

RESEARCH

Open Access



When testing stakes are no longer high: impact on the Chinese College English learners and their learning

Qiuxian Chen^{1*} , Chenxian Hao² and Yangyu Xiao³

* Correspondence: chenqx@sxu.edu.cn

¹Foreign Language School, Shanxi University, 92# Wucheng Road, Xiaodian District, Taiyuan City 030006, Shanxi Province, China
Full list of author information is available at the end of the article

Abstract

It is now a consensus that high-stakes testing exerts powerful impact in controlling curriculum and pedagogy, and shaping students' learning. The power of testing is especially strong in contexts where testing culture prevails. The Chinese educational context is such a context which is dominated by competitive tests. The language testing results from large-scale tests, such as CET-4 (College English Test—Band 4), used to be included as a compulsory prerequisite for undergraduates to get their bachelor's degree. Such a requirement has been removed recently to reduce testing stakes for leverage purposes. Thus, it is worthwhile to investigate whether the change will possibly help to impact students and their learning in a positive way. Our study aimed to explore the impact of changes in the stakes of a language test in the Chinese university context. Data were collected through questionnaires from 544 College English students at a university in north-western China. Our analysis via SPSS 2.3 showed that test-taking was still well accepted and practiced; while reducing the stakes helped to redirect the students towards English learning, it had less impact on their English learning motivation and approaches. The findings also show that the test-dominated environment that is deep-rooted in the macro- and micro-contexts decided the students' responses to this policy change.

Keywords: CET-4, High-stakes testing, Chinese College English learners, Impact

Introduction

Standardized testing, which has been prevalent since the 1970s, is faced with sharp and poignant criticism for its “power” (Shohamy, 2007). This is mostly because testing is often attached with high stakes so as to make it an effective policy “tool” for transforming teaching and learning practice (Cheng 2018; Hamilton 2003). The focus on high-stakes testing, however, has entailed severe consequences, intended as well as unintended (Rea-Dickins and Scott 2007; Stobart 2003). The unintended consequences such as “narrowed curriculum,” “teaching to tests,” and “learning for tests” (Madaus 1988; Shohamy 1993) are admittedly major impetuses for ongoing educational assessment innovations which target at improved quality around the globe (Broadfoot 2007; Chen 2011; Scott et al. 2016).



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

While many an initiative enthusiastically embraces alternative assessment such as formative assessment (also referred to as assessment for learning with minor differences), policy efforts to reduce or remove the stakes of testing are observed in some contexts as well so as to redirect teachers', students', and other stake-holders' attention from tests to learning and its process. Wales in 2004, for example, abandoned the national testing regime for Key Stage 2 and 3, and England discontinued Key Stage 3 testing in 2009 to lever off the undesired washback effects (Collins et al. 2010; Daugherty 2008).

Likewise, the Chinese higher education sector via a press conference (Wu, 2005) explicitly cautioned misuses and overuses of the test results of College English Test Band 4 or 6 (CET-4/6), an external large-scale standardized testing system in China, in the education context, and beyond. In particular, it strongly advocated to abandon the use of test results as a compulsory prerequisite for undergraduates to get their bachelor's degree. Thus, the stakes of CET have been reduced and students can choose if they would like to take this test on their own initiatives.

Along with the stake-reducing initiative, changes were also made to the CET-4/6 itself in terms of test format (fewer multiple choice questions), test content (35% of listening instead of 20%), reporting system (a 710-point scoring system to replace the original 100-point one, transcripts of results to all test-takers instead of certificates to the passers only) (CET Syllabus Committee, 2005; Jin and Yang 2018).

The intended positive washback effect on College English teaching and learning was clearly articulated in both initiatives. Nonetheless, in the Chinese context where the high-stakes testing has been part of educational system and tradition, it would be interesting to investigate how stake-holders would respond to the change and more importantly, if the reduced stakes will exert a positive impact on College English learners as intended. These are the very questions this present study seeks to address.

Given that the major influences to the impact degree of a test include not only the stakes and the uses to which the test scores are put, but also the status of the subject matter tested (Shohamy et al. 1996), College English, the core concept of this study, is briefly set up below.

Context

College English course: an overview

College English has been a compulsory English as a foreign language (EFL) course designed especially for Chinese undergraduates who take majors other than English since the mid-1980s. As one of the many routine courses in the Chinese higher educational institutions, it enjoys a special position. This is clearly manifested in its nationally unified curricula and testing syllabi, which have evolved for five generations over the past three decades (Xu and Fan 2017). Until the fourth, *College English Curriculum Requirements* [CECR] (CMoE 2007), the course is prescribed to last for 2 academic years at least and takes up 16 credits in most universities, namely, about 10% of the credit points required of a bachelor's degree for over 95% of the Chinese undergraduates. Despite that the coming fifth curriculum, *College English Teaching Guidelines* (2017 draft) has tried to empower individual universities to reduce the credit points and to individualize their curriculum arrangements according to their actual situations, it is undeniable that College English is still a very important course for the target students.

This is, firstly, attributable to the still noticeable credit points of the course (in most cases 10 to 12 currently). Another reason is the wide acceptance of the CET-4/6 as an external evaluation benchmark of Chinese undergraduates' English proficiency achievements in stake-holders' ideology. As is known, the test results of CET-4/6 have been widely, beyond the initial intention of test-developers though, used in school and societal settings for multiple purposes such as a requisite for graduation and/or degree, further education, employment, and residential permit in big cities (Jin 2008, 2014) so that they are perceived as "key" to the Chinese undergraduates' success in academic, career, and even life in general (Cheng 2008). These overuses or misuses have entailed severe negative washback effects onto College English classrooms (Jin 2008; Yang 2015) and consequences of social dimension, which, in turn, made CET a rather "powerful" test, yet left test-designers "powerless" (Jin 2008).

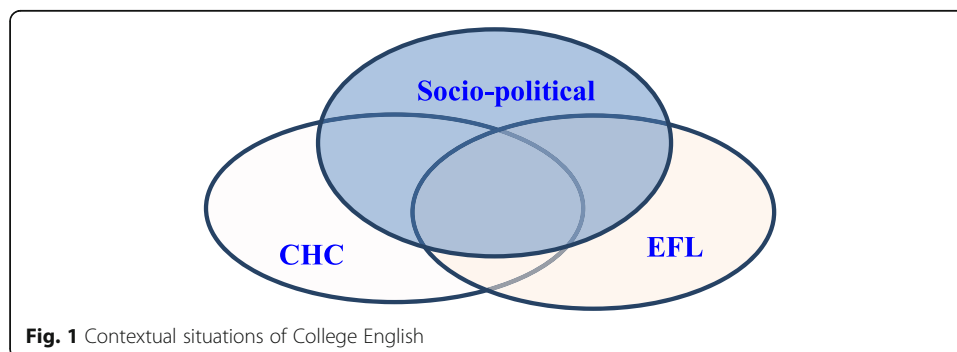
The present stake-reducing initiative is, with no doubt, an effort to counterbalance these consequences and an attempt to enhance the Chinese tertiary EFL learning outcomes. In response to the CMOE's advocacy to abandon these unintended uses, a number of universities adjusted their policies regarding the uses of test results. For example, CET-4 results were disconnected from the requirement of obtaining a bachelor's degree, and the test became optional rather than required to the students (Chen et al. 2013; Huang and Luo 2014).

The intended positive impacts, however, are not presumably to happen as a direct result of the changes to the test itself and its prestige (i.e., stakes and status of the test). Rather, it is "an interactive multi-directional process," which involves a constant interplay of varying factors pertaining to the macro-context (the specific society) and micro-context (school and university settings) as well (Tzagari and Cheng 2017 p.364). The macro and micro-context of this study hence follow.

Macro-contexts of the study

The macro-context that College English is situated in is tri-layered: the Confucian Heritage Cultural (CHC) context, the Chinese socio-political context, and the EFL disciplinary context (as illustrated in Fig. 1 below). Each layer of the contexts tends to foreground the high-stakes of testing.

To start with, China, as the birth place of CHC, has been embedded with its ideology since the Han Dynasty (206 BC–220 AD) (Cheng and Curtis 2010), when Confucianism has been exclusively revered and Confucian classics were adopted as the materials for



Kejǔ—a civil service examination system to select candidate for state offices. This millennium-old tradition strongly stresses the key role of examinations in education and the utilitarian values of its one-off result (Han and Yang 2001). Specifically, success in the examinations can bring multiple benefits such as fame, wealth, and higher social status to the examinees themselves and their families, whereas failure in the examinations might mean sheer poverty and humiliation (Chen 2016). These added-values, in turn, have reinforced the examination orientation and makes high-stakes testing a deep-rooted practice within the Chinese context.

The socio-political context in China similarly attaches considerable stakes to its tertiary EFL education. This is because higher education, as the elite rung of the educational ladder, is perceived by the Chinese government as an important contributing force to the nation's strength and identity, and its quality a key index of the nation's competitiveness in the global context of knowledge economy (Ouyang 2004). University graduates are accordingly expected to have sufficient subject matter knowledge, generic skills, and adequate English proficiency for instance, to meet the future needs (Mok 2018). For another, the increasingly global and competitive markets, along with the enrollment expansion since 1998, has made employment upon graduation really difficult (Wang 2007). Most often a degree from privileged universities and high-English-proficiency level (good CET-4/6 test results as a proof) mean good jobs and high salaries, whereas graduates from universities of low standards (Wu 2017) and low English proficiency (Cheng 2008) probably have to face up with unemployment or low-pay jobs. These expectations and practical issues have in time added a lot of stakes to the higher education in general and College English particularly in the contemporary China.

Discipline-wise, language testing is the only field of study in which “the measurement of a particular thing has evolved into a separate academic discipline” and developed into a worldwide business (Davison 2004, p.85). In this discipline, the use of large-scale proficiency tests for high-stakes purposes is a commonplace practice. Outer and expanding circle countries, such as Japan and Korea, often use tests as a policy tool to promote English learning outcomes and as gate-keeping devices for multiple purposes. Overall, the CET-4/6 and its high-stakes uses are of disciplinary characteristics, so are the negative washback, impact, and consequences. It is noteworthy, however, that the discipline, in the past decade, has taken a “social turn” and endeavored to incorporate alternative assessment (Davison and Leung 2009; Block 2003).

Context is important in that it constitutes not only the locale but also the “community of practice” for those within and shapes their ideology as well as their modes of behavior (Wenger 1998). This macro-context, where high-stakes is multi-folded and ideology about it deep-rooted, will most likely complicate this policy attempt to reduce the stakes, its implementation and effects.

Micro-context of the study

Regarding the micro-context of the study, this study was conducted in a comprehensive university located in an underdeveloped province of Midwest China. It belongs to the first tier of the Chinese higher educational institutions, but not one of the 211 or 985

Project, or Double First-ranking¹ universities. That is, it is one of the medium- to low-level universities among over 2000 Chinese universities. The researcher posits that students in such a university are those who are most seriously affected by the high stakes of the CET because the good test results could most probably act as a bonus for the scarce employment opportunities available to them.

This university, with a history of over 100 years, is one of the earliest universities within the country. Yet, limited by the location, the local development level, and its status within the Chinese higher educational circle, this inland university is overall not adaptive to the rapid educational changes. It enrolls around 3000 freshmen annually (the enrollment tends to increase these few years though), 95% of which are the target students of College English. By the time of data was collected in 2017², College English remains a compulsory course for undergraduate students of years one and two, and takes up 16 credit points required for a bachelor's degree, which means that College English still carries considerable weight for these students.

In response to the stake-reducing initiative, this university adjusted their policies regarding the uses of test results in 2006. The pegging practices of the CET certificates with bachelor's degrees and graduation were stopped and students were given the free choice of taking the CET or not at their own will. That is, the stakes were indeed reduced, as many other universities did (Huang and Luo 2014).

Meanwhile, this university innovated the internal assessment framework of College English by incorporating a "process assessment," which was made up of attendance, assignment, and classroom performance, and aimed to direct students' attention in some way to the learning process, an oral test, which was conducted by their teachers at the end of term. That is, students in this university, to earn the 16 compulsory credit points for their bachelor's degree, need to go through the internal College English assessment in every term of the first two academic years in accordance with the institutional assessment policy, which is composed of a process assessment (10%) over the period of a term, an oral test (10%) and a CET-formatted achievement test (80%) at the end of it.

While this study's agenda aims to examine how this changed status of stakes has impacted students in this university, it is instructive to know what previous research has revealed in this regard.

Impact of assessment change on students and their learning

Assessment, as "one of the defining features of students' learning" (Struyven et al. 2006, p. 326), has a considerable impact on students and their learning. Specifically, it is capable of conditioning students' learning goals (Marton and Säljö 1997), shaping students' approaches to learning (Boud and Falchikov 2007), the learning outcomes and even the development of their future learning (Black & Wiliam, 1998).

These influences are interrelated to each other. For example, when students study for an exam, they will try to understand the learning materials in ways that they perceive will meet requirements and adopt the most convenient approaches to maximize their grades. In comparison, when students study for learning, they will lay their eyes on

¹211, 985, and Double First-ranking are projects that Chinese government launched in, respectively, 1994, 1998, and 2016 to promote the quality of the Chinese higher education up to the top world level.

²The university began to reduce the credit points of College English to 12 in the fall term of 2018 as advised by The National Inspection Group from the Chinese Ministry of Education.

improvement rather than the grades and accordingly adopt high-level cognitive skills (Entwistle and Entwistle 1991). That is, orientation towards examination or learning could entail students to take different approaches to learning, and further surface- or deep-learning outcomes (Tang and Biggs 1996). Yet this link may not be straightforward; other factors such as whether students are equipped with appropriate strategies, and their personal agenda (Tang and Biggs 1996) also might interfere in-between. A small group of Chinese students in Jiang and Sharpling (2011), for example, demonstrated a variety of approaches to cope with a different assessment mode in UK universities, as a combined result of individual, contextual, pragmatic, and cultural influences.

Given that students are most often “alert cue-seekers” and good “examination game play[ers]” (Miller and Parlett 1974), changes to assessment policy such as reducing stakes in this study are potentially to effect corresponding changes in students’ learning behaviors. Relevant studies, however, show that the College English students were caught in a dilemma.

To change or not, that is a question

The impact of the reduced stakes of CET-4 was reported to, first of all, be psychologically. An investigation of five universities in the Northeast China (Wang 2007), for instance, found its impact on teachers and students in affective terms. Specifically, as these universities disentangled the CET-4 certificate from degree, and the students’ CET-4 passing rate from the salary of English teachers, students’ and teachers’ pressure from CET-4 were released to some extent. The administrators still paid close attention to the students’ marks and passing rate in the test, though. It seems time is needed for the change be accepted in both theory and mind, by stake-holders as well as administrators.

The impact of this initiative on EFL classrooms is found to be more complicated. It is, in some cases, rather pessimistic. Sun (2008), for example, investigated 180 students and 40 teachers in two universities in Central China with triangulated data sources, only to find surprisingly that both teachers and learners were willing to change on the surface only. That is, they would like to change on what to teach and what to learn, but not on how to teach and how to learn. A more recent survey (Gu and Xiao 2013) of 284 students similarly revealed that students’ strategy use remains oriented more towards testing rather than to learning.

In other cases, some light was shed on the bright side. Gu et al.’s (2013) investigation of 110 students’ learning experience via learning dairies, for instance, provided somewhat optimistic findings that reducing the stakes did influence students’ English learning in a positive way, but for a short term only. This is because students’ motivation for achieving a high score in English language tests, such as seeking for better opportunities for employment and further study, and other elements such as curriculum and teachers’ requirements, and the use of textbook as teaching materials, remain largely unchanged. This finding was echoed in Chen and May (2016), which showed that, despite the disconnection of CET-4 from bachelor’s degree, the use of its test results for other purposes such as scholarship or rewards retained, and its impact on students and their EFL learning was accordingly minimal.

While previous studies seem to show that the reducing the stakes of the test has not brought about the impact as intended yet, after a decade's implementation, further and deeper investigation is needed to investigate what exactly the change means to students and their English-language learning. The proposed aimed to investigate the perceptions of changes from a student perspective, because as the important target group of policy change, it is their response to the innovation that has the most say as to the success or failure of it and it is their experience of the innovation that holds the key to the complexity of it (Hamp 1997).

Methodology

Research questions

Specifically, this study has tried to address four questions:

- (1) How, if any, has reducing the stakes of CET-4/6 impacted on students' taking of the test?
- (2) Has reducing the stakes of CET-4/6 helped to redirect students' orientation more towards English learning? If yes, to what degree?
- (3) How, if any, has reducing the stakes of CET-4/6 impacted on students' motivation for English-language learning?
- (4) How, if any, has reducing the stakes of CET-4/6 impacted on students' approaches to English-language learning?

It is hoped that this investigation into students' test-taking practices, English learning orientation, motivation, and approach could shed some lights on the big question: to what degree has the stake-reducing initiative achieved its intent?

Research instrument

The data from the study were collected through questionnaires, which was adopted from Chen (2011). The questionnaire has been piloted to ensure that questionnaire items and formats would be interpreted consistently in the same way by all survey respondents. Statistics experts were also consulted to ensure its vigor and validity.

This study used selectively the two parts from the questionnaires that relate to CET-4/6. Part one consisted of six items regarding students' demographics including gender, major, year, students' self-reported English proficiency level, and the regions where they are from. Part two was five major questions regarding students' knowledge of the policy change, their test-taking status, and the possible impact of the policy change on their English learning orientations, motivations, and approaches towards the English learning. These questions, in turn, became the guiding route for data analysis. The questionnaire was designed on a 5-point Likert scale, with 1 = strongly disagree and 5 = strongly agree.

An ethical declaration was attached to the questionnaires for participants to know all the necessary information about the research and the researcher. It was also made clear that if they were not willing to fill in the questionnaire, they were free to do so without any consequences.

Participants

Considering the university's annual enrollment of 3000 students and target students of two years, questionnaires were distributed to 300 freshmen and 300 sophomores from different schools. Five hundred and sixty-two questionnaires were returned, of which 544 were valid with no significant missing values, with a return rate of 90.6%. Tables 1 and 2 below showcase the demographic information of these participants.

As shown in Table 1, the participants were more or less balanced in terms of genders (58.6% female, 41.1% male), majors (arts 47.6%, science 47.1%), and year of study (53.8% year one, 42.6% year two). Given the comprehensive nature of this university, these figures make a good sampling for this study.

When it comes to students' English proficiency level and development level of hometown, majority of the participants ranked their English proficiency as "medium" (343, 62.8%) and "low" or "very low" (182 students, 33.3%), which leaves only 19 student (3.5%) confident in their English proficiency. Those from economically undeveloped regions (290, 53.1%) or intermediate developed regions (189, 34.6%) account for the lion's share of the participants.

A cross-tabulation of the two variables (Table 2) shows a close correlation between educational level and economic development degree. That is, the lower the economic development degree of the region where the students came from is, the lower the English proficiency level that the students reported, and vice versa.

Students' low perception of their English proficiency, hence, may be explained by the fact that majority of the students are from the local province, which is located in an undeveloped Midwest China where the English language education is less developed. In addition, as the sampled university is a middle- to low-ranking university in the country, the language proficiency of students enrolled in this university is lower compared with top-tier universities in China. While EFL educational imbalance is much acknowledged within the country (CMoE 2007; Xiang et al. 2019), a point needs noting is that the percentage of "medium" selectors (63%) are even more noticeable. Certainly, this self-reported level could be students' fair judgment of their English proficiency. Yet the Chinese people's traditional psychology, which tends follow "the doctrine of the mean" (中庸之道 in Chinese) Hao (2016), could possibly be a reason as well.

Data analysis

The collected data were input into SPSS 22.0 in strict accordance with the SPSS coding and inputting procedures (Morgan et al. 2011) and were then analyzed. Descriptive

Table 1 Participants' profile

| Variables | Categories | Frequency | Percentage |
|-----------|------------|-----------|------------|
| Genders | Male | 226 | 41.4 |
| | Female | 320 | 58.6 |
| Majors | Arts | 260 | 47.6 |
| | Science | 257 | 47.1 |
| | Other | 29 | 5.1 |
| Years | One | 294 | 53.8 |
| | Two | 252 | 46.2 |
| | Total | 544 | 100 |

Table 2 Correlation between the development degree and English proficiency level

| | | English proficiency level | | | | | Total | | | | | | |
|-----------------------------------|--------------|---------------------------|------|--------|------|----------|-------|-----|-----|----|----|-----|------|
| | | Very high | High | Medium | Low | Very low | | | | | | | |
| Development degree of the regions | Developed | 1 | 0.2% | 4 | 0.7% | 37 | 6.8% | 19 | 3% | 6 | 1% | 67 | 12% |
| | Intermediate | 1 | 0.2% | 10 | 1% | 12 | 23% | 45 | 8% | 6 | 1% | 189 | 35% |
| | Undeveloped | 0 | 0 | 3 | 0.6% | 17 | 33% | 84 | 15% | 22 | 4% | 288 | 53% |
| Total | | 2 | 0.4% | 17 | 2.3% | 343 | 62.8% | 148 | 26% | 34 | 6% | 544 | 100% |

statistics were firstly employed to calculate the specific frequency distributions for all the questionnaire items, for the sake of an overview of the data results. As the present study was designed to extract the very factors with significant difference among multiple potential factors, and also independent variables used in this set of data such as genders, academic years, English proficiency levels are mostly nominal rather than ordinal in nature (Leech et al. 2015), a Multinomial *logistic regression model* was hence established. To illustrate more clearly the relationship between two groups of categorical variables, a *correspondence analysis* was also conducted. Results from the analysis were then used to inform the designed research questions, as shown below.

Results and discussions

Changed status of CET 4

As mentioned earlier, passing CET-4 is no longer a prerequisite for a bachelor’s degree in the university where this study was conducted since 2006. Students were supposed to know the changed status of CET-4. In the questionnaire, the question (I know that CET-4 is optional, not required by the school) was asked as the first question for a preliminary understanding of students’ knowledge of the change. The authors regard this enquiry as a prerequisite of the later investigation about impact of the reduced stakes on students’ learning, and hold that, without this knowledge, no impact could be expected.

Table 3 shows 290 (53.1%) of the surveyed students were clear or quite clear about the optional status of CET-4 in school, which in a sense indicates that the university surveyed had put the initiative into practice. Nonetheless, the fact that the students who were not clear or not certain about the change amounts up to 155 (28.4%) and 101 (18.5%) seemed to raise the concern that if the university has deliver the information to students effectively. It was also possible that since the policy was implemented nearly 10 years ago, and it was now no longer emphasized in classrooms, students may not be fully aware of such a policy. Whatever, the authors perceived this big share of students’ uncertainty a potential hindrance to the desired impact on students and their College English learning. Later analysis is based on this knowledge.

Table 3 Students’ knowledge of the CET-4’s changed status

| | 1 + 2 | | 3 | | 4 + 5 | | Total | | Mean | SD |
|------------|-------|------|-----|------|-------|------|-------|-----|------|-------|
| | n | % | n | % | n | % | n | % | | |
| Question 1 | 153 | 28.4 | 101 | 18.5 | 290 | 53.1 | 544 | 100 | 3.39 | 1.268 |

1 strongly disagree, 2 disagree, 3 neutral, 4 agree, 5 strongly agree

Impact on students’ test-taking practices

Whether this change of status would lead to students’ action in taking or not taking CET was investigated to address the first research question. The analysis of students’ response to the question (I will take CET-4 no matter the school requires or not) indicates that a majority of the surveyed students (478 or 87.6%) agreed or strongly agreed that they would take the CET-4 regardless of the change. Only 6.9% (38) of students who responded that they would choose not to take it if the university did not require it, and 5.5% (30) of the participants hesitated over whether to take it or not (see Table 4). Compared with the compulsory status of the CET-4, the percentages of 6.9% and 5.5% were certainly a resultant change of the initiative in this university. That is to say, a small proportion of the students, on realizing that they did not have to sit for the test if they did not want to, still chose to take the test.

To find out if and to what degree the students’ taking or not of CET-4 was related to their knowledge/ignorance of the changed requirements about the CET-4 in the university, a correspondence analysis was conducted and revealed an inverse correlation between the two (Fig. 2).

Figure 2 shows that, compared with those who did not know the change, those who knew that CET-4/6 was no longer a precondition for degree certification were more likely to disregard the change and still take the test. This suggested that the reasons for students taking or not taking the CET-4 were more complicated than expected. It is likely that their concern went beyond university requirements to something more important.

Students’ concern over the university requirements was verified in later inquires, which revealed that further employment, certifying their English level, and proving their own capacity were three major reasons accounting for the students’ taking the CET-4 (Table 5).

Table 5, while showcasing CET-4’s wide acceptance in the university settings as a valid indicator for students’ English proficiency level and a boost for students’ self-confidence, revealed that employment, the high stakes in socio-political settings, was the most important factor affecting the Chinese tertiary students and their test taking. Securing a better career opportunity is the reason why majority of the students continue to take CET-4 regardless of the change. In this sense, the impact of the reduced stakes on students’ test-taking practices was limited to a minority group; the majority students’ test-taking practices remained largely the same.

Impact on students’ English learning orientation

Whether students’ focus on English-language learning, after CET’s stakes were reduced, has changed from testing towards learning was investigated to address the second research question. The analysis of students’ responses revealed that around half of the participants (49.6%) tended to focus more on English learning itself after CET-4 became optional, which left 19.6% of the participants viewing that their English learning

Table 4 Students’ taking of CET-4

| | 1 + 2 | | 3 | | 4 + 5 | | Total | | Mean | SD |
|------------|----------|-----|----------|-----|----------|------|----------|-----|------|------|
| | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % | <i>n</i> | % | | |
| Question 2 | 38 | 6.9 | 30 | 5.5 | 478 | 87.6 | 546 | 100 | 4.26 | .981 |

1 strongly disagree, 2 disagree, 3 neutral, 4 agree, 5 strongly agree

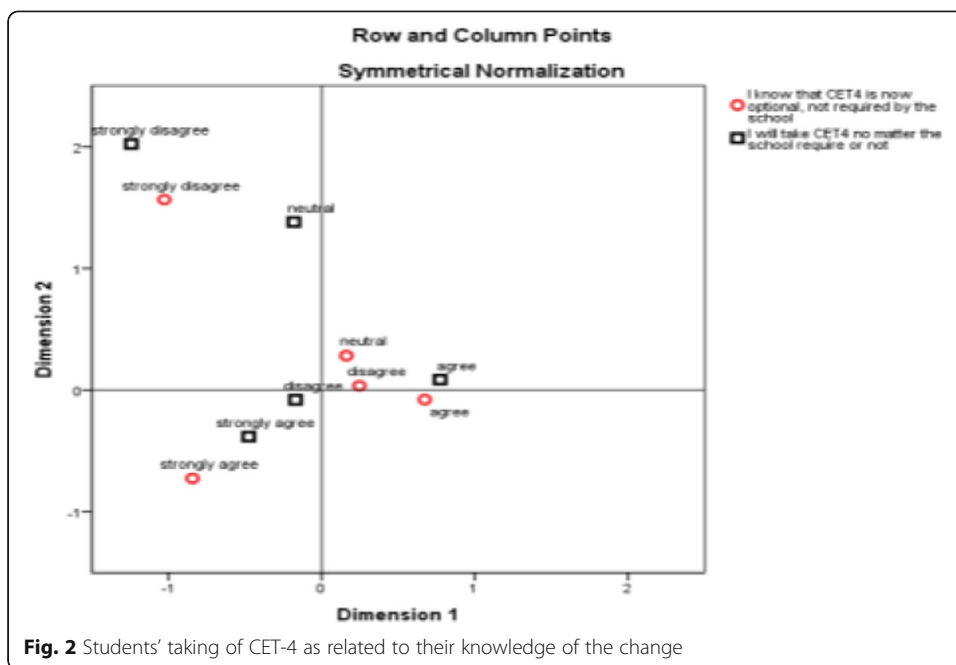


Fig. 2 Students' taking of CET-4 as related to their knowledge of the change

orientation has not been affected, and 30.8% of the participants responding neutrally to the question. Such being the case, then the reduced stakes indeed seemed to have affected at least half of the students' learning orientation in a positive way, as the policy intended to.

A logistic model was subsequently built to explore the factors that might influence students' responses to this question (Table 6).

The parameters show that factors such as majors, year of study, and genders have made no obvious statistical differences with the students' responses to this question ($\text{sig} > 0.05$), but students' English proficiency level was statistically significant with $\text{sig} < 0.05$ and the responding $\text{Wald } \chi^2 = 15.683$, $\text{Wald } \chi^2 = 6.536$. That means students' English proficiency level influenced their English learning orientation. A cross-tabulation (Table 7) was subsequently employed to further investigate the how students' English proficiency has influenced their English learning orientation.

As can be seen from Table 7, among 269 students (49.4%) who reported having been redirected towards learning, the medium- (53.6%) and low-achieving (43.4%) students

Table 5 Students' reasons for taking the CET-4

| Reasons | Response | | Percentage of cases |
|-----------------------|-----------|------------|---------------------|
| | Frequency | Percentage | |
| Habit | 40 | 5.5 | 7.4 |
| Prove myself | 136 | 18.9 | 25.1 |
| Useful for employment | 279 | 38.7 | 51.6 |
| Prove English level | 229 | 31.8 | 42.3 |
| Other | 37 | 5.1 | 6.8 |
| Total | 721 | 100 | 133.3 |

Note: For this question, students are allowed to choose as many as they find suitable from given items

were more affected than their high-achieving counterparts (31.5%). Moreover, more than half (52.6%) of the high-English-proficiency achievers tended to reserve their comments on this question, and 15.8% of them held against the possibility of being affected by the change.

This seems to resonate with Chen and May’s (2016) findings that mid- and low-achieving students were most affected by assessment policy change because a pass rather than fail mattered a lot for them; high-achieving students, on the other hand, usually achieved well in assessment of any kinds, and had no worry of the like. They hence were more likely to focus on learning and virtually unaffected by minor changes in assessment policy.

Impact on students’ English learning motivation

The question that “After CET-4/6 became optional, I am less motivated to learn English” was investigated to address the third research question, that is, whether and how the changing status of CET has led to any change in the students’ motivation of language learning. Participants’ responses to this question show that the optional status of CET-4 seemed to have resulted in some negative effects (Table 8).

Specifically, 13.7% of the participants explicitly stated that with this change, they became less motivated to learn English, and 21% of them hesitated or chose to reserve their opinions. Considering the originally compulsory nature and somewhat dominant position of College English in the Chinese higher education context, this was a rather noticeable impact. Fortunately, 64.9% of the participants opted for “disagree” or “strongly disagree”. That seemed to indicate that when passing the CET-4 and getting degree certificates were no longer the primary concern for the students, most of them

Table 6 Factors influencing students’ English learning orientation

| | | Estimate | Std. error | Wald | df | Sig. | 95% confidence interval | |
|-------------|----------------|----------------|------------|--------|------|------|-------------------------|-------------|
| | | | | | | | Lower bound | Upper bound |
| Threshold | Q3 = 1 | − 1.532 | .503 | 9.299 | 1 | .002 | − 2.517 | − .548 |
| | Q3 = 2 | .137 | .480 | .081 | 1 | .775 | − .803 | 1.077 |
| | Q3 = 3 | 1.616 | .485 | 11.100 | 1 | .001 | .665 | 2.567 |
| | Q3 = 4 | 3.248 | .498 | 42.558 | 1 | .000 | 2.272 | 4.224 |
| Location | Gender = 0 | .249 | .167 | 2.205 | 1 | .138 | − .080 | .577 |
| | Gender = 1 | 0 ^a | . | . | 0 | . | . | . |
| | Major = 1 | .128 | .370 | .119 | 1 | .730 | − .598 | .853 |
| | Major = 2 | .256 | .367 | .485 | 1 | .486 | − .464 | .976 |
| | Major = 3 | 0 ^a | . | . | 0 | . | . | . |
| | Year = 1 | .308 | .162 | 3.647 | 1 | .056 | − .008 | .625 |
| | Year = 2 | 0 ^a | . | . | 0 | . | . | . |
| | Eng pro = 1 | 22.539 | .000 | . | 1 | . | 22.539 | 22.539 |
| | Eng pro = 2 | .815 | .546 | 2.227 | 1 | .136 | − .255 | 1.885 |
| | Eng pro = 3 | 1.333 | .337 | 15.683 | 1 | .000 | .673 | 1.992 |
| Eng pro = 4 | .890 | .348 | 6.536 | 1 | .011 | .208 | 1.572 | |
| Eng pro = 5 | 0 ^a | . | . | 0 | . | . | . | |

Note: Eng pro, English proficiency; Link function, Logit
^aThis parameter is set to zero because it is redundant

Table 7 Students’ English proficiency level and their English learning orientation

| | | My English proficiency level | | | | | | Total | |
|------------|------------------------------|------------------------------|-------|--------|-------|----------------|-------|-------|-------|
| | | Very high + high | | Medium | | Low + very low | | | |
| Question 3 | Strongly disagree + disagree | 3 | 15.8% | 53 | 15.5% | 51 | 28% | 107 | 19.6% |
| | Neutral | 10 | 52.6% | 106 | 30.9% | 52 | 28.6% | 168 | 31% |
| | Agree + strongly agree | 6 | 31.5% | 184 | 53.6% | 79 | 43.4% | 269 | 49.4% |
| | Total | 19 | 100% | 343 | 100% | 182 | 100% | 544 | 100% |

were still motivated to learn English. Given the utilitarian ideology of the Chinese learners, the majority students’ maintained high motivation for English learning could be attributed to the fact that all the students still needed to pass a CET-4 formatted achievement test at the end of each term to earn credit points for their degree, and the fact that they might need the CET-4 results for practical purposes such as employment and the like; yet, it is also possible that these students were regarding improving their English proficiency level as a central goal for their English learning, which is the exact intent of the initiative.

A subsequent logistic model (Table 9) set up to further examine the mechanics of this impact revealed that none of the designed variables such as majors, year of study, genders, and students’ proficiency level, were significant in a statistical sense (sig > 0.05). This suggests that the impact of CET-4’s lower stakes on students’ motivation for English learning is not distinguishable in all these regards.

Impact on students’ approaches to English learning

The enquiry (CET-4’s becoming optional does not influence my approach to learn English) aims to address the fourth research question, that is, how, if any, the CET-4’s reduced stakes have affected students’ English learning approaches.

As Table 10 showcases, 24.7% of the students claimed they had adjusted their approaches to learn English in response to the change; 29.9% of them chose neutral position, yet much more students (45.4%) held that this change had exerted no influence on their English learning approaches. While the largest proportion of students’ responses seemed to agree with previous findings (Sun 2008; Gu and Xiao 2013) of little impact on students as to how to learn, the nearly a quarter of the sampled students’ positive response tends to point to the opposite. That is, the changed status of CET-4 seemed to have affected some, though not many, of the students’ approaches to English learning (Table 11).

Logistic model was again built to explore the possible influence of this result.

The parameter revealed that genders, year of study, and majors were not statistically significant factors and did not seem to make many differences in students’ responses regarding the impact of the change on their English learning approaches. The students’ English proficiency level, however, seemed to be relevant statistically (sig < 0.05, Wald

Table 8 Impact on students’ English learning motivation

| | 1 + 2 | | 3 | | 4 + 5 | | Total | | Mean | SD |
|------------|-------|------|-----|------|-------|------|-------|-----|------|------|
| | n | % | n | % | n | % | n | % | | |
| Question 4 | 354 | 64.9 | 116 | 21.4 | 74 | 13.7 | 544 | 100 | 2.29 | .986 |

1 strongly disagree, 2 disagree, 3 neutral, 4 agree, 5 strongly agree

Table 9 Factors influencing students’ English learning motivation

| | | Estimate | Std. Error | Wald | df | Sig. | 95% confidence interval | |
|-------------|----------------|----------------|------------|--------|----|------|-------------------------|-------------|
| | | | | | | | Lower bound | Upper bound |
| Threshold | Q4 = 1 | − 1.667 | .493 | 11.449 | 1 | .001 | − 2.633 | − .701 |
| | Q4 = 2 | .308 | .487 | .400 | 1 | .527 | − .646 | 1.262 |
| | Q4 = 3 | 1.563 | .494 | 10.025 | 1 | .002 | .595 | 2.530 |
| | Q4 = 4 | 3.839 | .584 | 43.206 | 1 | .000 | 2.694 | 4.983 |
| Location | Gender = 0 | .186 | .170 | 1.206 | 1 | .272 | − .146 | .519 |
| | Gender = 1 | 0 ^a | . | . | 0 | . | . | . |
| | Major = 1 | − .161 | .376 | .183 | 1 | .669 | − .899 | .577 |
| | Major = 2 | − .273 | .374 | .534 | 1 | .465 | − 1.006 | .460 |
| | Major = 3 | 0 ^a | . | . | 0 | . | . | . |
| | Year = 1 | − .056 | .164 | .117 | 1 | .733 | − .377 | .265 |
| | Year = 2 | 0 ^a | . | . | 0 | . | . | . |
| | Eng pro = 1 | − .050 | 1.334 | .001 | 1 | .970 | − 2.664 | 2.563 |
| | Eng pro = 2 | − .986 | .562 | 3.075 | 1 | .080 | − 2.088 | .116 |
| | Eng pro = 3 | − .300 | .337 | .791 | 1 | .374 | − .961 | .361 |
| | Eng pro = 4 | .283 | .351 | .653 | 1 | .419 | − .404 | .970 |
| Eng pro = 5 | 0 ^a | . | . | 0 | . | . | . | |

Note: *Eng pro* English proficiency; Link function, Logit
^aThis parameter is set to zero because it is redundant

$\chi^2 = 3.913$), and the relevance resided largely with those students who ranked their English proficiency level as “medium.” Why students with “medium” English proficiency were sensitive to the change goes beyond the scope of this study; this, nonetheless, seems make a new agenda for further investigation.

Conclusions

The above analysis of students’ response points to a conclusion that reducing the stakes of CET-4/6 by way of unpegging the test from degree has brought about impact on the Chinese College English learners and their EFL learning. The impact is to a certain degree positive as the policy intended. For instance, quite a few students (49.6%), particularly low-achieving and average students, reported having oriented themselves more towards English learning and nearly a quarter of them adjusted their learning approaches accordingly.

In some other cases, the stake-reducing initiative has entailed some unintended consequences as well. For example, some students (13.7%) became less motivated to learn English, half of which (6.9%) dropped the intention of sitting for the test at all, which, as the institutional administrators and teachers were much concerned about, might lead to reduced English learning outcome. This again happens mostly to the middle- and low-achieving students.

The overall scenario of the College English learning and testing in this university, however, remains largely unchanged. Specifically, nearly half (45.4%) of the students

Table 10 Impact on students’ English learning approach

| | 1 + 2 | | 3 | | 4 + 5 | | Total | | Mean | SD |
|------------|-------|------|-----|------|-------|------|-------|-----|------|-------|
| | n | % | n | % | n | % | n | % | | |
| Question 5 | 135 | 24.7 | 163 | 29.9 | 248 | 45.4 | 544 | 100 | 3.29 | 1.101 |

1 strongly disagree, 2 disagree, 3 neutral, 4 agree, 5 strongly agree

Table 11 Factors influencing students' English learning approaches

| | | Estimate | Std. error | Wald | df | Sig. | 95% confidence interval | |
|-------------|----------------|----------------|------------|--------|----|------|-------------------------|-------------|
| | | | | | | | Lower bound | Upper bound |
| Threshold | Q5 = 1 | −2.344 | .501 | 21.869 | 1 | .000 | −3.326 | −1.361 |
| | Q5 = 2 | −.644 | .479 | 1.809 | 1 | .179 | −1.582 | .294 |
| | Q5 = 3 | .667 | .479 | 1.941 | 1 | .164 | −.271 | 1.605 |
| | Q5 = 4 | 2.313 | .489 | 22.382 | 1 | .000 | 1.355 | 3.271 |
| Location | Gender = 0 | −.020 | .166 | .014 | 1 | .905 | −.345 | .306 |
| | Gender = 1 | 0 ^a | . | . | 0 | . | . | . |
| | Major = 1 | −.086 | .369 | .054 | 1 | .816 | −.810 | .637 |
| | Major = 2 | .140 | .366 | .147 | 1 | .702 | −.578 | .859 |
| | Major = 3 | 0 ^a | . | . | 0 | . | . | . |
| | Year = 1 | −.107 | .160 | .441 | 1 | .507 | −.421 | .208 |
| | Year = 2 | 0 ^a | . | . | 0 | . | . | . |
| | Eng pro = 1 | 1.792 | 1.338 | 1.795 | 1 | .180 | −.830 | 4.414 |
| | Eng pro = 2 | .757 | .545 | 1.930 | 1 | .165 | −.311 | 1.824 |
| | Eng pro = 3 | .656 | .332 | 3.913 | 1 | .048 | .006 | 1.306 |
| | Eng pro = 4 | .315 | .345 | .835 | 1 | .361 | −.360 | .990 |
| Eng pro = 5 | 0 ^a | . | . | 0 | . | . | . | |

Note: *Eng pro* English proficiency. Link function, Logit

^aThis parameter is set to zero because it is redundant

still learned English in the way they are used to, a majority of students (64.9%) were still highly motivated to learn English as they used to, and most importantly, most of the students (87.6%) still took the test regardless of the change. Overall, it seems the policy intent is realized only in a limited degree.

The students' relatively stable learning approaches are understandable, considering the CET-formatted achievement test still took the lion's share (80%) in the internal assessment framework of College English, a coursework that was of significance to their academic achievement within the university and degree. Despite the changes to assessment, teachers were required to keep to a unified syllabus and curriculum arrangements prescribed by the school and to use appointed textbooks as they used to. They were not retrained in terms of assessment and pedagogy either when these changes were made. It is no surprise, then, the students' learning approaches remained stable and changed only slightly to cope with the changes.

Students' continued high motivation to learn English and sit for the CET-4/6 regardless of the reduced stakes is, first of all, attributable to the weights of the College English within the micro-context. As mentioned previously, students in the university need to earn the 16 credit points required of this coursework for their bachelor's degree. A second reason could be the significance of the CET-4/6 certificates at the societal context for employment and other practical purposes, as evidenced in this and other studies (Chen and May 2016; Gu et al. 2013). The third and the most important reason, in the authors' view, resides with the established mindset and practice in the macro-contexts, such as the emphasis on examination and its results, the utilitarian uses of the results of the CHC ideology (Han and Yang 2001; Poole, 2016), the elevated status of English as an international language and acceptance of certificates as proof of one's English proficiency in the socio-political area, and the yet-to-establish alternative

assessment methods within the EFL discipline. In other words, the initiative itself has reduced stakes from the micro-context alone and from the external assessment only, whereas the stakes were still high with the internal assessment framework and remained untouched within the tri-layered macro-contexts. That is, the initiative was virtually a slight touch on the tightly knitted web, which is working with systems and mechanics, which all the stake-holders and the society at large have been accustomed to and accepted. It might have caused some waves to the web, yet far from changing it in any substantial way. It is, therefore, no wonder to see little impact on students' motivation and practice regarding the CET-4/6.

Nearly half of the students' admitted shift of orientation from testing to learning, nonetheless, seems to shed some lights on the prospects of College English assessment innovation in the future, since it shows students, in a sense, were willing to divert from the examination-orientation practices and focus on improving their English proficiency instead when possible.

Implications and limitations

Findings and conclusions as such hold rich implications for policy-making and pedagogy in the classroom. To start, for assessment used as an effective tool for significant improvement in students' English learning behaviors and learning outcomes, policy-making at the top/national level needs to take into account not only the assessment itself, but also the social-political and cultural context it is situated in. These contextual influences and the resultant consequences might as well swallow up most of the positive possibilities, as findings of this study showcased. In addition, to level off the negative washback effects of high-stakes testing, stakes need to be totally eliminated, at least in the internal assessment system. Otherwise, most students are not likely or could not afford to ignore it, only if stakes are there—reduced or not.

Since students like to orient themselves away from testing and towards English learning, teachers need to adopt a pedagogy which encourages the orientation in and out of the classroom. More importantly, teachers are obliged to show students how to better their English for the sake of learning rather than for the testing only. Otherwise, they might fall back on the English learning approaches they are used to.

This study is limited in its data, which was collected via a survey in a single university. Data of wider coverage and varied sources such as qualitative interview could have been employed for a deeper and more comprehensive investigation into, for instance, why the students responded as they did, and how their learning orientation, motivation, and approach interacted with each other. These, in turn, become our research agenda in the coming future.

Abbreviations

CET: College English test; EFL: English as a foreign language; CECE: College English Curriculum Requirements; CHC: Confucian heritage cultural

Acknowledgements

This study is based on a larger research project sponsored by The National Social Science Fund of China [NSSF] (no.17BYY100). Comments and suggestions from Professor Lawrence Jun Zhang at the SSCI (IV) Conference are well-appreciated as well.

Authors' contributions

The first author was responsible for the design, data collection, and most of the writing up of the paper. The second author had helped with data input and analysis, and literature review part. The third author helped to proofread and polish. The authors read and approved the final manuscript.

Authors' information

Dr. Chen Q. is a professor of Applied Linguistics in the Foreign Language School of Shanxi University. With a Ph.D. degree from Queensland University of Technology, Australia, she is particularly interested and expertised in the EFL education and assessment in China and other CHC contexts.

Ms. Hao C. won her Master's Degree in the field of Language Testing and Assessment from the Foreign Language School of Shanxi University. She now works in Shanxi Transportation Instruction Inspection Center as a clerk and interpreter.

Dr. Xiao Y. is now a lecturer in English for Academic Purposes in The Chinese University of Hong Kong, Shenzhen. She obtained her Ph.D. in English language education and assessment from The University of Hong Kong. Her publication and research interests are in the fields of English language assessment and testing, English language curriculum, language teacher education, and discipline-specific language learning.

Funding

This study is based on data collected for a project funded by the National Social Science Fund of China [NSSF] (no.17BYY100). NSSF has approved the project and its design. The use and interpreting of this set of data and writing up of this manuscript is the authors' responsibility. Yet the funding agency strictly forbids using the money for journal publication charge.

Availability of data and materials

The data and its backup were stored in the authors' PC and other digital storing devices. The data were declared in ethical approval to be shared by researchers involved only. Editors or reviewers, whom we consider as involved, can have access to the data on request.

Competing interests

No, there are no conflicting interests. The author order was agreed by all the three authors.

Author details

¹Foreign Language School, Shanxi University, 92# Wucheng Road, Xiaodian District, Taiyuan City 030006, Shanxi Province, China. ²Shanxi Transportation Instruction Inspection Center, Taiyuan, China. ³School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen, China.

Received: 12 December 2019 Accepted: 23 April 2020

Published online: 02 July 2020

References

- Black, P., & William, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Block, D. (2003). *The social turn in second language acquisition*. Washington: Georgetown University Press.
- Boud, D., & Falchikov, N. (2007). Assessment for the longer term. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in Higher Education: Learning for the longer term* (pp. 3–25). New York: Routledge.
- Broadfoot, P. (2007). *An introduction to assessment*. London: Continuum.
- CET Syllabus Committee. (2005). The CET4/6: Syllabus and sample. Shanghai: Shanghai Foreign Language Education Press.
- Chen, Q. (2016). The cultural appropriateness and appropriation of formative assessment in the Chinese context. *Journal of Shanxi University*, 39(3), 80–90. [https://doi.org/10.13451/j.cnki.shanxi.univ\(phil.soc\).2016.03.011](https://doi.org/10.13451/j.cnki.shanxi.univ(phil.soc).2016.03.011) (In Chinese).
- Chen, Q. X. (2011). *Assessment policy change in relation to English language teaching and learning in China: A study of perspective from two universities (Unpublished doctoral dissertation)*. Brisbane: Queensland University of Technology.
- Chen, Q. X., Kettle, M., Klenowski, V., & May, L. (2013). Interpretation of formative assessment in the teaching of English at two Chinese universities: A sociocultural perspective. *Assessment & Evaluation in Higher Education*, 38(7), 831–846. <https://doi.org/10.1080/02602938.2012.726963>.
- Chen, Q. X., & May, C. L. (2016). Chinese EFL students' response to an assessment policy change. In G. X. Yu & Y. Jin (Eds.), *Assessing Chinese learners of English: Language constructs, consequences and conundrums* (pp. 199–218). London: Palgrave MacMillian.
- Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, 25(1), 15–37. <https://doi.org/10.1177/0265532207083743>.
- Cheng, L. (2018). Geopolitics of assessment. In J. I. Lontos (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–12). John Wiley & Sons, Inc.
- Cheng, L., & Curtis, A. (Eds.). (2010). *English language assessment and the Chinese learner*. New York: Routledge.
- China Ministry of Education. (2007). College English curriculum requirements. Retrieved from <http://www.moe.edu.cn/edoas/website18/info34295.htm>.
- Collins, S., Reiss, M., & Stobart, G. (2010). What happens when high-stakes testing stops? Teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales. *Assessment in Education: Principles, Policy & Practice*, 17(3), 273–286. <https://doi.org/10.1080/0969594X.2010.496205>.
- Daugherty, R. (2008). Reviewing national curriculum assessment in Wales: How can evidence inform the development of policy? *Cambridge Journal of Education*, 38(1), 73–87. <https://doi.org/10.1080/03057640801889998>.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305–334. <https://doi.org/10.1191/0265532204lt286oa>.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415. <https://doi.org/10.1002/j.1545-7249.2009.tb00242.x>.
- Entwistle, N., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: The student experience and its implications. *Higher Education*, 22(3), 205–227. <https://doi.org/10.1007/BF00132288>.
- Gu, X., & Xiao, W. (2013). CET washback on the test-taking strategy use of non-English majors in China. *Foreign Language Testing and Teaching*, 1, 30–38. In Chinese.

- Gu, X., Yang, Z., & Liu, X. (2013). A longitudinal study of the washback of CET on college English classroom teaching and learning—Three College English teachers' classrooms revisited. *Foreign Language Testing and Teaching*, 1, 18–29 [In Chinese].
- Hamilton, L. (2003). Assessment as a policy tool. *Review of Research in Education*, 27(1), 25–68 <https://doi.org/10.3102/0091732X027001025>.
- Hamp, L. L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295–303 <https://doi.org/10.1177/026553229701400306>.
- Han, M., & Yang, X. (2001). Educational assessment in China: Lessons from history and future prospects. *Assessment in Education: Principles, Policy & Practice*, 8(1), 5–10 <https://doi.org/10.1080/09695940120033216>.
- Hao, C. (2016). *The washback effect of the innovated CET4*. (Unpublished Master thesis). Taiyuan: Shanxi University.
- Huang, J., & Luo, S. (2014). Formative assessment in L2 classroom in China: the current situation, predicament and future. *Indonesian Journal of Applied Linguistics*, 3(2), 18–34.
- Jiang, X., & Sharpling, G. (2011). The impact of assessment change on language learning strategies: The views of a small group of Chinese graduate students studying in the UK. *Asian EFL Journal*, 13(4), 33–68 <https://www.asian-efl-journal.com/main-journals/>.
- Jin, Y. (2008). Powerful tests, powerless test designers? Challenges facing the College English Test. *CELEA Journal*, 31(5), 3–11 <https://doi.org/10.1016/j.healthpol.2007.08.006>.
- Jin, Y. (2014). The limits of language tests and language testing: Challenges and opportunities facing the College English Test. In *English Language Education and Assessment* (pp. 155–169). Singapore: Springer.
- Jin, Y., & Yang, H. (2018). Language testing of Chinese Characteristics: CET-4/6 of the past three decades and its implications. *Foreign Language World*, 02, 29–39 [In Chinese].
- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2015). *SPSS for Introductory and Intermediate Statistics: IBM SPSS for Introductory Statistics Use and Interpretation*. New York: Routledge.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum: English-seventh yearbook of the National Society for the Study of Education* (pp. 83–121). Chicago: University of Chicago Press.
- Marton, F., & Säljö, R. (1997). Approaches to learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (pp. 39–58). Edinburgh: Scottish Academic Press.
- Miller, C. M. I., & Parlett, M. (1974). *Up to the mark: A study of the examination game*. Guildford: Society for Research into Higher Education.
- Mok, K. H. (2018). Promoting national identity through higher education and graduate employment: Reality in the responses and implementation of government policy in China. *Journal of Higher Education Policy and Management*, 40(6), 583–597 <https://doi.org/10.1080/1360080X.2018.1529127>.
- Morgan, G. A., Leech, N. L., Gloeckner, G. W., & Barrett, K. C. (2011). *SPSS for Introductory and Intermediate Statistics: IBM SPSS for Introductory Statistics Use and Interpretation*. New York: Routledge.
- Ouyang, O. (2004). Higher education reform in China today. *Policy Futures in Education*, 2(1), 141–149 <https://doi.org/10.2304/pfie.2004.2.1.4>.
- Poole, A. (2016). 'Complex teaching realities' and 'deep rooted cultural traditions': Barriers to the implementation and internalisation of formative assessment in China. *Cogent Education*, 3(1), 1156242. <https://doi.org/10.1080/2331186X.2016.1156242>.
- Rea-Dickins, P., & Scott, C. (2007). Washback from language tests on teaching, learning and policy: Evidence from diverse settings. *Assessment in Education: Principles Policy & Practice*, 14(1), 1–7 <https://doi.org/10.1080/09695940701272682>.
- Scott, S., Scott, D. E., & Webber, C. F. (2016). *Assessment in education: Implications for leadership*. Heidelberg: Springer.
- Shohamy, E. (1993). The power of tests: The impact of language tests on teaching and learning. *NFLC Occasional Papers*.
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education: Principles Policy & Practice*, 14(1), 117–130 <https://doi.org/10.1080/09695940701272948>.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298–317 <https://doi.org/10.1177/026553229601300305>.
- Stobart, G. (2003). Editorial. The impact of assessment: Intended and unintended consequences. *Assessment in Education: Principles, Policy & Practice*, 10(2), 139–140 <https://doi.org/10.1080/0969594032000121243>.
- Struyven, K., Dochy, F., Janssens, S., Schelfhout, W., & Gielen, S. (2006). The overall effects of end-of-course assessment on student performance: A comparison between multiple choice testing, peer-assessment and portfolio assessment. *Studies in Educational Evaluation*, 32(3), 202–222 <https://doi.org/10.1016/j.stueduc.2006.08.002>.
- Sun, H. (2008). *Intention and reality: An investigation of Washback effect of the new CET4 on College English teaching and learning*. China: Master Thesis. Central South University.
- Tang, C., & Biggs, J. (1996). How Hong Kong students cope with assessment. In D. A. Watkins & J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 159–182). Hong Kong and Melbourne: The Central Printing Press.
- Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of Language and Education* (3rd ed., pp. 359–372). Cham: Springer.
- Wang, J. (2007). The College English Test in China: Challenges and suggestions. *Asian Journal of English Language Teaching*, 17, 137–144 [In Chinese].
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.
- Wu, Q. (2005). Talk on CET-4/6 reform press conference by Wu Qidi, Deputy Minister of CMOE. *Foreign Language World*, 2, 2–4 [In Chinese].
- Wu, X. (2017). Higher education, elite formation and social stratification in contemporary China: Preliminary findings from the Beijing college students panel survey. *Chinese Journal of Sociology*, 3(1), 3–31 <https://doi.org/10.1177/2057150X16688144>.
- Xiang, L., Stillwell, J., Burns, L., & Heppenstall, A. (2019). Measuring and assessing regional education inequalities in China under changing policy regimes. *Applied Spatial Analysis and Policy*, 1–22 <https://doi.org/10.1007/s12061-019-09293-8>.
- Xu, J., & Fan, Y. (2017). The evolution of College English curriculum in China (1985–2015): Changes, trends and conflicts. *Language Policy*, 16(3), 267–289 <https://doi.org/10.1007/s10993-016-9407-1>.
- Yang, H. Z. (2015). Valid testing, effective teaching, and valid test use. *Journal of Foreign Languages*, 38(1), 2–26 [In Chinese].

Publisher's Note

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