. Wuhan, Hubei, China: The Central China Normal University Press.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modelling. *Psychological Methods*, 1(2), 130-149.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111(3), 490-504.
- MacCallum, R. C., Roznowski, M., & Reith, J. V. (1994). Alternative strategies for cross-validation of covariance structure models. *Multivariate Behavioral Research*, 29(1), 1-32.
- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44(2), 283-305.
- MacIntyre, P. D., Mackinnon, S. P., & Clement, R. (2009). The baby, the bathwater, and the future of language learning motivation research. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 43-65). Bristol, UK: Multilingual Matters.
- Maehr, M. L., & Midgley, C. (1991). Enhancing student motivation: A school-wide approach. *Educational Psychologist*, 26(3-4), 399-427.
- Malallah, S. (2000). English in an Arabic environment: Current attitudes to English among Kuwait University students. *International Journal of Bilingual Education and Bilingualism*, 3(1), 19-43.
- Marcoulides, G. A., & Schumacker, R. E. (1996). Introduction. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modelling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Markus, H., & Nurius, P. (1986). Possible Selves. *American Psychologist*, 41(9), 954-969. doi: 10.1037//0003-066x.41.9.954
- Marshall, G. (1994). *The concise Oxford dictionary of sociology*. Oxford, UK: Oxford University Press.



- Mason, J. (1996). *Qualitative researching*. London, UK: SAGE Publications Ltd.
- McArdele, J. J., & Bell, R. Q. (2000). An introduction to latent growth models for developmental data analysis. In T. D. Little, K. U. Schnabel & J. Baumert (Eds.), *Modelling longitudinal and multilevel data: Practical issues, applied approaches and specific examples* (pp. 69-107). Mahwah, NJ: Lawrence Erlbaum Associates.
- McKenzie, R. M. (2004). Attitudes of Japanese nationals towards standard and non-standard varieties of Scottish English speech. *The East Asian Learner*, 1, 17-27. Retrieved from <https://www.northumbria.ac.uk/media/3744496/mckenzie-2004-east-asian-learner-publication.pdf>
- McKenzie, R. M. (2008). Social factors and non-native attitudes towards varieties of spoken English: A Japanese case study. *International Journal of Applied Linguistics*, 18(1), 63-88.
- McKenzie, R. M. (2010). *The social psychology of English as a global language: Attitudes, awareness and identity in the Japanese context*. New York, NY: Springer.
- McNamara, T. (2000). *Language testing*. Oxford, UK: Oxford University Press.
- McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, 3(1), 31-51.
- McNamara, T. (2008). The socio-political and power dimensions of tests. In N. H. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 7, pp. 415-428). Boston, MA: Springer.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.
- McNamara, T., & Ryan, K. (2011). Fairness versus justice in language testing: The place of English literacy in the Australian citizenship test. *Language Assessment Quarterly*, 8(2), 161-179.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: American Council on Education & Macmillan.
- Messick, S. (1993). Trait equivalence as construct validity of score interpretation across multiple methods of measurement. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 61-73). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(1), 243-256.



- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156-166.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. London, UK: SAGE Publications Ltd.
- Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement*, 25(3), 707-726.
- Ministry of Education. (2004). *College English curriculum requirements (trial implementation)*. Shanghai: Shanghai Foreign Language Education Press.
- Ministry of Education. (2007). *College English Curriculum Requirements*. Shanghai: Shanghai Foreign Language Education Press.
- Mosier, C. I. (1947). A critical examination of the concepts of face validity. *Educational and Psychological Measurement*, 7(2), 191-205.
- Murry, J. C., Riazi, A. M., & Cross, J. L. (2012). Test candidates' attitudes and their relationship to demographic and experiential variables: The case of overseas trained teachers in NSW, Australia. *Language Testing*, 29(4), 577-595.
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189.
- National Center for Education Statistics. (2003). *NCES statistical standards*. Washington, DC.: Department of Education.
- National College English Test Committee. (2006). *CET4 test syllabus*. Shanghai: Shanghai Foreign Language Education Press.
- Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modelling. *Structural Equation Modelling*, 8(3), 353-377.
- On, L. W. (1996). The cultural context for Chinese learners: Conceptions of learning in the Confucian tradition. In D. A. Watkins & J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences*. Hong Kong, China: CERC and ACER.
- Ormrod, J. E. (2004). *Human learning* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Oxford, R. L. (2002). Sources of variation in language learning. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 245-252). Oxford, UK: Oxford University Press.



- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). Berkshire, UK: Open University Press.
- Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System*, 38(3), 467-479. doi: 10.1016/j.system.2010.06.011
- Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (2000). Developing disillusionment: Students' perceptions of academic achievement tests. *Issues in Education*, 1(2), 17-46.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: SAGE Publication, Inc.
- Pearson, I. (1988). Tests as levers for change. In D. Chapm-Berland & R. Baumgartner (Eds.), *ESP in the classroom: Practice and evaluation* (pp. 98-107). London, UK: Modern English Publications.
- Peng, J. (2007). Willingness to communicate in the Chinese EFL classroom: A cultural perspective. In J. Liu (Ed.), *English language teaching in China: New approaches, perspectives and standards*. London, UK: Continuum International Publishing Group.
- Phakiti, A. (2003). *An empirical investigation into the relationships of state-trait strategy use of EFL reading comprehension test performance: A structural equation modeling approach* (Unpublished doctoral dissertation). The University of Melbourne, Australia.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory and research. *Contemporary Educational Psychology*, 25(1), 92-104.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *The Phi Delta Kappa*, 68(9), 679-682.
- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, 16(2), 1-23.
- Rasti, I. (2009). Iranian candidates' attitudes towards IELTS. *Asian EFL Journal*, 11(3), 110-155.
- Riazi, A. M. (2016). Innovative mixed-methods research: Moving beyond design technicalities to epistemological and methodological realisations. *Applied Linguistics*, 37(1), 33-49.
- Riazi, A. M., & Candlin, C. N. (2014). Mixed methods research in language teaching and learning: Opportunities, issues and challenges. *Language Teaching*, 47(2), 135-173.
- Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge, UK: Cambridge University Press.



- Roulston, K. (2014). Analysing interviews. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 297-312). London, UK: SAGE.
- Rowley, G. L. (1974). Which examinees are favoured by the use of multiple choice tests? *Journal of Educational Measurement*, 11(1), 15-23.
- Ryan, S. (2009). Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English. In E. Ushioda & Z. Dörnyei (Eds.), *Motivation, language identity and the L2 Self* (pp. 120-144). Bristol, UK: Multilingual Matters.
- Saif, S. (2006). Aiming for positive washback: A case study of International teaching assistants. *Language Testing*, 23(1), 1-34.
- Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. *Review of Educational Research*, 49(2), 252-279.
- Saville, N. (2009). *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider* (Unpublished doctoral dissertation). University of Beffordshire, UK.
- Saville, N. (2012). Applying a model for investigating the impact of language assessment within educational contexts: The Cambridge ESOL approach. *Cambridge ESOL: Research notes*, 50, 4-7.
- Schmitt, N. (2002). Do reactions to tests produce changes in the construct measured? *Multivariate Behavioural Research*, 37(1), 105-126.
- Schreier, M. (2014). Qualitative content analysis. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 170-183). London, UK: SAGE.
- Schumann, J. H. (2015). Foreword. In Z. Dörnyei, P. D. MacIntyre & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 11-14). Bristol, UK: Multilingual Matters.
- Secolsky, C. (1987). On the direct measurement of face validity: A comment on Nevo. *Journal of Educational Measurement*, 24(1), 82-83.
- Seidel, J., & Kelle, U. (1995). Different functions of coding in the analysis of textual data. In U. Kelle (Ed.), *Computer-aided qualitative data analysis: Theory, methods and practice* (pp. 52-61). London, UK: SAGE Publications Ltd.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London, UK: Pearson.
- Shohamy, E. (2006). *Language policy: Hidden agendas and new approaches*. New York, NY: Routledge.



- Spolsky, B. (1989). *Conditions for second language learning: Introduction to a general theory*. Oxford, UK: Oxford University Press.
- Stake, R. (1995). *The art of case study research*. London, UK: SAGE Publications Ltd.
- Starks, D., & Paltridge, B. (1996). A note on using sociolinguistic methods to study non-native attitudes towards English. *World Englishes*, 15(2), 217-224.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stricker, L. J., Wilder, G. Z., & Bridgeman, B. (2006). Test takers' attitudes and beliefs about the Graduate Management Admission Test. *International Journal of Testing*, 6(3), 255-268.
- Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In E. Ushioda & Z. Dörnyei (Eds.), *Motivation, language identity and the L2 self* (pp. 66-98). Bristol, UK: Multilingual Matters.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed method research*. Thousands Oaks, CA: SAGE.
- Thelen, E., & Smith, L. B. (1994). *A dynamic system approach to the development of cognition and action*. Cambridge, MA: The MIT Press.
- Thompson, B. (2000). Ten commandments of structural equation modelling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 261-283). Washington, DC: American Psychological Association.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Ushioda, E., & Dörnyei, Z. (2009). Motivation, language identities and the L2 self: A theoretical overview. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 1-8). Bristol, UK: Multilingual Matters.
- Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473-481.
- Valette, R. A. (1967). *Modern language testing*. New York, NY: Harcourt, Brace and World.



- Van Geert, P. (2007). Dynamic systems in second language learning: Some general methodological reflections. *Bilingualism: Language and Cognition*, 10(1), 47-49.
- Wall, D. (1997). Impact and washback in language testing. In C. M. Clapham & D. Corson (Eds.), *Language testing and assessment: Encyclopedia of language and education* (Vol. 7, pp. 291-302). Dordrecht, the Netherlands: Kluwer.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- Wang, H. (2008). A system approach to the reform of College English Testing - Reports of the survey of College English Reform. *Foreign Languages in China*, 4, 4-12. Retrieved from <http://ncpssd.org/Literature/readurl.aspx?id=4000044227&type=1>
- Wang, J. (2003). *The history of education in China [中国教育史]*. Guangzhou, Guangdong, China: Guangzhou Higher Education Press.
- Wang, X., & Cheng, L. (2010). Chinese EFL students' perceptions of the classroom assessment environment and their goal orientations. In L. Cheng & A. Curtis (Eds.), *English language assessment and Chinese learner* (pp. 202-218). New York, NY: Routledge.
- Warden, C. A., & Lin, H. (2000). Existence of integrative motivation in an Asian EFL setting. *Foreign Language Annals*, 33(5), 535-545.
- Watanabe, Y. (1996). Investigating washback in Japanese EFL classrooms: Problems of methodology. In G. Wigglesworth & C. Elder (Eds.), *The language testing circle: From inception to washback* (pp. 208-239). Melbourne, Australia: Applied Linguistics Association of Australia.
- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Y. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129-147). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Wesche, M. B., & Skehan, P. (2002). Communicative, task-based, and content-based language instruction. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 207-288). Oxford, UK: Oxford University Press.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables. In R. H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues and applications* (pp. 56-75). Thousand Oaks, CA: SAGE.
- William, D. (1996). National curriculum assessments and programmes of study: Validity and impact. *British Educational Journal*, 22(1), 129-141.
- Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals. *American Psychologist*, 54(8), 594-604.



- Wilkinson, S. (1998). Focus groups in feminist research: Power, interaction, and the co-production of meaning. *Women's Studies International Forum*, 21(1), 111-125.
- Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *The Modern Language Journal*, 97(1), 109-130.
- Winke, P. (2014). Testing hypotheses about language learning using structural equation modelling. *Annual Review of Applied Linguistics*, 34, 102-122.
- Woodside, A., & Elman, B. A. (1994). The expansion of education in Ch'ing China. In B. A. Elman & A. Woodside (Eds.), *Education and society in late imperial China 1600-1900* (pp. 525-560). Berkeley, CA: University of California Press.
- Wu, S. (2003). College English teaching: Problems and solutions. *Foreign Language Teaching and Research*, 3, 226-228. Retrieved from <http://ncpsd.org/Literature/readurl.aspx?id=8018083&type=1>
- Xie, Q. (2011). Is test taker perception of assessment related to construct validity? *International Journal of Testing*, 11(4), 324-348.
- Xie, Q., & Andrews, S. (2012). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49-70.
- Yang, H., & Weir, C. J. (1998). *The CET validation study*. Shanghai: Shanghai Foreign Language Education Press.
- Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *Modern Language Journal*, 86(1), 54-66. doi: 10.1111/1540-4781.00136
- Yashima, T. (2009). International posture and the ideal L2 self in the Japanese EFL context. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 144-163). Bristol, UK: Multilingual Matters.
- Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning*, 54(1), 119-152.
- Yung, Y. F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modelling: Issues and techniques* (pp. 195-226). Mahwah, NJ: Lawrence Erlbaum Associates.
- Zhao, J., & Cheng, L. (2010). Exploring the relationship between Chinese university students' attitudes towards the College English Test and their test performance. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 190-201). New York, NY: Routledge.



Zheng, Y., & Cheng, L. (2008). Test review: College English Test (CET) in China. *Language Testing*, 25(3), 408-417.

Zhou, L. (2003). Further development of the College English teaching. *Foreign Language Teaching and Research*, 3, 224-226. Retrieved from <http://ncpssd.org/Literature/readurl.aspx?id=8018083&type=1>