

RESEARCH

Open Access



The relationship of test takers' learning motivation, attitudes towards the actual test use and test performance of the College English Test in China

Bai Ying

Correspondence: by@baiying.ch
Yantai Research Institute, China
Agricultural University, 2006 Binhai
Road, Laishan District, Yantai,
Shandong Province, China

Abstract

Test takers' individual characteristics may have a significant impact on their test performance (e.g. Bachman & Palmer, *Language assessment in practice: Developing language assessments and justifying their use in the real world*, 2010). However, given their dynamic and context-sensitive nature, different variables of test takers' individual characteristics were explored in previous empirical research and the multivariate relationships among these variables and test performance varied in different research contexts. This study explored the relationships of test takers' English learning motivation, attitudes towards the actual test use of the College English Test (CET), a high-stakes test in China, and their test performance. The findings revealed two distinctive dimensions with regards to English leaning motivation and embedded value implications in test use. Meanwhile, test takers' test performance was positively and negatively affected by their supportive attitudes towards the actual test use of the CET and test-related learning motivation, respectively.

Keywords: Test takers' attitude and English learning motivation, Actual test use in reality, Test performance, The College English Test in China

Introduction

China has witnessed the development of examinations for nearly 2000 years since the Han Dynasty (202 BC–220 AD) (Liu, Tian, Zhang, & Zheng, 2002; J. Wang, 2003). This long history of testing has strongly moulded Chinese people's trust in the value and justice of the examination system. Testing in China generally enjoys a societal acceptance and recognition as a fair and effective method to select talent, promote the development of education, improve the performance of schools and colleges, and even counter nepotism and outright corruption in the allocation of scarce social resources (Bray & Steward, 1998; Cheng, 2010; Cheng & Qi, 2006; Eckstein & Noah, 1992). Due to the powerful forces that testing and examinations have upon Chinese society, tests,

especially high-stakes tests, are sometimes used for purposes beyond that for which they are intended (Cheng, 2010).

The College English Test (CET) is a nation-wide English proficiency test specifically designed for non-English majors in tertiary education in China. The implementation of the CET, according to the CET syllabus (National College English Test Committee, 2006, p. 3), aims to “accurately examine the English proficiency of undergraduate students in China in accordance with the language skills specified in the College English Curriculum Requirements (CECR), and to promote effective College English teaching and learning”. However, reality has demonstrated a more complex picture. As a large-scale high-stakes exam in China, the CET and its subsequent test result has been widely used in practice to make decisions for multiple unintended purposes. Test takers’ CET results are used by employers of various professions as a criterion for selection (Garner & Huang, 2014; Xie, Han, Lin, & Sun, 2007; Yang & Weir, 1998). The CET test result is a prerequisite for many students to obtain their bachelor degree or even to graduate from Chinese universities (e.g. Garner & Huang, 2014; Huang, 2002; Li & Wang, 2003; H. Wang, 2008; Wu, 2003; Zhou, 2003), and in some of the most developed cities in China (e.g. Beijing, Shanghai, Guangzhou and Shenzhen), this certificate is even named as one criterion for issuing a residence permit (Garner & Huang, 2014; Jin, 2008). The CET test results, under the influence of a historically rooted testing culture, was regarded as a kind of cultural capital representing opportunities to compete for scarce social resources, change lives and win success (Garner & Huang, 2014). Even though the NCETC argues that the CET result provides an objective assessment of test takers’ English competence and thus explains why it has been widely used as a criterion for employment, many of the high-stakes decisions on test takers are considered to be misuse of the test results (Jin, 2008).

Test takers are one group of the most directly affected stakeholders of the consequences of using a test and of the decisions made about them on the basis of test results (Bachman & Palmer, 2010). Their attitudes and perceptions are considered as a valuable source of evidence in re-evaluating the promoted standards and intended results and in discovering any unintended and unpredicted occurrences in practice (Bachman & Palmer, 2010; Davies, 2008; Hawkey, 2006; Saville, 2012). Given their unique position in testing, test takers’ attitudes and perceptions have also been recognised as a potential construct-irrelevant factor that may either positively or negatively affect test performance (e.g. Bachman & Palmer, 2010). Until now, however, the questions of how test takers perceive the actual test use of the CET in reality, including unintended test use by test design, and how this attitudinal factor affects test performance have not been explored.

Literature review

Test takers’ attitude, language learning motivation and performance

The key tenet in social psychological theories of action (Ajzen, 2005; Eagly & Chaiken, 1993) is the assumption that it is attitudes that exert a directive influence on people’s behaviour, since people’s attitude towards a target influences the overall pattern of their responses to the target (Gardner, 1985). The term “attitude” in this study refers to an affective and evaluative response to an object, institution or event, elicited from

subjective views or answers to a number of questions (Ajzen, 1988, 2005; Allport, 1971; Spolsky, 1989). An attitude in this definition is a hypothetical construct, which may not be directly observable but can be inferred from observable responses (Eagly & Chaiken, 1993). Language learning motivation is a widely explored complex issue in Second Language Acquisition (SLA). Located within the discipline of social psychology in attitude theory, language learning motivation provides the driving force to initiate second language learning, sustain the tedious learning process and subsequently affect success in learning language (Dörnyei, 2005). The socio-educational (SE) model of SLA (Gardner, 1985; Gardner & Lambert, 1972), created and applied in the Canadian context, has for decades influenced international conceptualisations of motivation for second language learning. As English has gradually spread throughout the world as an international lingua franca, a more substantial reconceptualisation of the L2 motivation construct is Dörnyei's (2005, 2009) L2 Motivational Self System, established on the basis of the understandings acquired from the SE Model, the concepts from "Possible Selves" (Markus & Nurius, 1986) and "Self-Discrepancy" Theory (Higgins, 1987), and his own empirical motivation research (Dörnyei & Csizér, 2002; Dörnyei, Csizér, & Németh, 2006). The most recent development on motivational theories have introduced motivational dynamics into language learning (e.g. Dörnyei, MacIntyre, & Henry, 2015), highlighting the point that learners' motivation is dynamic and context sensitive, and the exploration of motivators in a particular research context requires "a localised, scientific, research-based approach" (Chen, Warden, & Chang, 2005, p. 626).

Some Chinese researchers proposed a concept of Required Motivation or Chinese Imperative in an attempt to capture the cultural element in a Chinese learning context (Chen et al., 2005; Warden & Lin, 2000). They noted that most motivation theories were established on the basis of North American and European cultural values, which emphasised individualism rather than collectivism commonly upheld in Chinese Confucian culture. This concept highlighted the culturally internalised requirements, emphasising on the influence of the imperial examination system in history as well as Confucius collectivism in that students often felt obliged to obtain a good test result to win success for themselves and to bring glory to the whole family (Chen et al., 2005; Leung, 1994; Warden & Lin, 2000).

When dealing with individual differences concerning attitudes and motivation, however, one problematic issue is the cause and effect sequence between the two: There is no unequivocal generalisations as to which are the cause, and which are the effect (Gardner & Clement, 1990; Lambert, 1963, 1967). Xie (2011) in exploring the relationship between students' perceptions of examined skills and CET test performance noted a significant impact of students' instrumental learning motivation on their perceptions, whilst in Wu and Lee's study (J. Wu & Lee, 2017), students' attitudes towards the policy of English benchmark requirements for graduation in Taiwan universities positively affected both of their extrinsic and intrinsic English learning motivations.

A major concern in the design and development of language tests is to minimise 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 relationship between learners' language attitudes, motivations and their language achievement in the socio-psychological approach of SLA has been a focal point in research, and the relationship among these factors has been, to a certain extent, established in empirical studies in SLA (e.g. McKenzie, 2008). However, what are test takers' motivational construct in a test-oriented learning environment, whether and how test takers' learning motivation and attitudes interact with their test performance and what attitude variables may significantly affect their test performance are greatly under-investigated in language testing (e.g. J. Wu & Lee, 2017). Bachman (1990) specified four factors that influence language test performance, one of which is test takers' characteristics or personal attributes such as cultural background, background knowledge, cognitive abilities, sex and age. In Bachman's later work (Bachman & Palmer, 2010), test takers' affective responses to test content and test tasks were explicitly highlighted as such a potential factor. Following Bachman's guide, Kunnan (1995) particularly explored the relationship of test takers' characteristics and their test performance using Structural Equation Modelling (SEM). Test takers' characteristics in his empirical study integrated factors such as cultural background, exposure to the target language which were stressed in the model of language testing (Bachman, 1990) and monitoring (Krashen, 1981) and attitude (Gardner, 1985) in language learning models in SLA. Among other attitude-test performance studies in language testing, test takers' attitude factors included perceptions of examined abilities (Stricker, Wilder, & Bridgeman, 2006), quality of each test components (Rasti, 2009), test bias, test-taking motivation and anxiety (Zhao & Cheng, 2010) and testing policy (J. Wu & Lee, 2017) to name just a few. The findings of these studies demonstrated that the relationships between test performance and test takers' individual characteristics varied in different research contexts with different attitude factors. To date, there is still a lack of theoretical guidance on what aspects of a test that test takers' attitudes should be investigated to examine the possible interaction with test performance. Test use, until now, has not been explored as an attitudinal factor.

Test use

Test use refers to the use of test-score-based interpretations to make decisions about the stakeholders (Bachman & Palmer, 2010), in most cases, the test takers. To refer to a test or test score as valid, the fundamental claim should be built upon "the specific ability or abilities the test is designed to measure and the uses for which the test is intended" (Bachman, 1990, p. 238). However, the actual test interpretation and test use at local levels are invariably more complex, varied, nuanced than intended, usually shaped and decided by local purposes (Bachman & Palmer, 2010; Moss, 2016; Saville, 2012). The construct of the test could be re-interpreted by test users as evidence of hard work, discipline, intelligence or other valued qualities relevant to these decisions being made about the test takers, rather than that of measured abilities (Akiyama, 2004; McNamara & Roever, 2006); test-based information could also be used in allocating resources, setting new educational targets or goals and improving administrative management (Moss, 2016). Even though there are still arguments over definitions of test validity and the critical evidence to gather for test validation, modern validity theories agree that test uses and test consequences cannot be ignored and are an indispensable

component of validity discussions (Cizek, 2016; Moss, 2016; Newton & Baird, 2016; Sir-eci, 2016). One major concern of test validity is that even though tests/assessments are valid indicators of the abilities they are intended to measure, the test results are used inappropriately and decisions based on test results are irrelevant to the measured abilities.

When tests are not used as designed, how would test takers, one of the most directly affected stakeholders, perceive the score-based decisions made about them? Given their unique position in testing, would their attitudes towards this consequential aspect of tests affect test performance and hence become a potential construct-irrelevant factor?

Research questions and explored relationships

This study is part of a bigger project which explores the multivariate relationships of test takers' individual characteristics, attitudes towards the CET and test performance, in an attempt to shed light on the complexity and dynamism of test takers in testing. This study focuses on reporting the construct of test takers' English learning motivation in preparation for the CET, their attitudes towards and perceptions of the actual test use of the CET in practice and their potential interactions with test performance. The research questions (RQ) are as follows:

1. What is the structure of test takers' motivational construct in preparation for the CET?
2. What is the structure of test takers' attitude towards the actual test use of the CET in practice?
3. What are the relationships between test takers' English learning motivation and their attitudes towards the actual test use of the CET? How do these individual characteristics affect test performance?

Methods

The process of test justification is specific to each particular testing context (Bachman & Palmer, 2010). This study, therefore, adopted the Exploratory Sequential Design (ESD), a hybrid research design, to facilitate the understanding of local factors and attempt to capture their complex and dynamic interplay in the research context (Creswell, 2014). The ESD is a two-phase mixed methods design, which starts with qualitative data for an exploratory purpose and follows up with quantitative data in order to generalise results within a population. Test takers' motivation and attitude, given their dynamic and context-sensitive nature, were first collected in the qualitative phase using semi-structured focus group interviews, face-to-face interviews and eMail interviews. The themes revealed in the qualitative analysis were then integrated into a questionnaire, the data collection instrument for the quantitative phase, using the same wordings whenever appropriate. Given the word limit, this study only reports the relevant results of quantitative analysis in answering the raised research questions.

Quantitative data collection

The questionnaire, a 5-point Likert scale of agreement, consisted of two major parts, with part I collecting test takers' demographic information including name, department,

university and CET test result, which was applied as the factor of test performance in examining the hypothesised relationships, and part II, measuring test takers' characteristics such as English learning motivation and attitudes towards the actual test uses of the CET (see Appendix). The questionnaire was first piloted among five colleagues and a targeted sample group of 10 students. The items that generated ambiguity and misunderstanding were reworded and modified before the actual data collection was conducted.

The quantitative data (same as qualitative data) were collected at two universities in China, which were selected due to data accessibility and financial practicality of this study. The universities are of similar characteristics in terms of their size, number of students, recruitment procedure and financial standing. Both of the universities are top-ranking universities directly under the Ministry of Education (MOE) and located in the eastern provinces, culturally and economically more developed areas in China. The questionnaire was administered both online and in hardcopy form to maximise participant accessibility. A total of 369 participants submitted a completed questionnaire, of which, 11 were detected as outliers and thus deleted from the data set, making a total of 358 valid questionnaire responses for the subsequent quantitative analysis (42.5% female; 57.5% male).

Quantitative analysis

The quantitative data was analysed using SEM. The programmes SPSS 20 and AMOS 22 were applied in this phase. A number of item-level pre-test evaluations were first conducted, examining how well-identified items elicited from qualitative analysis represent their associated latent constructs, i.e. test takers motivation and attitude in this study. The pre-examinations included item descriptive statistics, data distribution, exploratory factor analysis (EFA) and item internal consistency (e.g. Cronbach's Alpha). The detailed item-level analyses are not included in the result section due to the word limit.

Based on results of the EFA and reliability analysis, a confirmatory factor analysis (CFA) was then conducted, evaluating the factorial validity of each latent construct. The maximum likelihood (ML) estimation was applied in the model estimations. The evaluation of the model adequacy was based on an inspection of the values of standardised residuals, the chi-square statistics, other fit indices (e.g. CFI, GFI, SRMR & RMSEA) and theoretical and conceptual aspects of constructs under study (Hair, Black, Babin, Anderson, & Tatham, 2006). Post hoc model re-specification and re-estimation based on the assessment of model fit indices, the standardised residual covariance matrix and modification indices, were conducted when necessary, to identify the most adequate representation of the factorial structure of each latent construct. The squared multiple correlations (SMC or R^2), which determines the amount of variance accounted for in each dependent variable by the predictor variables (Kline, 2011) was also reported at this stage.

The causal validity of hypothesised structural relationships were then examined, integrating the results derived from the CFA of measurement models (i.e. test takers' motivation and attitude). The significance of the hypothesised relationships, i.e. whether the hypothesised relationships were supported by the data or not; the direction, i.e.

whether the hypothesised relationships were positive or negative; and effect sizes, including SMC and Cohen's f^2 , were evaluated at this stage.

Moreover, prior to the examination of the causal validity, the multivariate normality for each model was assessed for a reliable ML estimation. Based on the findings of studies which particularly compared and explored the acceptable range of multivariate non-normalities for ML estimation to be appropriate (Bentler, 2005; Gao, Mokhtarian, & Johnston, 2008; Harlow, 1985; Lei & Lomax, 2005; Nevitt & Hancock, 2001), when the multivariate kurtosis value is no larger than 28.76, the ML estimation still produces a trustworthy analysis. When the multivariate kurtosis value was greater than the acceptable range (28.76), bootstrap estimates were then performed on 500 samples using the ML estimator, to provide bias-corrected confidence intervals (95%) for each of the parameter bootstrap estimates (Byrne, 2010).

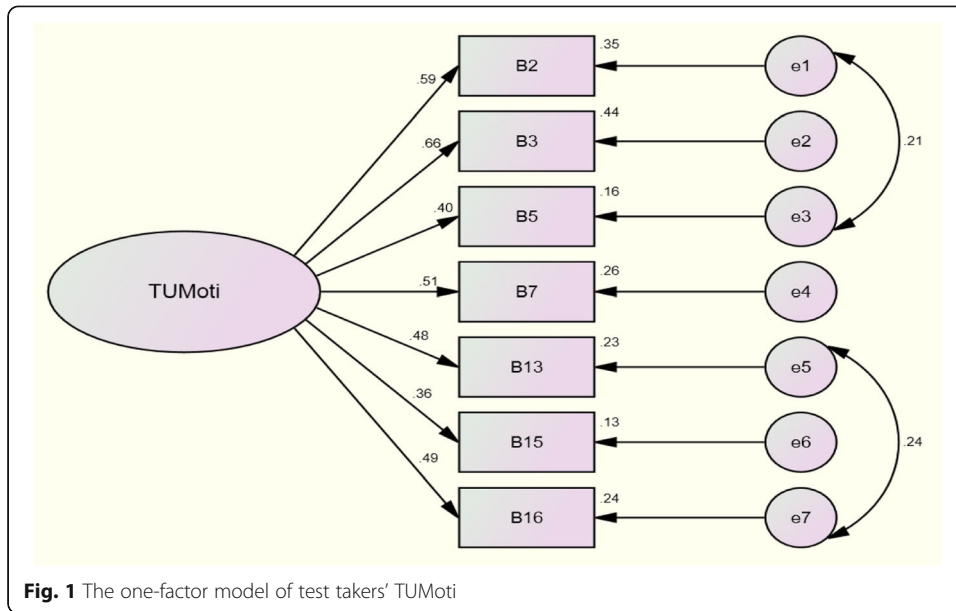
Results

Examining the structure of test takers' motivational construct in preparation for the CET

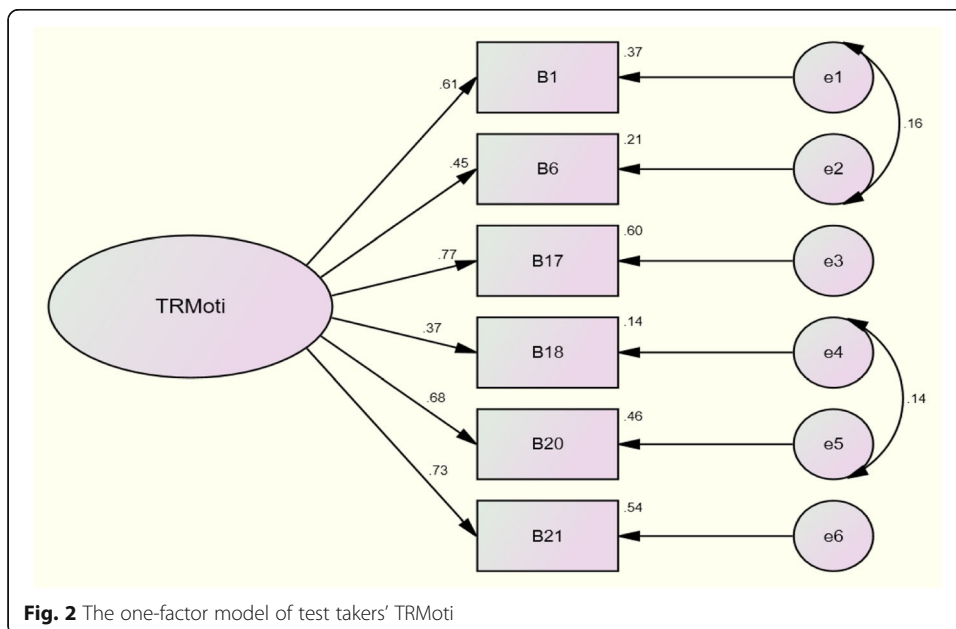
Test takers' English learning motivation revealed in the qualitative analysis, and later identified in EFA as well, demonstrated two distinctive dimensions: test-unrelated English learning motivation (TUMoti) and test-related English learning motivation (TRMoti).

Test takers' TUMoti comprised 7 items (items B2, B3, B5, B7, B13, B15 and B16, Cronbach's Alpha = 0.699). The Mardia's multivariate kurtosis value was 20.618 with an associated critical ratio (C.R.) of 17.377, suggesting an acceptable multivariate non-normality in the sample, supportive of a sound ML estimation (e.g. Gao et al., 2008). Based on the post hoc analyses, error terms associated with items B2 and B5; B13 and B16 were correlated as a free parameter (see Bentler & Chou, 1987 for details with regards to error correlation). The confirmatory examination of the factorial validity of the TUMoti, with error terms associated with items B2 and B5 and B13 and B16 correlated as a free parameter, demonstrated a good model fit: $\chi^2_{(12)} = 13.280$, $p = 0.349$; CFI = 0.996; GFI = 0.990; SRMR = 0.027; RMSEA = 0.017 (0.000–0.058, 90%CI), suggesting an appropriate representation of its factorial structure. The unstandardised estimates were within an admissible range and statistically significant (with C.R. $> \pm 1.96$ and $p < 0.001$); all standard errors appeared to be in good order as well. The loadings in the standardised solution ranged from 0.359 for item B15 to 0.660 for item B3. All factor loadings (i.e. estimate values) were statistically significant at the 0.05 level. The SMC (R^2) ranged from 0.129 (B15) to 0.436 (B3). Figure 1 provides a diagrammatic representation of this one-factor model.

Test takers' TRMoti comprised 6 items (items B1, B6, B17, B18, B20 and B21; Cronbach's Alpha = 0.776). The Mardia's multivariate kurtosis value (6.344, with an associated C.R. of 6.125) suggested an acceptable non-normality in the sample, supportive of a trustworthy ML estimation. The factorial validity of the TRMoti in CFA, with error terms associated with items B1 and B6 and B18 and B20 correlated as a free parameter (Bentler & Chou, 1987), demonstrated a good model fit: $\chi^2_{(7)} = 9.646$, $p = 0.210$; CFI = 0.995; GFI = 0.991; SRMR = 0.021; RMSEA = 0.033 (0.000–0.078, 90%CI), suggesting an appropriate representation of its factorial structure. The unstandardised estimates



were within an admissible range and statistically significant (with C.R. $> \pm 1.96$ and $p < 0.001$); all standard errors appeared to be in good order as well. The loadings in the standardised solution ranged from 0.374 for item B18 to 0.775 for item B17. All factor loadings were statistically significant at the 0.05 level. The SMC (R^2) values ranged from 0.140 (B18) to 0.600 (B17). Figure 2 provides a diagrammatic representation of this one-factor model.



Examining the structure of test takers’ attitude towards the actual test use of the CET in practice

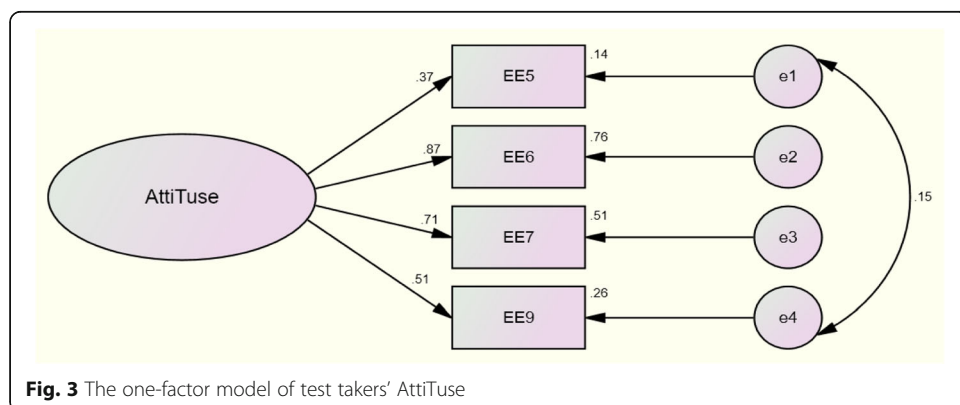
Test takers’ attitudes towards the actual test use of the CET (AttiTuse) were composed of 4 items (items EE5, EE6, EE7 and EE9; Cronbach’s Alpha = 0.713), portraying test takers’ perceived societal values embedded in the actual test use of the CET. The Mardia’s multivariate kurtosis value (1.579 with an associated C.R. of 2.155) was within an acceptable range, suggesting a multivariate normal distribution in the data, supportive of a trustworthy ML estimation. The test of the one-factor model of test takers’ AttiTuse, with errors associated with EE5 and EE9 correlated as a free parameter (Bentler & Chou, 1987), demonstrated a good model fit: $\chi^2_{(1)} = 0.499$, $p = 0.480$; CFI = 1.000; GFI = 0.999; SRMR = 0.007; RMSEA = 0.000 (0.000–0.124, 90%CI), suggesting an appropriate representation of its factorial structure. A review of the unstandardised solutions revealed that all estimates were within an admissible range and statistically significant (with C.R. > ± 1.96 and $p < 0.001$); all standard errors appeared to be in good order as well. The loadings in the standardised solution ranged from 0.368 for item EE5 to 0.872 for item EE6. All factor loadings were statistically significant at the 0.05 level. The SMC (R^2) ranged from 0.136 (EE5) to 0.761 (EE6). Figure 3 provides a diagrammatic representation of this one-factor model.

Examining the hypothesised structural relationships

Among the 358 participants in the quantitative phase, the highest CET test score (TS) was 613 and the lowest, 310. Both skewness and kurtosis values were within a good range (smaller than |1|), indicating a normal distribution in the data.

The findings derived from the measurement model evaluation of the AttiTuse, TRMoti and TUMoti were then integrated into the examination of the hypothesised structural relationships. Figure 4 depicts the hypothesised relationships of test takers’ learning motivations, attitude and their test performance examined at this stage, including:

H1: Test takers’ test-unrelated English learning motivation affects their attitudes towards the actual test use of the CET and vice versa.



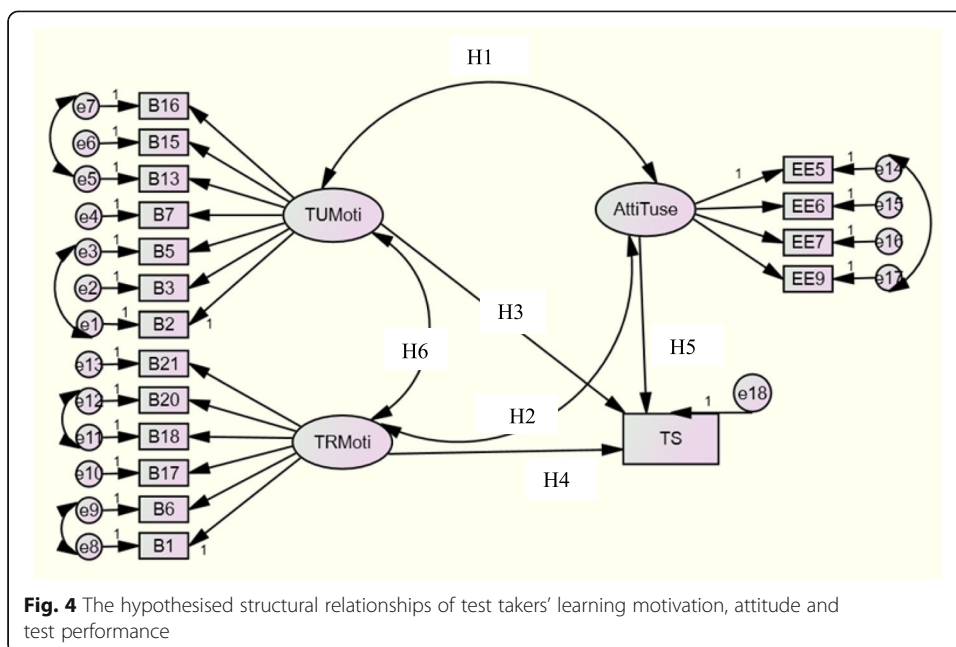


Fig. 4 The hypothesised structural relationships of test takers’ learning motivation, attitude and test performance

H2: Test takers’ test-related English learning motivation affects their attitudes towards the actual test use of the CET and vice versa.

H3: Test takers’ test-unrelated learning motivation affected their test performance.

H4: Test takers’ test-related learning motivation affected their test performance.

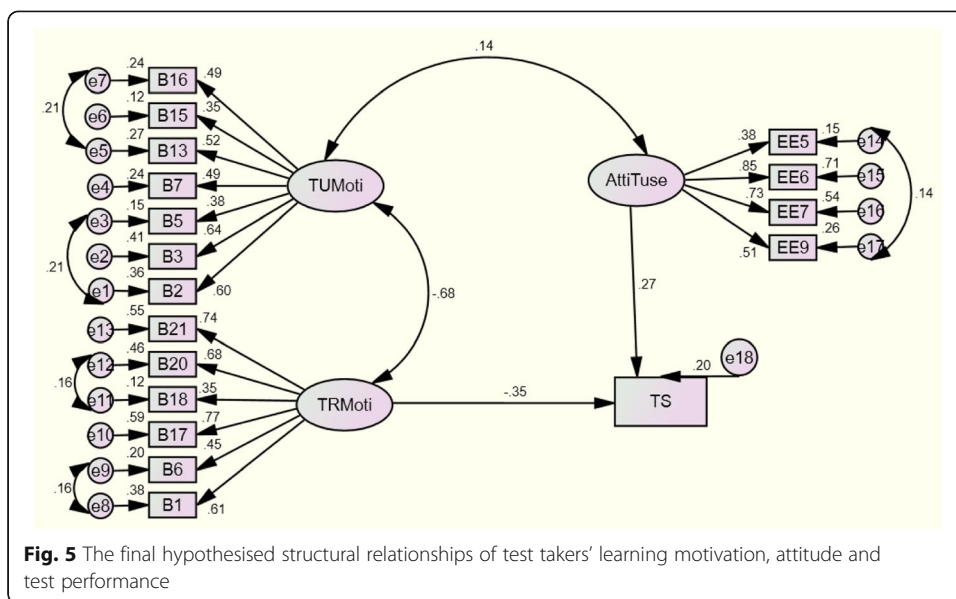
H5: Test takers’ attitudes towards the actual test use of the CET affected their test performance.

H6: Test takers’ learning motivations were interrelated with each other.

The Mardia’s multivariate kurtosis value for the hypothesised model was 52.236, with an associated C.R. of 18.417, suggesting a multivariate non-normality in the sample. A bootstrap analysis was then conducted, the results of which indicated the multivariate non-normality was still at an acceptable level to support a sound ML estimation (Byrne, 2010). All the fit indices were within a desirable range: $\chi^2_{(125)} = 155.373, p = 0.034$; CFI = 0.978; GFI = 0.956; SRMR = 0.044; RMSEA = 0.026 (0.008–0.039, 90%CI). An examination of the feasibility and the statistical significance of parameter estimates indicated that all estimates were reasonable and statistically significant at the 0.05 level. The SMC¹ value for TS was 0.200, indicating 20% of test takers’ test performance was explained by the constructs of AttiTuse and TRMoti (Cohen’s $f^2 = 0.250$, medium effect size).

In reviewing the significance of each hypothesised relationship, however, two of the hypothesised structural relationships, paths H2 and H3, were not supported by the sample, both falling out of the significant range (with C.R. < ± 1.96 and $p > 0.05$) and were then removed from the model. As a result, four out of six estimates were

¹The squared multiple correlations (SMC or R^2) in a structural model represent the proportion of each dependence construct that is explained by the predictors or independence constructs in question (Byrne, 2010; Hair et al., 2006).



consistent with the hypothesis. Figure 5 provides a diagrammatic representation of the modified model.

The unstandardised, standardised total effects and correlation estimate of the hypothesised relationships are presented in Table 1.

Discussion

RQ1: What is the structure of test takers' motivational construct in preparation for the CET?

Test takers' English learning motivation in this research context revealed two distinctive dimensions: test-unrelated English learning motivation (TUMoti) and test-related English learning motivation (TRMoti). The items in test takers' TUMoti were closely in line with the L2 Ideal Self in Dörnyei's L2 Motivational Self System (2005, 2009). Test

Table 1 The unstandardised, standardised effects and correlation estimates of hypothesised relationships

DR	UnSt TE	SE	C.R.	St TE	St DE
AttiTuse→TS	37.953	9.241	4.107	0.274	0.274
TRMoti→TS	- 29.478	4.940	- 5.968	- 0.353	- 0.353
IR	CE	SE	C.R.	r	
TUMoti↔TRMoti	- 0.215	0.034	- 6.399	- 0.676	
TUMoti↔AttiTuse	0.027	0.012	2.161	0.141	
DC	SMC or R ²		^a Cohen's f ² (f ² = R ² /1 - R ²)		
TS	0.200		0.250		

DR dependence relationships, UnSt TE unstandardised total effects, SE standard error, C.R. critical ration, St TE standardised total effects, St DE standardised direct effects, IR interrelationship, CE covariance estimate, r correlation estimate, DC dependence constructs, SMC squared multiple correlations, TUMoti test-unrelated motivation, TRMoti test-related motivation, AttiTuse attitudes towards actual test use of the CET, TS test score (test performance)
^af² = 0.02 as small effect size; f² = 0.15 as medium effect size; f² = 0.35 as large effect size

takers/students were inspired to work hard on English to achieve their goal of an ideal image (e.g. B7: It is cool to speak fluent English) or motivated to learn English because they had a good understanding that English as a global language would provide them with a broader range of options and opportunities for either their academic or professional development in the future (e.g. B16: Learning English will help me to be more competitive in the future). The components explaining Yashima's International Posture (2002, 2009) were also identified in this research context. Test takers/students expressed their desires in travelling and studying abroad and making friends with people from different cultures (e.g. B2: I would like to communicate and make friends with foreigners). Test takers' TUMoti on the whole is promotive in nature, positively motivating test takers/students to learn English.

Test takers' TRMoti appeared to have been greatly affected by their test-related learning experiences (e.g. B20: Test-oriented English learning in high school discouraged my English studies), high-stakes decisions attached to the CET test results, opportunities generated by test use (e.g. B1: I have to learn English to pass the test), and subsequent competition among peer groups (e.g. B18: I am always nervous in the English class). Test takers were well aware of the influence of the test results on their future lives. They felt strongly obliged to obtain a good test score to win success for themselves, but the internalised obligation to bring glory for the whole family, as highlighted in Required Motivation or Chinese Imperative (e.g. Chen et al., 2005), was not detected among test takers in this study. The wording of the items in TRMoti, in general, demonstrated a preventive nature, discouraging rather than encouraging test takers to learn.

Test takers' TUMoti was significantly correlated with their TRMoti negatively (Path H6). The correlation coefficient was high ($r = -0.68$, large effect size). This strong negative correlation indicated that when test takers felt compelled/obliged to learn English due to a "mere sense of obligation, duty or a fear of punishment" (Dörnyei et al., 2006, p. 93), or of gaining the benefits and values attached to the tests, they were less encouraged to learn out of sense of self-achievement or self-fulfilment.

RQ3: How does test takers' learning motivation affect their test performance? (H3 and H4)

Test takers' test-related preventive, rather than test-unrelated promotive learning motivation, had a significant, direct and negative effect on their test performance (StTE = -0.35 , $p < 0.05$). Test takers' TRMoti demonstrated two distinctive characteristics: First, it was generally test-related; second, it was preventive rather than promotive in nature. In other words, most of the items used to measure this type of motivation demonstrated a negative and de-motivating characteristic. Chen et al. (2005) expressed concerns over test-related required motivation, claiming that this type of motivation directed students' efforts in learning for tests, rather than meaningful language use abilities, and thus played a role as a de-motivator instead of motivator in English studies. The negative relationship identified in this study appeared to suggest that in addition to language use abilities, this motivation did not play a positive role in boosting test scores either.

Given the power of the CET in making multiple decisions on the basis of its test results, its high stakes, and the subsequent impact on English teaching, learning and on society as a whole, the CET has been widely criticised by scholars and English educators in China (Cai, 2006). The MOE, who is responsible for overseeing the promotion of English teaching and learning at tertiary level in China, has attempted to minimise the influence of the CET on English teaching and learning through softening its decisive role in tertiary education. The NCET C, since the inception of the test, has made a consistent effort to reduce test items examining students' isolated language knowledge, and to increase those that examine English usage, in the hope of directing students' learning motivation and efforts in improving their language use abilities. These modifications and efforts, however, did not appear to have made an immediate progress in achieving their intended goals. Bai (2019) noted that the central language policies at the macro-level were largely subject to the local interpretation and practice. The way in which central language policy was interpreted and implemented at university level had a great impact on English teaching, and subsequently on students' learning motivation and attitudes. When localised central language policy was more competence-oriented, as a result, so was their English teaching, and students in general believed that language use abilities instead of test training should be the focus of English education; when the localised language policy was more test-oriented, students demonstrated a conflicting attitude: They were anxious to obtain a good test score on the one hand and questioned the effectiveness of the teaching approach on the other hand.

Nevertheless, when perceived importance and high value of the test results exists among test takers, it would be inevitable for students/test takers, and even their teachers, to apply whatever strategies they believe effective in achieving a good test score. A more pragmatic question for them is what teaching and learning approaches are beneficial and more effective in improving test performance. When tests have become a focal point in education, as they have in China, apart from criticisms of its negative impact, more in-depth empirical investigations are needed to better inform the related stakeholders how and what to teach and learn in preparation for a test so as to facilitate their competence in competing for this cultural capital.

RQ3: What are the relationships between test takers' English learning motivations and their attitudes towards the actual test use of the CET? (H1 and H2)

Test takers' TRMoti had no significant effect on their AttiTuse and neither did the other way round (with C.R. $< \pm 1.96$ and $p > 0.05$). Test takers' TUMoti was positively correlated with their AttiTuse but the correlation coefficient was low ($r = 0.141$, small effect size), indicating a weak influence between the two factors. This finding was not unexpected given that test takers' attitudes towards test use appeared to be more closely related to their values and beliefs associated with test use, which in this study conformed to the commonly accepted social values of test use by the general public.

RQ2: What is the structure of test takers' attitude towards the actual test use of the CET in practice?

Test takers in general demonstrated a supportive attitude towards the actual test use of the CET (mean = 3.44), including the test uses not intended by test design. The NCETC acknowledges that CET test results have been widely used by businesses in China as one of their major employment criteria, claiming such decisions are based on the test quality, in that the CET is able to examine and successfully predict students' English proficiency level. However, test takers' perceived rationale that interpreted such use of the CET test scores, differed from the official claims made by the NCETC. They believed the CET test use was closely in line with commonly accepted societal values of tests in China: Given the impact of a long-lasting testing culture, tests and test results were regarded by the general public as a fair, effective and practical method for selection in society (e.g. EE5: I believe that it is China's tradition to use test results for selection). They also believed that in present-day of China, using high-stakes test results such as the CET for social selection assisted in achieving fairness and social balance (e.g. EE6: I believe that using test results for the allocation of limited social resources is so far the fairest and most effective method in China; EE7: I believe using test results for selection is an effective method to achieve social balance).

Tests, instead of being unbiased and neutral, are deeply rooted in political, social, educational, ideological and economic contexts (e.g. Messick, 1989; Shohamy, 2001). "Values come into play at all junctures in the testing process" (Cizek, 2016, p. 217). It is acknowledged among many scholars in testing that psychometric evidence is insufficient in justifying test score use validity (e.g. Cizek, 2016). The value systems that inform a particular test use, as well as that of each main stakeholder, cannot be neglected or taken for granted in the process of test validation (e.g. Bachman & Palmer, 2010; McNamara & Roever, 2006). The values and interests of authorities could be expressed in testing at the expense of the values of other major stakeholders (McNamara & Roever, 2006). The findings of this study suggest that when score-based information is not interpreted and used as designed, multiple sources of feedback from major test stakeholders are needed to assist in identifying: (1) What are the underlying value implications that interpret the actual use of the test scores? (2) Do the value implications conform to or conflict with the values and beliefs of major stakeholders such as test takers? An understanding of whether the values embedded in tests and test use are consistent with those of the stakeholders is a critical step in promoting positive test consequences and making fair and equitable decisions on the basis of test results (e.g. Bachman & Palmer, 2010).

RQ3: How does test takers' attitude towards the actual test use of the CET affect their test performance? (H5)

Test takers' supportive attitudes towards the actual test use of the CET had a direct, significant and positive impact on their test performance (StTE = 0.27, $p < 0.05$). Bachman and Palmer (2010, p. 107) claimed that engaging test takers in test development, collecting their perception of the "assessment and assessment

tasks” would facilitate test takers’ perceived authenticity of the test; test takers would hence be more “motivated” and “perform better”. The finding of this study appeared to suggest that test takers’ perceptions of the actual test use in practice could also be identified as a potential factor affecting test performance. Research in language testing rarely includes value implications in a validity discussion; the value implications appear to have been taken for granted in test use, deeming that values embedded in test use naturally conform to the societal values and the values of stakeholders. This finding, although confined by its inherent limitations (e.g. a relatively small and homogenous data set), appeared to have provided empirical evidence that should call for test developers and researchers to rethink the significance and impact of actual test use in practice, especially in contexts where (1) a test is used for purposes other than intended (e.g. CET, IELTS) and (2) competing values exist among different test stakeholders. An urgent question for future research is when the uses of a test do not support, or even conflict with test takers’ interests, how would test takers perceive its test use? More importantly, if test takers’ supportive attitudes towards the actual test use have a positive impact on test performance, would test takers’ negative attitudes towards test use impede their test performance, hence compromising the validity of inferences derived from test scores?

Conclusion

This study identified two main dimensions in terms of English learning motivation, perceived value implications in test takers’ attitudes towards the actual test use of the CET in practice and the potential effects of test takers’ characteristics on test performance. However, given the process of test justification and validation should be local and research based, how would test takers and other major stakeholders perceive and react to a high-stakes test and its test use are inevitably sensitive to each unique local cultural and social context.

Moreover, limited data diversity may have also affected the generalisability of the findings in this research. Both qualitative and quantitative data of this study were collected from the same two universities in Eastern China, which shared similar characteristics in many aspects. Given the large scale and diversity of Higher Education in China, the items measuring the construct of test takers’ motivations and attitudes are highly likely to be subject to a lack of input of test takers/students from universities of other areas and with different characteristics. More empirical research is needed to explore the remaining questions such as: How would test takers/students from lower ranking universities and/or from economically less developed areas perceive the CET test use in reality? What would be their motivational composition? Would the inclusion of their views alter the identified causal relationship in this study? In summary, the complex and value-sensitive relationships among test takers’ learning motivation, attitudes towards test use and other test-related factors and test performance deserve appropriate attention in validity studies of the language testing as to better inform test users and other stakeholders in making fairer score-based decisions, manufacture positive consequential effects and maintain test accountability.

Appendix

Table 2

Table 2 A questionnaire that collects test takers’ personal information and measures their English learning motivation and attitudes towards the actual test use of the CET

Name	Gender
University	
College/Department	
What is your CET total score?	
What is your motivation in learning English?	
1 = strongly disagree, 2 = disagree, 3 = not sure, 4 = agree, 5 = strongly agree	1 2 3 4 5
B1- I have to learn English to pass the test.	
B2- I would like to communicate and make friends with foreigners.	
B3- I would like to travel abroad.	
B5- I plan to study abroad in the future.	
B6- I have to learn English because it is a compulsory subject.	
B7- I think it is cool to speak fluent English.	
B13- It is useful to learn English because it is an important international language.	
B15- I am learning English because it will help me to read English academic papers and journals.	
B16- I am learning English because it should help me become more competitive in the future.	
B17- I would not learn English if there were no high-stakes test like the CET.	
B18- I often feel nervous in English classes.	
B20- My past test-orientated language learning experience has de-motivated me to learn English.	
B21- I do not need English in my future career, nor life, so it is a waste of time to learn English.	
What is your attitude toward the actual test use of the CET in practice?	
1 = strongly disagree, 2 = disagree, 3 = not sure, 4 = agree, 5 = strongly agree.	1 2 3 4 5
EE5- It is China’s tradition to use test results for selection.	
EE6- I believe it is so far the most effective and fair method to allocate limited social resources.	
EE7- I think it is appropriate to use test results for selection because it allows students from disadvantaged families to move forward.	
EE9- In present-day China, I do not think there is a better method than using the test result for selection.	

Abbreviations

AttiTuse: Attitudes towards the actual test use of the CET; CET: College English Test; CFA: Confirmatory factor analysis; CFI: The comparative fit index; C.R.: Critical ratio; EFA: Exploratory factor analysis; ESD: Exploratory Sequential Design; GFI: The goodness-of-fit index; IELTS: International English Language Testing System; ML: The maximum likelihood; MOE: Ministry of Education; NCETC: National College English Test Committee; RMSEA: The root mean square error of approximation; SEM: Structural Equation Modelling; SLA: Second Language Acquisition; SMC: The squared multiple correlations; SRMR: The standardised root mean square residual; StTE: Standardised total effects; TRMoti: Test-related English learning motivation; TUMoti: Test-unrelated English learning motivation

Acknowledgements

The author would like to thank Ms de Veer for her review of this paper.

Author’s contributions

Bai Ying is the only author of this study. The author read and approved the final manuscript.

Funding

The author would like to acknowledge The International Research Funds (TIRF) for providing financial support for this study.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The author declares there are no competing interests.

Received: 24 June 2020 Accepted: 17 August 2020

Published online: 28 August 2020

References

- Ajzen, I. (1988). *Attitudes, personality and behaviour*. Milton Keynes: Open University Press.
- Ajzen, I. (2005). *Attitudes, personality and behaviour*, (2nd ed.,). England, UK: Open University Press.
- Akiyama, T. (2004). *Introducing speaking tests into Japanese Senior High School entrance examination*. The University of Melbourne: Unpublished doctoral dissertation.
- Allport, G. W. (1971). Attitudes. In K. Thomas (Ed.), *Attitudes and behaviour*. Harmondsworth: Penguin.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004). Statistical analyses for language assessment. Cambridge, UK: Cambridge 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: Oxford University Press.
- Bai, Y. (2019). Intended goals and reality in practice: A test-takers' perspective on the College English Test in China. In S. Papageorgiou, & K. M. Bailey (Eds.), *Global perspective on language assessment*. New York, NY: Routledge.
- 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 & Research*, 16, 78–117.
- Bray, M., & Steward, L. (1998). *Examination systems in small states: Comparative perspectives on policies, models and operations*. London: Commonwealth Secretariat.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*, (2nd ed.,). New York: Psychology Press.
- Cai, J. (2006). *ELT at tertiary level in China: Review, reflection and research*. Shanghai: Fudan University Press.
- 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. (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: Routledge.
- Cheng, L., & Qi, L. (2006). Description and examination of the National Matriculation English Test in China. *Language Assessment Quarterly: An International Journal*, 3, 53–70.
- Cizek, G. J. (2016). Validating test score meaning and defending test score use: Different aims, different methods. *Assessment in Education: Principles, Policy & Practice*, 23(2), 212–225.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*, (4th ed.,). Thousand Oaks, CA: SAGE.
- 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, US: Springer.
- 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, 421–462.
- Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: A Hungarian Perspective*. Clevedon: Multilingual Matters.
- Dörnyei, Z., MacIntyre, P. D., & Henry, A. (Eds.) (2015). *Motivational dynamics in language learning*. UK: Multilingual Matters.
- Eagly, A., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando: Harcourt Bruce.
- Eckstein, M. A., & Noah, H. J. (1992). *Examinations: Comparative and international studies*. Oxford: Pergamon Press.
- Gao, S., Mokhtarian, P. L., & Johnston, R. A. (2008). Non-normality of data in structural equation models. *Transportation Research Record*, 2082, 116–124.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- 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: John Wiley.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.
- Garner, M., & Huang, D. (2014). *Testing a nation: The social and educational impact of the College English Test in China*. Bern, Switzerland: Peter Lang.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis*, (Sixth ed.,). New Jersey: Pearson Education.
- Harlow, L. L. (1985). Behaviour of some elliptical theory estimators with non-normal data in a covariance structures framework: A Monte Carlo study. Unpublished PhD thesis. University of California, Los Angeles.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge: Cambridge University Press.

- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94(3), 319–340. <https://doi.org/10.1037//0033-295x.94.3.319>.
- Huang, D. (2002). *A preliminary investigation of CET4 washback. Paper presented at the Paper presented at the International Conference on Language Testing and Language Teaching*. Shanghai: Shanghai Jiaotong Studies University.
- Jin, Y. (2008). Powerful tests, powerless test designers: Challenges facing the College English Test. *CELEA Journal (Bimonthly)*, 31(5), 3–11.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford, UK: Pergamon.
- Kunnan, A. J. (1995). *Test taker characteristics and test performance: A structural modeling approach*, (vol. 2). Cambridge: Cambridge University Press.
- Lambert, W. E. (1963). Psychological approaches to the study of language 1: On learning, thinking and human abilities. *Modern Language Journal*, 14, 51–62.
- Lambert, W. E. (1967). A social psychology of bilingualism. *Journal of Social Issues*, 23, 91–109.
- Lei, M., & Lomax, R. G. (2005). The effects 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.
- Liu, H., Tian, J., Zhang, Y., & Zheng, R. (2002). *The history of examination in China*. Wuhan, Hubei, China: The Central China Normal University Press.
- Markus, H., & Nurius, P. (1986). Possible Selves. *American Psychologist*, 41(9), 954–969. <https://doi.org/10.1037//0003-066x.41.9.954>.
- 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.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension* (Malden, MA: Blackwell).
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*, (3rd ed., pp. 13–103). New York: American Council on Education & Macmillan.
- Moss, P. A. (2016). Shifting the focus of validity for test use. *Assessment in Education: Principles, Policy & Practice*, 23(2), 236–251.
- 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.
- Newton, P. E., & Baird, J. (2016). The great validity debate. *Assessment in Education: Principles, Policy & Practice*, 23(2), 173–177.
- Rasti, I. (2009). Iranian candidates' attitudes towards IELTS. *Asian EFL Journal*, 11(3), 110–155.
- 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.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Pearson.
- Sireci, S. G. (2016). On the validity of useless tests. *Assessment in Education: Principles, Policy & Practice*, 23(2), 226–235.
- Spolsky, B. (1989). *Conditions for second language learning: Introduction to a general theory*. Oxford: Oxford University Press.
- 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, 255–268.
- Wang, J. (2003). *The history of education in China [中国教育史]*. Guangzhou, Guangdong, China: Guangzhou Higher Education Press.
- 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*, 5(4).
- Warden, C. A., & Lin, H. (2000). Existence of integrative motivation in an Asian EFL setting. *Foreign Language Annals*, 33(5), 535–545.
- Wu, S. (2003). College English teaching: Problems and solutions. *Foreign Language Teaching and Research*, 3.
- Wu, J., & Lee, M. C. (2017). The relationships between test performance and students' perceptions of learning motivation, test value, and test anxiety in the context of the English benchmark requirement for graduation in Taiwan's universities. *Language Testing in Asia*, 7(9). <https://doi.org/10.1186/s40468-017-0041-4>.
- Xie, Q. (2011). Is test taker perception of assessment related to construct validity? *International Journal of Testing*, 11(4), 324–348.
- Xie, Y., Han, T., Lin, F., & Sun, J. (2007). A survey and reflection on English teaching for postgraduates. *Foreign Language World*, 1, 53–60.
- Yang, H., & Weir, C. J. (1998). *The CET validation study*. Shanghai: Shanghai Foreign Language Education Press.
- 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: Routledge.
- Zhou, L. (2003). Further development of the College English teaching. *Foreign Language Teaching and Research*, 3.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.