

RESEARCH

Open Access



Do school status and classroom English learning practices influence Chinese learners' perceptions of a high-stakes English test?

Manxia Dong*

*Correspondence:
dmanxia2013@163.com

Sichuan International Studies
University, No.33, Zhuangzhi
Road, Lieshimu, Shapingba
District, Chongqing 400031,
China

Abstract

Learners' views of a test have been regarded as a crucial source of evidence of a test's construct validity as well as a measurement of its impact. This study investigates students' ($N = 3105$) perceptions of the National Matriculation English Test (NMET), a high-stakes test in China, and explores the influences of school status and classroom English learning practices on their perceptions. The sample was randomly split in half to conduct cross-validations. Data analysis involved descriptive statistics, exploratory factor analysis, *t*-test, correlation analysis, standard multiple regression, and structural equation modeling. The survey results demonstrated that students viewed test validity and importance positively and test impact negatively. The results also revealed that school status and classroom English learning practices significantly influenced students' perceptions of a test. Specifically, those in more prestigious schools held more positive perceptions of NMET validity and importance and more negative perceptions of its impact. Classroom English teaching-related learning practice was more associated with students' positive perceptions of NMET validity and importance, whereas classroom test preparation was more associated with negative views of NMET impact. This study has important implications for test design, school administration, and learning and teaching.

Keywords: Test perceptions, School status, Classroom learning practices

Introduction

In the field of education, stakeholders' perception of a test and the relationship between test and performance has received considerable attention (e.g., Hughes, 1993; Brown & Hirschfeld, 2005; Stricker et al., 2006; Cheng et al., 2011). Furthermore, students have been regarded as direct, ultimate, and important stakeholders in any testing situation (Hamp-Lyons, 1997; Rea-Dickins, 1997), and their perceptions of a test are considered a crucial source of evidence for its construct validity (Messick, 1989). This notion has become a major concern of both test constructors and test users (Stricker et al., 2006). For instance, Sato and Ikeda (2015) investigated the test takers' perceptions of ability measured by certain items (i.e., face validity) in a high-stakes test with the purpose of examining the extent to which test taker perceptions meet test developer intention.

Their study provided some useful implications for test development. Additionally, Murray et al. (2012) suggested that test takers' views of a measure could be used to partly assess the test impact. For instance, Fan et al. (2014) found that students' positive perceptions of a test tend to produce beneficial effects on their learning. Thus, it is critical to understand learners' views of a test and what factors contribute to their views. This study investigates the perceptions of the National Matriculation English Test (NMET) from senior high school students in China and explores the factors that influenced such perceptions. The NMET as a high-stakes English test has garnered much concern from researchers. Numerous studies have been conducted on the NMET through the analysis of test papers or investigating test stakeholders' views of the NMET to examine its validity (e.g., Dong, 2008; Pan & Qian, 2017). They found that the NMET does have satisfactory validity (Dong, 2020). However, there are fewer studies on the NMET's impact through investigating learners' perceptions of the NMET and the factors that influence these perceptions. This study has strong implications that can benefit the field of education. In particular, this study's findings will be helpful to those concerned with understanding more about the nature and origins of students' perceptions of a high-stakes test. In addition, the study offers essential insights for test designers, schools, students, and teachers.

Literature review

Considering the importance of learners' perceptions of tests, certain researchers have investigated such perceptions and the factors that influence it. For instance, Scott (2007) employed a qualitative method to explore the stakeholders' perceptions of test impact and found that they varied across different grades. Cheng et al. (2011) assessed the students' views of a school-based assessment in Hong Kong and suggested that the students' observed language competencies were directly and significantly related to their test perceptions. Li et al. (2012) administered a survey on 150 undergraduate students to acquire their perceptions of the impact of the College English Test (CET) in China. They found that some students viewed the CET's impact as pervasive and affirmed that the CET motivated their efforts to learn English. In contrast, many students expressed that the CET made them more anxious and stressed regarding learning English. Fan and Ji (2014) have investigated the test candidates' attitudes toward a school-based test and explored the relationships between test candidates' view and their characteristics. Their results indicated that gender and academic background do not significantly affect the test takers' reported attitudes. These studies examined the students' perceptions of tests and explored the factors (i.e., gender, grade, proficiency level, and academic background) that influenced their perceptions. Xie and Andrews (2013) suggested that school tier or ranking (or school status) mediates the intensity and length of the test impact. Most recently, Zhang and Bournot-Trites (2021) demonstrated that school status significantly influences students' perceptions of a test's purpose and preparation, apart from which test validity, importance, and impact have been significant concerns for researchers. However, it remains unclear whether learners' perceptions of a high-stakes test, as related to test validity, impact, and importance, are also affected by school status. Thus, this study investigates the effects of school status on students' perceptions of the NMET, a high-stakes test.

Apart from the demographic variables, Murray et al. (2012) also found that test candidates' personal experience correlates with their perceptions of a test. For instance, Elder et al. (2002) indicated that test takers' negative views of an assessment are likely to affect their performance. Xie (2015) utilized structural equation modeling (SEM) to examine the relationship between test takers' views of test validity and their language learning activities and test preparation. Their results revealed that a favorable perception of test validity exerts a greater influence on test preparation than language learning activities do. Based on Xie's (2015) research, Dong (2020) extended learning practices to four types of after-class learning activities; additionally, they explored the structural relationship between learners' perceptions of NMET validity, importance, and impact and the four types of English learning activities through SEM. Their study revealed that learners' perceptions of test validity, impact, and importance affect various types of after-class English learning activities. These studies examined the influence of test perceptions on learning practices or performance. However, whether test takers' engagement with learning practices or learning behaviors affects their perceptions of a test remains unknown. In the context of education, classroom learning, as a reflection of students' educational experience as well as educators' teaching practices, is an important and indispensable channel for students' learning. Moreover, it influences whether their learning has been successful. Thus, this study aims to explore the exact relationship between classroom English learning practices and learners' test perceptions to provide useful suggestions and implications for learning as well as teaching.

Considering the research gaps in this literature, this study primarily investigates the students' perceptions of the NMET in China. Additionally, it examines the influences of school status and classroom English learning practices on such perceptions. The study addresses three research questions (RQs):

RQ1. What are Chinese high school students' perceptions of the NMET regarding test validity, importance, and impact?

RQ2. Does school status influence students' views of the NMET? If so, how?

RQ3. Do students' classroom English learning practices impact their reported perceptions of the NMET? If so, how?

Research context

The current Chinese education system is a centralized and public system, involving "pre-school, primary, secondary (junior secondary and senior secondary), and tertiary levels" (Dong, 2020, p.3). Education at the primary and junior secondary levels is compulsory. After receiving a 9-year mandatory education program, over half of graduates continue to pursue 3 years of high school learning; the rest enter either vocational high schools or job markets. After completing 3 years of general senior high schooling, students take the National Matriculation Test (NMT) to compete for admission to universities or colleges.

The NMT battery has been regarded as the most competitive high-stakes selection test. Its importance is so significant in China that it impacts thousands of households in addition to test takers' future careers (Liu, 2010). The NMT battery includes five to six subjects. The English subject, namely the NMET, is one of three compulsory tests

for all candidates; it is also a research focus for this study. Every year, approximately ten million secondary school graduates attempt this examination. The current NMET consists of written and oral assessments. The written assessment adopts the pen-and-paper format to examine the test takers' abilities in reading, listening, English language usage, and writing (Zhang & Bournot-Trites, 2021). Approximately 60–70% of the test items follow a multiple-choice format that is automatically graded by a computer. The speaking subtest is seldom assessed and usually excluded from the NMET total score owing to the difficulty in conducting a face-to-face speaking subtest within a short period with approximately 10 million test takers per year.

Research method

Participants

Participants included students from six senior high schools located in four districts and counties of a municipality in southwestern China. School rankings were considered in the selection process; there were three ordinary ($N = 1318$, 42.4%) and key ($N = 1787$, 57.6%) schools each. Overall, 3105 participants were recruited, and their responses to the questionnaire survey were recorded. These participants consisted of 1402 male (45.2%) and 1703 female (54.8%) students. Of the participants, 1199 (38.6%), 1098 (35.4%), and 808 (26%) were year 1, 2, and 3 students, respectively. Valid responses were divided into three levels of groups that self-reported the range of their English scores. They included high- (scoring above 120 points out of 150) ($N = 789$, 25.4%), medium- (scoring from 91 to 119 points) ($N = 1422$, 45.8%), and low-level groups (scoring below 90 points) ($N=894$, 28.8%).

Instrumentation

This study uses two questionnaires: the Students' Test Perceptions Questionnaire and the Classroom English Learning Practices Questionnaire. The questionnaire items were generated based on group interviews with six students from senior high schools and a pilot study with a semi-structured classroom English learning log table. The questionnaires were assessed with careful scrutiny by the researcher, experts, peers, and learners to improve the appropriateness, accuracy, and clarity of each item's wording. The two questionnaires were subsequently piloted on 179 students. Some items were removed or revised before the questionnaires were utilized in the main study. The detailed development and validation of both questionnaires have been reported in the authors' linked studies (Dong et al., 2021).

The questionnaires adopted in the main study consisted of 22 questions that were administrated in Chinese. The first section consisted of three demographic items, including gender, grade, and self-reported English proficiency score. The questionnaire regarding students' perceptions of the NMET consisted of 10 items ($\alpha = 0.79$), rated from "1 = disagree" to "5 = agree." The Classroom English Learning Practices Questionnaire included nine items that reflected English learning activities within the classroom, which was rated from "1 = never" to "5 = always."

Data collection and analysis

Printed hard copies of the questionnaire and an administration manual were mailed to the designated contact teachers in charge of administering, collecting, and mailing back the completed questionnaires to the researcher. This manual provided precise instructions on how to administer the questionnaires and clarify uncertainties. Before administering the questionnaire survey, teachers explained the purpose of the research, assured the participants that their responses would be treated confidentially and anonymously, and obtained informed consent from the participants. The questionnaire survey was administered in the classroom with the support of the English teachers and was not time-bound.

SPSS 21.0 was used to screen and analyze the data. The data analysis involved an exploratory factor analysis (EFA), *t*-test, correlation analysis, standard multiple regression (SMR), and SEM. SEM, also known as covariance structure analysis and causal modeling, is a statistical technique used to examine the various relationships that are hypothesized among sets of variables that apply confirmatory factor analysis, structural regression, path, growth, multiple-groups, and multitrait-multimethod models (Byrne, 2006; Ockey & Choi, 2015; Ullman, 2007).

Before initiating data analysis, the data was reviewed for outliers and missing values. Then, descriptive statistics at the item and factor levels were calculated to answer the first research question. To validate the effect of school status and classroom English learning practices on test perceptions, the entire sample was randomly split in two ($N = 1552$, $N = 1553$) to conduct cross-validations. Regarding the second research question, a *t*-test was conducted on each of the two halves of the sample to explore the students' different perceptions of the NMET across higher and lower prestige schools, thereby verifying the effect of school status on learners' perceptions. To address the third research question, we primarily employed the first half of the sample ($N = 1552$) to perform a SMR for assessing the influence of students' engagement in classroom English learning practices on their test perceptions. Subsequently, SEM was conducted on the second half of the sample ($N = 1553$) to verify the relationship between classroom English learning practices and test perceptions that were explored in the SMR. In this study, the following fit indices were utilized to evaluate the model fit: chi-square, comparative fit index (CFI), Tucker-Lewis index (TLI), Bollen's incremental fit index (IFI) ≥ 0.9 , standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA) ≤ 0.08 with their confidence intervals (CI).

Results and discussion

Students' responses to the NMET (RQ1)

The item-level EFA was conducted to explore the patterns underlying the participants' perceptions of the NMET. Exploratory factor analysis using principal component analysis with a varimax rotation was also conducted. The results indicated that three factors were extracted, and they explained 65.83% of the total variance. They involved students' perceptions of test validity, importance, and impact. The descriptive statistics results of students' responses to the NMET at the factor and item levels are presented in Table 1.

Table 1 Descriptive statistics results of test perceptions

Factors	Variables	Means	S.D
Perception of test: validity (Pvali) ($M = 3.25$, $SD = 1.15$)	Pvali1: The NMET reflects my English level scientifically and objectively.	3.08	1.35
	Pvali2: The NMET examines my senior high school English learning.	3.38	1.40
	Pvali3: The NMET facilitates my setting a clear goal with respect to learning English.	3.28	1.31
Perception of test: importance (Pimpo) ($M = 3.28$, $SD = 1.14$)	Pimpo1: My NMET score gives me a feeling of pride.	2.97	1.43
	Pimpo2: My NMET score is important to my future English learning.	3.41	1.41
	Pimpo3: My NMET score enhances my self-confidence in English learning.	3.47	1.34
Perception of test: impact (Pimpa) ($M = 3.41$, $SD = .99$)	Pimpa1: The NMET orients my English learning to the test.	3.65	1.30
	Pimpa2: The NMET preparation takes time away from regular English learning.	3.11	1.28
	Pimpa3: The NMET pressures me to review test-taking skills.	3.60	1.26
	Pimpa4: The NMET content focuses my English learning.	3.29	1.43

Table 1 demonstrates that most items of the three factors of test perceptions had mean values greater than 3.0, indicating the participants' overall endorsement of these statements. A further analysis at the factor level revealed that test impact had a greater mean value ($M = 3.41$, $SD = .99$) than test importance ($M = 3.28$, $SD = 1.14$) and validity ($M = 3.25$, $SD = 1.15$). These findings demonstrated that the participants had more positive perceptions of test impact compared with test validity and importance.

Students' perceptions of test validity (Pvali) consisted of three items with a mean value ranging from 3.08 to 3.38, suggesting that the students tended to endorse test validity. Pvali2 (The NMET examines my senior high school English learning) had the greatest mean value ($M = 3.38$, $SD = 1.40$), followed by Pvali3 (The NMET facilitates my setting a clear goal with respect to learning English) ($M = 3.28$, $SD = 1.31$). Compared to Pvali2 and Pvali3, Pvali1 (The NMET reflects my English level scientifically and objectively) had a lower mean ($M = 3.08$, $SD = 1.35$), indicating that the students endorsed Pvali1 less. This result is partly because teachers and students tend to pay insufficient attention to speaking in a regular teaching and learning curriculum as it is unassessed in the NMET (Dong, 2018). This result may also be attributed to the fact that the NMET's test format is primarily (60–70%) multiple-choice, which has been criticized for being considered to encourage test takers' guessing.

The students' perceptions of test importance (Pimpo) contained three items, with mean values ranging from 2.97 to 3.47. Pimpo 3 (My NMET score enhances my self-confidence in English learning) had the greatest mean value ($M = 3.47$, $SD = 1.34$), followed by Pimpo2 (My NMET score is important to my future English learning) ($M = 3.41$, $SD = 1.41$). The results indicated that the students supported the importance of the NMET. Compared with Pimpo2 and Pimpo3, Pimpo1 (My NMET score gives me a feeling of pride) had a lower mean ($M = 2.97$, $SD = 1.43$). This result suggests that

although the Chinese tradition of emphasizing face-value perceptions influences some students' learning, most students recognize that they commit to their studies for themselves rather than for their reputation (Dong, 2018).

The students' perceptions of test impact (Pimpa) contained four items, with mean values ranging from 3.11 to 3.65, suggesting that students perceived that the NMET influenced their learning. According to some researchers' definitions and descriptions of values or directions of test influence (e.g., Cooley, 1991; Jin, 2006), test impact in this study was negative, specifically for Pimpa1 (The NMET orients my English learning to the test) ($M = 3.65$, $SD = 1.30$) and Pimpa3 (The NMET pressures me to review test-taking skills) ($M = 3.60$, $SD = 1.26$). The outcomes indicated that the students' views of test impact are overall moderately negative.

In summary, participants generally spoke positively of the NMET in terms of its validity and importance. They admitted its effectiveness in assessing their learning content and English proficiency levels supported its positive role in facilitating their English learning and agreed to its importance regarding their learning and career. Conversely, they tended to hold more negative perceptions of test impact.

The influence of school status on students' test perceptions (RQ2)

This section compares the students' views at the factor level across school status. To compare whether students from key high schools differed from those from ordinary schools in their view of the NMET, we performed an independent sample *t*-test. We also applied the Bonferroni adjustment procedure to reduce the margin of error by dividing the traditional alpha value of 0.05 by the three factors of test validity, importance, and impact ($0.05/3 = 0.017$).

The descriptive statistics results at factor levels and the *t*-test results are presented in Table 2. The findings demonstrated that the mean values of students' perceptions of the NMET from key high schools were significantly greater than their counterparts' perceptions from ordinary schools. Moreover, students from higher prestige schools held more positive perceptions in terms of test validity and importance than their counterparts in lower prestige schools. This result was supported by Zhang and Bournot-Trites's (2021) finding that learners in higher-ranking schools view the NMET more objectively.

Notably, a surprising result regarding test impact is that the students in higher prestige schools tended to hold more negative perceptions. Specifically, they maintained that the NMET exerts stronger negative effects on them than it does on their counterparts in lower prestige schools. The primary reason for this result was that nearly all key senior high schools in China aimed to cultivate candidates for prestigious universities (i.e.,

Table 2 *T*-test of students' perceptions by school status across two halves of samples

Variables	The first half sample ($N = 1552$)			The second half sample ($N = 1553$)		
	Ordinary ($N = 652$), mean, SD	Key ($N = 900$), mean, SD	Sig.	Ordinary ($N = 666$), mean, SD	Key ($N = 887$), mean, SD	Sig.
Pvalidity	3.10, 1.190	3.38, 1.133	.000	3.07, 1.157	3.34, 1.121	.000
Pimpact	3.27, 1.012	3.48, .963	.000	3.33, 1.000	3.50, .961	.001
Pimportance	3.16, 1.157	3.36, 1.105	.001	3.20, 1.126	3.35, 1.148	.010

universities of 211 Project and 985 Project). Project 211 was proposed by the Chinese State Council in 1992 and initiated in 1995 to build approximately 100 key universities with key disciplines and dominant majors pertaining to the 21st century. The council initiated Project 985 (which includes, at present, 39 universities) in 1998 to build world-class universities, which are selected from former Project 211 universities. An exceedingly small proportion of these candidates could be admitted by the universities of Project 211 (about 5%) and Project 985 (about 2%). Therefore, higher prestige schools experience fiercer competition and greater pressure regarding student admission to prestigious universities, compared with ordinary universities that primarily aim at cultivating candidates for ordinary universities and colleges. Thus, higher prestige schools tend to engage in more frequent test-oriented learning practices.

Influences of classroom English learning practices on students' test perceptions (RQ3)

Descriptive statistics of classroom English learning practices

Classroom learning practices describing representative learning activities pursued in the English class generated two factors that explained 67.41% of the total variance. These factors represented two types of learning activities: classroom teaching-related learning (CTL) and classroom test preparations (CTP). Table 3 reveals that CTL generated a mean value of 3.73 ($SD = .87$), consisted of five variables (CTL1 to CTL5) with the mean values ranging from 3.14 to 4.02. CTP had a mean value of 3.44 ($SD = 1.06$) and consisted of four variables (CTP1–CTP4) ranging from 3.27 to 3.63. Compared with CTP, CTL exhibited a greater mean value, suggesting that teaching-related English learning activities were more frequently conducted than test preparation in the classrooms.

SMR analysis

SMR analysis was conducted on the first half of the sample ($N = 1552$) to explore classroom English learning practices' effect on learners' test perceptions. Prior to conducting the SMR analysis, we performed the Pearson bivariate correlation analysis of test perceptions and classroom English learning practices scale at the factor level. The results

Table 3 Descriptive statistics results of classroom learning practices

Factors	Variables	M	SD
Class teaching-related English learning (CTL) (M = 3.73, SD = .87)	CTL1: Learning vocabulary and grammar in the textbook	4.02	1.06
	CTL2: Training listening, speaking, reading, and writing skills	3.64	1.14
	CTL3: Participating in class discussion and role plays and so on.	3.14	1.29
	CTL4: Doing exercises assigned by teachers	3.91	1.13
	CTL5: Listening to teachers' comments on exercises in the textbook	3.92	1.11
Class test preparation (CTP) (M = 3.44, SD = 1.06)	CTP1: Drilling the NMET sample papers	3.45	1.21
	CTP2: Listening to teachers' comments on test items on the NMET sample papers	3.63	1.16
	CTP3: Drilling the past NMET papers	3.27	1.24
	CTP4: Listening to teachers' comments on test items on the past NMET papers	3.42	1.23

Table 4 Correlation analysis of the examined variables ($N = 1552$)

PVali	PImpa	PImpo	CTL	CTP
–	.140**	.491**	.358**	.252**
	–	.296**	.164**	.234**
		–	.281**	.233**
			–	.536**
				–

* $p < .05$; ** $p < .01$; *** $p < .001$ ** (two tails)

Table 5 Standard multiple regression statistics ($N = 1552$)

Variables	R	R^2	Adjusted R^2	$F(2/1549)$	β	t	p	Tolerance	VIF
DV Pvali	.327	.107	.106	92.968***					
IV CTL					.303	10.888	.000	.744	1.344
IV CTP					.044	1.577	.115	.744	1.344
DV Pimpa	.277	.077	.075	64.216***				.744	1.344
IV CTL					.091	3.214	.001	.744	1.344
IV CTP					.219	7.750	.000	.744	1.344
DV Pimpo	.334	.112	.110	97.307***				.744	1.344
IV CTL					.268	9.667	.000	.744	1.344
IV CTP					.105	3.789	.000	.744	1.344

Table 6 Standardized path coefficients of the structural model ($N = 1553$)

Exogenous variable	Endogenous variable	Path coefficient	P
CTL	Pvali	.42	.000
CTL	Pimpo	.37	.000
CTL	Pimpa	.16	.000
CTP	Pvali	.01	.867
CTP	Pimpo	.03	.405
CTP	Pimpa	.23	.000

(see Table 4) demonstrated that all the variables had significantly low to moderate correlation ($p < .01$). Following the correlation analysis, we conducted a SMR analysis (see Table 5 for the results). The two diagnostics indices of the tolerance value ($> .20$) and the variance inflation factor (< 5.0) were within a satisfactory range, indicating that a multi-collinearity problem did not arise in the variables within this data set (Xie, 2013). Thus, these variables could be used in the regression analysis.

The results demonstrated that CTL and CTP had significant contributions to students' perception of test validity ($R^2 = .107$, $p < .05$) and importance ($R^2 = .112$, $p < .05$), explaining 10.7% and 11.2% of the additional variance, respectively (Table 6). CTP's impact on Pvali and Pimpo ($\beta = .044$, $\beta = .10$, respectively) was lower than CTL's ($\beta = .303$, $\beta = .268$, respectively).

The SMR analysis results indicated that CTP and CTL significantly contributed to the students' perception of test impact ($R^2 = .077$, $p < .05$), explaining 7.7% of the additional

variance. Regarding the variables of classroom learning practices, CTP had a significant and intermediate influence on Pimpa ($\beta = .219, p = .000$), whereas CTL had more marginal effects on Pimpa ($\beta = .091, p = .001$).

Structural equation modeling

Based on the SMR analysis result of the first half of the sample ($N = 1552$), we conducted a structural equation modeling on the second half ($N = 1553$). Figure 1 presents the final model with standardized coefficients that estimated the relationship between classroom learning practices and learners' test perceptions. Based on the results of the correlation analysis of Pvali, Pimpa, and Pimpo, we fixed the error terms of test perception variables to be correlated. In addition, Byrne (2011) suggests that fixing error terms correlating among the latent variables indicates that these variables in the structural model may be influenced by factors not included in the model.

The fit indices of the hypothesized model are as follows: $\chi^2 = 1162.6$, $df = 142$, CFI = .923, TLI = .907, IFI = .923, SRMR = .059, and RMSEA = .068 with 90% CI (.064, .072). Considering all the indices, the model achieved a satisfactory fit despite the large chi-square statistic owing to the big sample size. The hypothesized model verified the

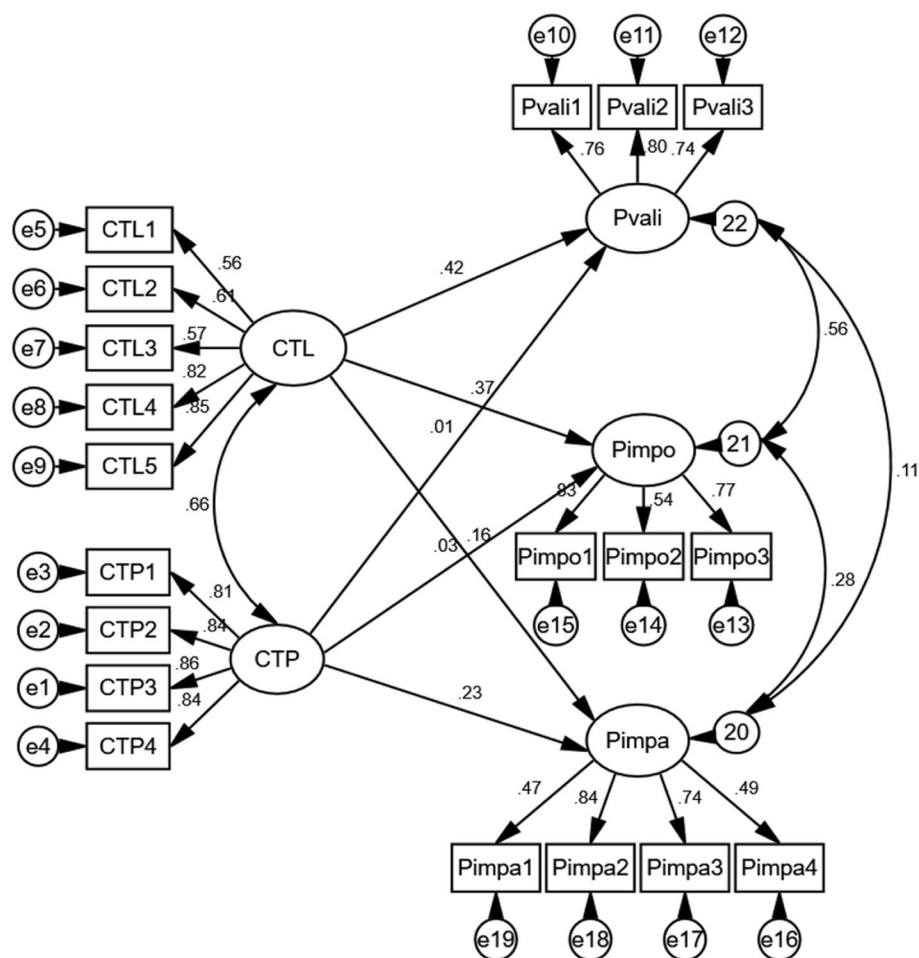


Fig. 1 Structural model with standardized parameters estimation

underlying pattern of the classroom English learning practices' impact on learners' test perceptions. Comparing the results between the SMR and the SEM found that nearly all of the results were consistent except for one path (CTP→Pimpo). The regression coefficients of CTP→Pimpo in the SEM ($\beta = .03, p = .405$) were not significant, although it was significant in the SMR analysis ($\beta = .105, p = .000$). The small difference between the SMR analysis and the SEM was likely because of the sample and the analysis method effect. To clarify the reason for the difference, we repeatedly split the sample into halves and conducted the SEM with these halves. The findings of each SEM analysis were all consistent. Thus, we intended to support the SEM results to analyze the influence paths.

The regression result demonstrated that CTL had positive and significant effects on the learners' test perceptions. Compared with the influence on the students' perceived test impact ($\beta = .16$), the CTL exerted considerably greater effects on students' views regarding test validity ($\beta = .42$) and importance ($\beta = .37$). Specifically, classroom teaching-related activities better predicted students' positive test perceptions better than they did negative perceptions, the primary reason being that. CTL was more consistent with teaching syllabus and curriculum objectives, which were additionally beneficial in improving students' learning. This reason is supported by Dong's (2020) study that teaching-related English learning practices are more helpful in boosting learners' learning outcomes.

Conversely, CTP had a positive and significant effect merely on students' perceived test impact ($\beta = .23, p < .001$); however, it lacked significant effects on test validity ($\beta = .01, p = .867$) and importance ($\beta = .03, p = .405$). The results indicated that CTP had a stronger association with students' negative test perceptions. Specifically, the more frequently test preparation occurred in class, the more negatively students held perceptions of test impact. This might be because test preparation activities aimed at improving test scores, which easily led to deviating from curriculum objectives, thereby negating the cultivation of students' language ability and reduce their English learning interest (e.g., Jin, 2006; Xie, 2013). Moreover, some empirical studies also verified that test preparation only slightly improves learning outcomes or does not significantly boost them at all (Dong, 2020; Robb & Ercanbrack, 1999; Xie, 2013). Thus, test preparation is usually regarded as a less beneficial learning practice. Therefore, students who engage in it more frequently tended to view test impact negatively.

Conclusion

This study investigated the learners' perceptions of the NMET by surveying a large sample of students and exploring the impact of school status and classroom English learning practices on students' test perceptions. It found that overall, students tended to speak positively of the NMET in terms of its validity and importance; however, they had negative perceptions of test impact. Moreover, this study found that school status and classroom English learning practices significantly affected the learners' test perceptions. Specifically, those in higher prestige schools held more positive perceptions of NMET validity and importance, and more negative ones of its impact. CTL was more associated with students' positive perceptions of test validity and importance, whereas CTP was more associated with their negative views of test impact. These findings may be helpful for those concerned with understanding more about

the nature of students' perceptions of a high-stakes test and the origins of such perceptions. Based on our findings, some implications and suggestions were provided for test designers, schools, students, and teachers.

This study found that students held relatively lower endorsement of test validity that was likely attributed to the NMET's preference in test design toward the multiple-choice test format and to the speaking test being excluded from the NMET. A test, especially a high-stakes one, plays a crucial guiding role in teaching and learning and even determines teaching and learning content, pace, materials, and methods. Thus, test designers attempt to improve test design by fully considering a test' positive guiding roles in teaching and learning.

This study found that students in higher prestige schools tended to hold more negative perceptions of test impact than their counterparts in lower prestige schools, which indicated that the NMET exerted a stronger negative influence on students' learning. According to Shih (2007), the way students view a test greatly affects their perceptions of it, as well as their learning behaviors (i.e., regular learning, test preparation). We suggest that school administrators and teachers, as significant relevant parties, should guide their students' opinion of tests, hold an appropriate opinion of a test and long-term vision of learning goals, and improve students' ability rather than test scores and a test-oriented education.

This study demonstrated that students' CTL practices (or regular learning practices in the classroom) were more associated with students' positive perceptions of test validity and importance. Classroom learning practices reflected the educators' teaching practices and concepts. Moreover, teachers have the most important effect on their students' perceptions of a test. Thus, teachers may help their students benefit by attaching greater importance to regular teaching practices and reducing the significance of test preparation.

A major limitation of this study was that the data included students' self-reported information. There was insufficient data from other sources to verify the results. Thus, the findings of this study should be interpreted with caution.

Appendix

English Translation of Questionnaire

Part 1: Personal data

1.	Your gender	A. Female	B. Male	
2.	Your grade	A. Grade 1	B. Grade 2	C. Grade 3
3.	Which score range best represents your English proficiency in the recent academic year? Please select the closest option.			
A.	Below 90 points	B. 91 to 119 points	C. Above 120 points	

Part 2: Perceptions of the NMET

Please grade the following statements of senior high students' opinions regarding NMET using a 5-point scale format: 1 = disagree, 2 = somewhat disagree, 3 = undecided, 4 = somewhat agree, and 5 = agree.

Pvali1	The NMET reflects my English level scientifically and objectively.	1	2	3	4	5
Pvali2	The NMET examines my senior high English learning.	1	2	3	4	5
Pvali3	The NMET facilitates my setting a clear goal with respect to learning English.	1	2	3	4	5
Pimpa1	The NMET orients my English learning to the test.	1	2	3	4	5
Pimpa2	The NMET preparation takes time away from regular English learning.	1	2	3	4	5
Pimpa3	The NMET pressures me to review test-taking skills.	1	2	3	4	5
Pimpa4	The NMET content focuses my English learning.	1	2	3	4	5
Pimpo1	My NMET score gives me a feeling of pride.	1	2	3	4	5
Pimpo2	My NMET score is important to my future English learning.	1	2	3	4	5
Pimpo3	My NMET score enhances my self-confidence in English learning.	1	2	3	4	5

Part 3: Classroom English learning practices

Please grade the following statements of senior high students' English learning activities using a 5-point scale format: 1 = never, 2 = seldom, 3 = sometimes, 4 = often, and 5 = always.

CTL1	Learning vocabulary and grammar in the textbook	1	2	3	4	5
CTL2	Training listening, speaking, reading, and writing skills	1	2	3	4	5
CTL3	Participating in class discussion and role plays and so on	1	2	3	4	5
CTL4	Doing exercises assigned by teachers	1	2	3	4	5
CTL5	Listening to teachers' comments on exercises in the textbook	1	2	3	4	5
CTP1	Drilling the NMET sample papers	1	2	3	4	5
CTP2	Listening to teachers' comments on test items on the NMET sample papers	1	2	3	4	5
CTP3	Drilling the past NMET papers	1	2	3	4	5
CTP4	Listening to teachers' comments on test items on the past NMET papers	1	2	3	4	5

Acknowledgements

I would like to express my heartfelt thanks to Professor Shen Zou for her guidance in conducting this research and Fei Liu for her support in collecting the data.

Author's contributions

MD is the only author of this manuscript. The author(s) read and approved the final manuscript.

Author's information

Manxia Dong, Ph.D., is a professor of language assessment and English education at Sichuan International Studies University, Chongqing, China. Her research interests are language assessment, language teaching, and learning. She has conducted research in the area of language assessment, L2 teaching, and learning and has published more than thirty journal articles of the washback and validity of high-stakes tests and English teaching and learning.

Funding

This research work had no funding to support it.

Availability of data and materials

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

Declarations

Competing interests

The author declares that there are no competing interests.

Received: 25 April 2022 Accepted: 12 August 2022

Published online: 19 September 2022

References

Brown, G. T. L., & Hirschfeld, G. H. F. (2005). *Secondary school students' conceptions of assessment*. In *Conceptions of assessment and feedback project report #4*. University of Auckland.

- Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming*, (2nd ed.,). Erlbaum.
- Byrne, B. M. (2011). *Structural equation modelling with Mplus: Basic concepts, applications, and programming*. Routledge.
- Cheng, L. Y., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, 28(2), 221–249. <https://doi.org/10.1177/0265532210384253>.
- Cooley, W. W. (1991). State-wide student assessment. *Educational Measurement: Issues and Practice*, 10(4), 3–6. <https://doi.org/10.1111/j.1745-3992.1991.tb00209.x>.
- Dong, M. X. (2008). Content validity study on reading comprehension tests of NMET. *Teaching English China*, 31(4), 29–39.
- Dong, M. X. (2018). What perceptions do senior high school students hold of the NMET under the background of NMET reform? *Foreign Language Testing and Teaching*, 1, 19–26.
- Dong, M. X. (2020). Structural relationship between learners' perceptions of a test, learning, and learning outcomes: A study on the washback mechanism of a high-stakes test. *Studies in Educational Evaluation*, 64, 1–11. <https://doi.org/10.1016/j.stueduc.2019.100824>.
- Dong, M. X., Fan, J., & Xu, J. (2021). Differential washback effects of a high-stakes test on students' English learning process: evidence from a large-scale stratified survey in China. *Asia Pacific Journal of Education*, 1–18. <https://doi.org/10.1080/02188791.2021.1918057>.
- Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: What does the test taker have to offer? *Language Testing*, 19(4), 347–368. <https://doi.org/10.1191/0265532202lt235oa>.
- Fan, J., & Ji, P. (2014). Test candidates' attitudes and their test performance: The case of the Fudan English test. *University of Sydney Papers in TESOL*, 9, 1–35.
- Fan, J., Ji, P., & Song, X. (2014). Washback of university-based English language tests on students' learning: A case study. *The Asian Journal of Applied Linguistics*, 1(2), 178–191. <https://caes.hku.hk/ajal/index.php/ajal/article/view/49>.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295–303. <https://doi.org/10.1177/026553229701400306>.
- Hughes, A. (1993). *Backwash and TOEFL 2000 (unpublished manuscript)*. University of Reading.
- Jin, Y. (2006). Improving test validity and washback: A proposed washback study on CET4/6. *Foreign Language World*, 6, 65–73.
- Li, H., Zhong, Q., & Suen, H. K. (2012). Students' perceptions of the impact of the college English test. *Language Testing in Asia*, 2(3), 1–18. <https://doi.org/10.1186/2229-0443-2-3-77>.
- Liu, H. (2010). Where should the national college entrance examination go? *Beijing University Education Review*, 8(2), 2–13.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*, (3rd ed., pp. 13–104). American Council on education and Macmillan.
- Murray, 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. <https://doi.org/10.1177/0265532212440690>.
- Ockey, G. J., & Choi, I. (2015). Structural equation modeling reporting practices for language assessment. *Language Assessment Quarterly*, 12(3), 305–319. <https://doi.org/10.1080/15434303.2015.1050101>.
- Pan, M., & Qian, D. (2017). Embedding corpora into the content validation of the grammar test of the National Matriculation English Test (NMET) in China. *Language Assessment Quarterly*, 14(2), 101–119. <https://doi.org/10.1080/15434303.2017.1303703>.
- Rea-Dickins, P. (1997). So, why do we need relationships with stakeholders in language testing? A view from the UK. *Language Testing*, 14(3), 304–314. <https://doi.org/10.1177/026553229701400307>.
- Robb, T. N., & Ercanbrack, J. (1999). A study of the effect of direct test preparation on the TOEIC scores of Japanese university students. *Teaching English as a Second or Foreign Language*, 3, 25–55.
- Sato, T., & Ikeda, N. (2015). Test-taker perception of what test items measure: A potential impact of face validity on student learning. *Language Testing in Asia*, 5(1), 1–16. <https://doi.org/10.1186/s40468-015-0019-z>.
- Scott, C. (2007). Stakeholder perceptions of test impact. *Assessment in Education: Principles, Policy & Practice*, 14(1), 27–49. <https://doi.org/10.1080/09695940701272807>.
- Shih, C. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135–161. <https://doi.org/10.3138/cmlr.64.1.135>.
- 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. https://doi.org/10.1207/s15327574ijt0603_3.
- Ullman, J. E. (2007). Structural equation modeling. In B. G. Tabachnick, & L. S. Fidell (Eds.), *Using multivariate statistics*, (5th ed., pp. 676–780). Allyn and Bacon.
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10(2), 196–218. <https://doi.org/10.1080/15434303.2012.721423>.
- Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, 50, 56–68. <https://doi.org/10.1016/j.system.2015.03.002>.
- Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation: Testing a model of washback with structural equation modelling. *Language Testing*, 30(1), 49–70. <https://doi.org/10.1177/0265532212442634>.
- Zhang, H., & Bournot-Trites, M. (2021). The long-term washback effects of the National Matriculation English Test on college English learning in China: Tertiary student perspectives. *Studies in Educational Evaluation*, 68, 1–21. <https://doi.org/10.1016/j.stueduc.2021.100977>.

Publisher's Note

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