



Received: 06 May 2016
Accepted: 16 August 2016
Published: 31 August 2016

*Corresponding author: Chuqiao Zhao, Graduate School of Education, University of Pennsylvania, Philadelphia, PA, USA
E-mail: chuqiaoz@gse.upenn.edu

Reviewing editor:
Kris Gritter, Seattle Pacific University, USA

Additional information is available at the end of the article

EDUCATIONAL ASSESSMENT & EVALUATION | REVIEW ARTICLE

Motivational effects of standardized language assessment on Chinese young learners

Chuqiao Zhao^{1*}

Abstract: This review paper examines how standardized language assessment affects Chinese young learners' motivation for second-language learning. By presenting the historical and contemporary contexts of the testing system in China, this paper seeks to demonstrate the interrelationship among cultural, social, familial, and individual factors, which collectively contribute to test impacts on young learners' motivation. In order to call for future research on standardized language assessment's motivational effects on Chinese young learners, this paper synthesizes findings and arguments from established research on learners of different age groups in different social contexts, which serve as a theoretical foundation for emerging studies on Chinese young test takers. The author concludes that Chinese young learners' motivation for second-language learning is shaped by the sociocultural and historical context of the testing system, subject to age-related characteristics, and vulnerable to adults' attitudes toward test results.

Subjects: Educational Change & School Reform; Educational Research; Testing, Measurement, and Assessment

Keywords: standardized language assessment; motivation; Chinese young learners; second-language learning

1. Introduction

In the past two decades, standardized language assessment as an interdisciplinary domain has continuously invited controversy over its pros and cons among scholars in the fields of educational measurement and educational linguistics (e.g. Flores & Schissel, 2014; Llosa, 2011; McNamara, 2001; Taylor, 1994, etc.). Whereas the practicality of test administration and the quantifiability of test results are considered as advantages, standardized language assessment is often accused of its high-stakes characteristics, which tend to induce test anxiety and therefore impede test takers'

ABOUT THE AUTHOR

Chuqiao Zhao graduated from the Educational Linguistics division at the University of Pennsylvania, Graduate School of Education. His research interests include classroom and policy uses of standardized language assessment, language planning and policy among indigenous and minority language speakers, and bilingual education in sociolinguistic perspective.

PUBLIC INTEREST STATEMENT

The use of standardized language assessment among young learners has been increasingly prevalent and controversial in China. By reviewing existing literature on multiple aspects of standardized tests, this paper seeks to provide a fuller understanding of how current high-stakes testing practices influence Chinese children's motivation for second-language learning. Viewing motivation as a dynamic and fluid construct, the author argues that test motivational effects on Chinese children can be enhanced by carefully examining various developmental, social, historical, and cultural factors.

motivation for future studies (Cheng et al., 2014; Wren & Benson, 2004). Particularly in China and some other Asian countries, there has been an increasingly prevalent use of standardized English tests to assess young learners' (YLLs)¹ language proficiency, along with the trend of teaching English to younger and younger children (Butler, 2014, 2015). Although such tests seem to have motivated parents to have their children receive early English education, it remains uncertain whether they can contribute to YLLs' motivation for their own second-language (L2) learning.

Motivation has been widely recognized as one of the key factors that influence the success of L2 learning by both teachers and researchers (Dörnyei, 1998). Indeed, it provides learners with the primary impetus to initiate their language learning and sustain the long-term learning process over time. As Gardner and Lambert (1972) claim, learners' L2 motivation can often override their aptitude effects, even though the latter may largely account for individual differences in learning achievement. Building on this argument, Dörnyei (1998) asserts that (1) even individual learners with the most remarkable language aptitudes can hardly accomplish their long-term goals in L2 development without motivation, and that (2) even the best possible curriculum and teaching method cannot substitute a strong L2 motivation. Hence, aware of the vital importance of motivation in affecting L2 learning progress and achievement, I argue that it is essential for teachers and policy-makers to take into account how standardized language assessment influences YLLs' L2 motivation in order to make informed decisions for test development and administration.

Regarding motivation as a crucial factor for YLLs' L2 learning is especially meaningful in China's education system. Since the high-stakes consequences brought by standardized language tests received scholars' examination and criticism internationally (Flores & Schissel, 2014; Menken, 2008), China has been declaratively promoting less test-driven and more student-centered education reform in grade schools (Liu & Dunne, 2009; Lockette, 2012). Accordingly, the reform is supposed to avoid putting YLLs' motivation at risk, due to its central role in guiding learners toward their academic goals. However, according to Liu and Dunne (2009), all educational decisions in Chinese schools still focus on the standardized testing system. Moreover, students are pressured to take numerous large-scale and high-stakes tests as soon as they start their schooling (Qi, 2004), whereas their motivation for learning is often overlooked. This paradox between policy-makers' announcements and real-life educational practices rings the alarm to educators who are concerned with YLLs' learning progress in the long run. Further, I hypothesize that compared with older learners, YLLs are more vulnerable to negative motivational effects of tests. If this hypothesis holds true, it will be necessary for researchers, policy-makers, educators, and parents to provide YLLs with extra support to strengthen their motivation and make adaptations to current test uses.

Although efforts have been made to investigate the impacts of some powerful standardized tests for secondary schoolers or adults, such as the National Matriculation English Test (Cheng & Qi, 2006; Qi, 2004), very few studies have examined how standardized tests influence the motivation of young children, especially elementary students. In actuality, however, the Middle School Entrance English Test (MSEET) for many elementary students can be as demanding as the college entrance tests for high school graduates. It should be noted that MSEET is not a nation-mandated standardized test for advancement in the public education system. Yet, it has been administrated for a long time and become popular among students who plan to enter prestigious private middle schools. Test formats and items are typically determined by those local private schools, while test difficulty and intensity can vary according to each district. Further, those private schools regard test scores as the primary criteria for high-stakes enrollment decisions, thereby exacerbating the sense of competition among YLLs. For example, as *China News* reported on 4 July 2015, approximately 2,530 children in Nanjing city took the MSEET held at Nanjing Foreign Language School, competing for 320 available positions for incoming students. Similar reports can be easily found on all kinds of official news websites every year, along with pictures that capture thousands of children's parents and grandparents waiting outside the test centers. In sum, although elementary schoolers are not required to take any

nation-mandated test for graduation, the strong parental and familial desire for sending children into prestigious middle schools have resulted in various local implementations of standardized assessments, which in turn affect YLLs' learning motivation.

Due to teachers' frequent use of standardized assessment and parents' high expectations for test outcomes, the debilitating effects of academic stress are commonly perceived by Chinese YLLs (Zhao, Selman, & Haste, 2015). Likewise, Hesketh et al. (2010) conducted a survey among 2,191 Chinese children of 9–12 years old and found that 81% of the participants were frequently worried about their test results of summative assessments; more than half of them were afraid of being punished by parents or teachers for test failures and became less motivated by learning itself. Since a large-scale, high-stakes test like MSEET is much more intense and anxiety-inducing than regular summative assessments at the end of each semester, its motivational effects on YLLs are highly likely to be more negative. Still, there is little empirical data elicited from Chinese YLLs to reveal how standardized language tests, such as MSEET, affect their motivation for English learning. What has been clearly identified is that both teachers and parents of YLLs in China have associated high-stakes tests with the possibilities of their children's future (Cheng, 2008), even if the mass media constantly report how much burden of test preparation has been imposed on YLLs. Indeed, large-scale and high-stakes tests continuously receive warm support from the public as fair measurements for selecting the best learners into the social hierarchy (Cheng & Qi, 2006), or as cultural symbols of "order and control, desired educational outcomes, and traditional moral values." (Airasian, 1988, p. 301) Because of these symbolic aspects, which I will discuss in detail later, test motivational effects become even more complex and dynamic.

Up until now, quite a few scholars in Western countries have taken into account test impacts on YLLs' L2 development and academic achievement (McKay, 2006; Menken, 2008; Wortham, 2005), intended and unintended effects on teaching and learning (Cheng & Curtis, 2004; Qi, 2004), as well as critical and sociopolitical dimensions (Lynch, 2001; McNamara & Roever, 2006; Shohamy, 1998). Nonetheless, a narrowly focused research on the motivational effects of standardized English tests on Chinese YLLs has yet to be conducted. Furthermore, studies that explicitly explore the relationship between standardized language testing and L2 motivation in Asian contexts remain scarce (Haggerty, 2011; Haggerty & Fox, 2015). Additionally, test impacts perceived by YLLs themselves have not been adequately investigated. In other words, little emphasis has been given to specific *washback* effects² perceived by learners themselves (Watanabe, 2004). Therefore, whether to maintain, modify, or remove standardized language tests for YLLs is largely debatable, owing to the exclusion of the most direct stakeholders' voices. In order to partially fill in this gap of inquiry, I selected and reviewed 59 existing studies from theory-based edited volumes and context-specific journal articles, whose research foci include contextual factors of standardized assessment in China, assessing children's language ability, and test washback effects on L2 learning. By highlighting these interrelated factors above, I expect that this review paper will contribute to a more comprehensive understanding of historical and current testing practices in Chinese school contexts, thereby inviting international scholars to participate in the discussion on how to strategically help Chinese children maintain their L2 motivation. I first give a general overview of the historical and contemporary contexts of standardized assessment in China and discuss its symbolic aspects. Second, I justify the hypothesis that YLLs tend to be more vulnerable to test demotivation than older learners by addressing some key considerations about YLLs' language assessment. Third, I adopt Dörnyei's (2005, 2009) framework of *L2 Motivational Self System* as a model to analyze the relationship between L2 motivation and testing experiences. I conclude that there is an important distinction between YLLs and older learners in terms of their susceptibility to test motivational effects, and that YLLs are particularly vulnerable to test demotivation due to contextual factors, age-related characteristics, and adults' attitudes toward their test results.

2. Cultural motivation and the evolution of standardized assessment in China

In this section, I begin by tracing the origin of standardized assessment and its primary function in ancient China. Next, I build on Hu's (2003) characterization of the standardized testing system in modern China, highlighting the fact that there has been a long-established criticism about the negative impacts of standardized assessment on test takers. As I illustrate below, such criticism coexists with societal acceptance and public support for continued implementation of large-scale, high-stakes assessments. Indeed, I find that standardized assessment in China has been culturally symbolized as a fair measurement for selecting talented people over time. As a result, Chinese test takers are generally situated in and influenced by a testing culture. Standardized language assessment, as a subset of standardized assessment, is assumed to be subject to common contextual factors that affect standardized assessment at large. Thereafter, I argue that YLLs' learning motivation is closely related to and shaped by the testing culture, and that the test-driven motivation is shown to be dynamic and complex across generation.

China has been historically known for the tremendous influence of examinations on its educational system (Qi, 2004). Many scholars believe that Chinese test-oriented and outcome-based education has existed for more than 1,000 years, since the earliest large-scale, standardized assessment (i.e. *keju*) was used to select civil officials for administration purposes in the Han Dynasty (Chen, 2011; Cheng & Curtis, 2004; Lockette, 2012). According to Chen (2011), the continuous practice of large-scale, high-stakes assessments has strengthened the utilitarian values of education and emphasized the one-off result of testing. As a consequence, extended practice of standardized assessment has contributed to a socially and culturally accepted testing tradition, which in turn shapes the Chinese culture of education and social life. Throughout the history, most Chinese test takers have been taught by teachers and parents to make every effort to succeed in large-scale, high-stakes assessments from their early childhood. Moreover, a number of mottos has been continuously used by adults to motivate YLLs to be diligent learners and competitive test takers. For example, *shí nián hán chuāng wú rén wèn, yī jǔ chéng míng tiān xià zhī* (Nobody cares about your 10 years of sweat and swot, but the world will know you once you succeed [from the standardized assessment]). While such mottos convey adults' utilitarian values about education, it is critical to recognize that the extrinsic motivation associated with fame and fortune brought by test success has greatly influenced Chinese families and their perceptions of education and assessment. Embedded in cultural values that are routinely instilled into human minds, YLLs' motivation to study for tests is inevitably affected by their parents' and teachers' attitudes toward their test performances. Thus, YLLs' learning motivation should be viewed as ultimately related to and shaped by the testing culture.

In contemporary time, the standardized testing system still plays a fundamental role in affecting the operation of the education system, which is characterized as "centralization, stratification, and selection" (Hu, 2003). First, centralization refers to the central role of the Chinese Ministry of Education in policy-making, including appointing textbooks, setting standards for a national curriculum, administering national matriculation tests, and so forth. Thus, standardized testing is frequently used as a policy tool, thereby revealing its sociocultural impacts from a top-down approach without paying close attention to whether test takers are motivated to learn by the testing system. Second, stratification refers to the regional gaps in terms of education investment, which is correlated with local economic development of different districts. As a result, the access to high-quality standardized assessment is not evenly distributed across the country. Arguably, YLLs from underdeveloped districts are confronted with much more difficulty in test preparation than their peers in urban areas, which can negatively affect their motivation for learning. Finally, selection refers to the decision-making process based on highly competitive tests among a large student population (Chen, 2011; Hu, 2003). Similar to the imperial examination system that was often accused of restricting test takers' learning and thinking ability, the contemporary testing system has been widely regarded as leading to a common paradox of students with "high score but low ability" (Chen, 2011, p. 63). In other words, scholars are anxious about the fact that Chinese YLLs' motivation to develop multiple competencies can

be hindered by the excessive focus on increasing test scores in the test-driven education system. Nevertheless, although the negative effects of standardized assessment have been acknowledged for centuries, it has been receiving support from the public society unceasingly.

Within the backdrop of testing culture, standardized assessment has been playing a symbolic role in the Chinese society. As Airasian (1988) explains, people's reaction to a symbol depends on the commonality that rests on shared experiences within social groups. Such common experiences may result in an increased social validity of symbols, thereby contributing to the social acceptance of a culturally symbolized phenomenon. As I mentioned earlier, in China, there has been a commonly held belief that successful test results, which serve as the primary criteria for selective decision-making, can ultimately lead to a superior status in social classification (Chen, 2011; Cheng, 2008). Hence, standardized assessments continue to enjoy a wide social acceptance in China, due to the symbolization of such assessments as a "fair measurement for selection of the best talent into the social hierarchy" (Cheng, 2008, p. 16). With a growing motivation among parents to prepare their children for the future advancement, YLLs can start to take their earliest examinations immediately after they enter elementary school; They continue to take numerous standardized tests as they move on to receive secondary and college education (Cheng, 2008; Qi, 2004). In addition, it is worthwhile to note that the intergenerational motivation is a dynamic and complex issue. That is, most YLLs who are being pushed by parents or teachers to study for tests will become parents or teachers of YLLs in the future, while most parents and teachers of YLLs were influenced by their parents' and teachers' attitudes toward test results in the past. Suppose there were many YLLs truly motivated by tests to gain academic success, it is reasonable that they may want to motivate their children to succeed in a similar way. Alternatively, suppose that many YLLs were demotivated by tests in their childhood, the overwhelming majority of parents in every generation who are eagerly fostering their children to succeed from tests helps explain that motivation is not static but dynamic.

Thus far, it has been shown that Chinese parents and teachers are often found to have a strong motivation to make YLLs study for standardized tests. Contextualized in the era of globalization, English education and assessments have been increasingly popular among Chinese elementary, secondary, and tertiary schools. As a subset of standardized assessment, I assume that large-scale and high-stakes language assessment must be subject to the contextual factors that influence the standardized testing system as a whole. Consequently, standardized language assessment can also greatly motivate parents and teachers to prepare YLLs for test success by receiving extra language education in their childhood. I am not suggesting that adults' attitudes toward testing is detrimental for YLLs' motivation or achievement. In fact, numerous studies have proved that early bilingual or multilingual development can be considerably helpful for children's academic success (e.g. Schwarzer, 2001; Wortham, 2005, etc.). Yet, I am arguing that adults' attitudes and test-driven behaviors can have a strong influence on YLLs' learning motivation. Since it is almost unlikely to ensure YLLs without motivation to succeed and benefit from learning in the long run (Dörnyei, 1998), it is critical to examine the standardized language test's motivational effects on YLLs within the context of testing culture in China. Albeit it may not sound well grounded to presume that standardized language assessment demotivates Chinese YLLs for future L2 studies due to the shortage of empirical evidence, there are a number of reasons to expect for such a debilitating effect through a comparison between YLLs and older learners, which is presented in the following section.

3. Motivational effects of testing and young learners' age-related characteristics

In this section, I first provide an overview about YLLs' age-related characteristics, which should be taken into consideration when examining test impacts on YLLs' learning motivation. Particularly, I emphasize that YLLs are distinct from older learners and therefore, more likely to be demotivated by high-stakes assessments, because of their age-related characteristics such as vulnerability to adults' attitudes. Next, I take into account the relationship between test validity and washback effects to further justify the hypothesis that YLLs are more vulnerable to test demotivation than older learners.

Building on this sequential discussion, I argue that Chinese YLLs tend to be demotivated by the standardized English tests, which fail to achieve a high content validity and therefore lead to negative washback effects on YLLs' motivation.

Regarding the increasing number of young English learners within the socioeconomic context of globalization, the interest in understanding the role of language assessment to measure YLLs' language development has grown into an academic field full of potential (McKay, 2006). In school settings, standardized language tests are often used for high-stakes decision-making, such as promotion to the next grade (Wortham, 2005). As I mentioned earlier, Chinese YLLs routinely take summative assessments for all subjects at the end of each semester. For those who are planning to enter prestigious middle schools, it is often necessary to take MSEET developed and administrated by those schools in order to be considered for acceptance. Because of the large population of Chinese YLLs and the limited positions offered by prestigious schools, MSEET and similar tests are designed for schools to select the best candidates based on test scores. Thus, YLLs' test anxiety and academic stress to pursue higher and higher test scores have often been regarded as negative outcomes of standardized assessment (Zhao et al., 2015). Similar to secondary students who are often demotivated by the pressure caused by college entrance tests, testing YLLs through large-scale and high-stakes tests is highly likely to have negative impacts on their learning motivation. However, assessing YLLs can be fundamentally different from assessing older learners (e.g. secondary students, adult learners, etc.). Indeed, I emphasize that there can be a significant distinction between YLLs and older learners in terms of the susceptibility to motivational effects of standardized language assessment. This does not mean that we cannot gain any inspiration from research studies on older learners to make informed decisions about teaching and assessing YLLs. What I attempt to clarify is that conclusions drawn from research on older learners may not be generalizable and applicable for educational and testing practices involving YLLs. As Mihaljević Djigunović (2012) reviews, research studies focusing on YLLs' motivation suggest that YLLs' motivation has a distinct nature due to their age-related characteristics, besides some shared features with older learners. Such age-related characteristics include rapid cognitive, linguistic, physical, and psychological development (Butler, 2016; McKay, 2006; Wortham, 2005). Therefore, test development must be age-appropriate to avoid demotivating YLLs.

According to Butler (2016), distinctive features of YLLs' language development include non-linear developmental factors, individual differences, and vulnerability to adults' attitudes and test results. First of all, YLLs' information processing mechanism follows nonlinear developmental patterns. Various developmental and social changes occur not only in their first three to five years, but also between ages 6 and 14, when YLLs' personal identities, self-consciousness, and the orientation toward achievement will shape their school success in divergent ways (Eccles, 1999). Hence, individual variability is demonstrated among children with different information processing rates. Complicated by the fact that YLLs who are learning their L2 are still developing their first language (L1), the non-linear nature of emergent bilingual development has encouraged researchers to problematize assessing YLLs using the traditional standardized testing format, which normalizes monolingual native speakers' language performance (Flores & Schissel, 2014; Menken, 2008). In addition, YLLs are sensitive to parents' and teachers' attitudes toward their test results, which heavily influence their affective components including motivation, anxiety, and confidence (Butler, 2016). Congruent with my earlier hypothesis, Kim and Seo (2012) find that YLLs are vulnerable to test demotivation, which emerges from their test results, teachers' reactions to their scores, and the pressure from stakeholders (e.g. teachers, school administrators, parents, etc.).

In order to further justify the claim that YLLs are more vulnerable to test demotivation than older learners, I continue to elaborate the washback effects of standardized language assessment on YLLs. As I noted earlier, washback effects refer to test impacts on teaching and learning (Alderson & Wall, 1993; Brown & Abeywickrama, 2010; Cheng, 2004; Cheng & Watanabe, 2004). According to Messick (1996), a test's validity should be measured by to which degree it positively affects teaching and learning. Specifically, positive washback effects occur "when a test measure the same types of materials and

skills that are described in the objectives,” whereas a lack of correspondence to the content objectives results in negative effects (Brown & Hudson, 1998, p. 668). In terms of the washback effects on test takers’ motivation, McDonnell (2005) claims that policymakers generally expect to motivate students to strive for better performance when using assessment as a measurement tool. However, McDonnell (2005) also states that it has been recognized that the a flawed testing system can lead to negative effects on students’ motivation. Particularly in China’s testing system, test format and content are not flexible enough to cater to specific age-related features. For example, the language being used to construct English test items frequently exceeds the vocabulary taught at school, which results in excessive workbook exercises, practice tests, and supplementary classes in private sectors. For many decades, both students and parents have taken for granted that school curriculum can hardly cover all the unpredicted test items, which actually measure extracurricular language knowledge. However, unlike older learners with a fuller awareness of test impacts on their future academic attainment as well as a stronger capability of self-regulation, YLLs can have more difficulty in motivating themselves for intense learning under pressure. Thus, I argue that Chinese YLLs are very likely to be demotivated by standardized English tests, which often suffer from a low content validity and force YLLs to devote extra time and efforts to close the gap between what is learned at school and what is measured on test papers. In addition, I maintain that YLLs are more vulnerable to test demotivation than older learners.

4. Motivational effects of testing and the L2 motivational self system

This section focuses on synthesizing research findings that demonstrate how standardized language tests affect test takers’ motivation. I start from critiquing the predominant assumption that teachers and students are motivated to improve their teaching and learning performances because of the extrinsic rewards or sanctions attached to test results. Next, I integrate some researchers’ arguments that problematize the tendency to emphasize extrinsic motivation, while overlooking personal intrinsic motivation and self-perspectives. Then, I introduce Dörnyei’s (2005, 2009) *L2 Motivational Self System* as a theoretical framework to explore the relationship between YLLs’ testing experiences and their motivation for L2 learning. By detailing how this framework works effectively in a recent study, I suggest that this framework can be employed in future research on test motivation or demotivation among Chinese YLLs.

As Black et al. (2002) point out, there has been a shared view that testing provides extrinsic incentives to YLLs to improve their test performances, which in turn assist them to obtain rewards and avoid penalties. This view is parallel to the popular *motivational theory*, which assumes that extrinsic rewards and sanctions associated with high-stakes test results can motivate teachers to improve performance (Hamilton, Stecher, & Klein, 2002). Even if many scholars cast doubt on its motivational effects, some policy-makers think that the extrinsic motivation is necessary to change teaching behavior (McDonnell, 2005). As Supovitz (2009) suggests, the motivational theory unjustifiably assumes that teachers lack the intrinsic motivation to improve performance, which is challenged by the fact that many teachers internally share a strong sense of responsibility to their profession. Following this line of thought, it is also flawed to assume that test takers, in our case YLLs, are necessarily motivated by extrinsic consequences attached to high-stakes tests and consequently work harder to achieve a stronger test performance. Hence, I propose that it is essential to examine learners’ motivation as an interactive system, where both intrinsic and extrinsic motivations interact with each other in complex and dynamic ways.

Although very few studies have been conducted to examine standardized tests’ impacts on YLLs’ L2 motivation, some studies have investigated the relationship between testing practices and adolescent students’ learning motivation. For example, Crooks (1988) investigates test impacts on students’ self-efficacy, intrinsic motivation, and academic attainment. He finds it problematic to attribute continued learning to extrinsic motivation, which often leads to shallow rather than deep learning (Harlen & Crick, 2003). Furthermore, motivational effects are intertwined with a number of personal and social factors. McDonald (2001) examines how test anxiety affects school children’s performance and finds that age and gender factors also come into play, which adds more difficulty

to generalize the motivational effects of testing. In a cross-national study, Cheng et al. (2014) analyze responses elicited from test takers across three contexts (i.e. Canada, China, and Taiwan), and recognize the complex interrelationship between learning motivation and test anxiety. They further illustrate that differential motivation and test anxiety are closely related to social variables (i.e. perception of test importance and purposes) and personal variables (i.e. gender and age). With a specific focus on high-stakes tests, Madaus and Clarke (2001) argue that high-stakes standardized tests do not show a remarkably positive effect on motivating students in classroom learning, and that they can barely motivate the unmotivated students. Moreover, Amrein and Berliner (2003) argue that when extrinsic rewards or sanctions are attached to test scores, high-stakes tests often demotivate students, thereby leading to higher retention and dropout rates.

Based on the arguments reviewed above, I find that a number of scholars have implied the importance to emphasize personal intrinsic motivation for learning. Nevertheless, merely integrating extrinsic and intrinsic motivation as a dual system is still too oversimplified to capture the complexity and multidimensionality of motivation. Gardner (1985) attempts to use the concept of *integrativeness/integrative motive* to construct an inclusive umbrella term that can contain most, if not all, attitudinal and motivational variables that are considered as predictors for L2 learning success or failure. Nonetheless, his framework has been accused of failing to capture the complexity of social identity and cognitive theories of learning motivation (Dörnyei, 1994; McNamara, 1997). As a recently developed framework that draws on theoretical paradigms from both L2 motivation research studies and motivational psychology, Dörnyei's (2005, 2009) *L2 Motivational Self System* is more comprehensive than previous models in terms of its inclusion of various facets of L2 motivation and affective variables such as L2 anxiety. Particularly, its unprecedented emphasis on self-perspectives can help researchers elicit data from learners themselves and provide new insights for emerging studies. As I discussed in previous sections, standardized language assessment has triggered test anxiety among Chinese YLLs, but those young test takers' perceptions of how they are motivated or demotivated by the assessment remains under-researched. Thus, thanks to its inclusion of learners' voices, *L2 Motivational Self System* can serve as an effective model for examining the multi-faceted issue of test impacts on Chinese YLLs' L2 motivation.

The *L2 Motivational Self System* consists of three main components: (1) the Ideal L2 Self, (2) the Ought-to L2 Self, and (3) the L2 learning experience. According to Dörnyei (2009), the Ideal L2 Self is constituted by idealized images that learners hold about themselves as future L2 users. The Ought-to L2 Self is composed of idealized images that are used by L2 learners to please others or avoid having negative consequences. The L2 learning experience refers to situated motives responding to the immediate learning context and experience. From a self-perspective, this system can be well applied to examine the relationship between testing experience and learners' motivation to study their L2. As an example, Haggerty and Fox's (2015) study explore how standardized English testing experiences affect Korean middle school learners' motivation for English learning. Their findings suggest that the huge amount of time Korean middle school students have spent on test preparation (i.e. the L2 learning experience) may play a vital role in mediating their L2 motivation. Indeed, the less time students spend preparing for L2 tests, the less negative L2 motivation is likely to be identified. Further, they claim that the standardized English proficiency test has set too high standards and unrealistic expectations for a large number of young test takers, compared with their Ideal Self and Ought-to L2 Self. As a result, their motivation to keep learning English is weakened. As suggestions for standardized language test development, Haggerty and Fox (2015) urge test developers to be more concerned with the test's age-appropriateness, potentially detrimental sociocultural effects, and the social consequences of L2 testing practices. These suggestions are aligned with my earlier discussion about the importance of identifying YLLs' age-related characteristics, and that of increasing content validity to enable positive washback effects on learning motivation. As one of the latest studies that explicitly examine the motivational effects of standardized language assessment on Asian adolescent learners, Haggerty and Fox's (2015) study can serve as an inspiring model for conducting similar research on Chinese YLLs.

5. Conclusion and implication

In this paper, I review the historical and cultural context of standardized assessment in China, stressing that YLLs' motivation for learning has often been overlooked. Owing to the significant role that motivation plays in affecting learners' long-term achievement, it is essential to emphasize YLLs' motivation and examine the motivational effects of testing, especially in the test-oriented and outcome-based education system. Based on this overview, I hypothesize that Chinese YLLs, who are under great pressure caused by the standardized testing system, are more vulnerable to test demotivation than older learners. Also, I emphasize YLLs' age-related characteristics should be taken into account in order to develop appropriate language assessments without weakening their learning motivation. Further, I discuss the connection between test validity and washback effects of standardized language assessment on YLLs' learning motivation. As a conclusion, Chinese YLLs' motivation for L2 learning is shaped by the sociocultural and historical context of the testing system, influenced by their parents and teachers, vulnerable to adults' attitudes toward their test results. These factors collectively contribute to the complexity and multidimensionality of YLLs' L2 motivation. Accordingly, I introduce the L2 *Motivational Self System* as an effective theoretical model to aid in future research on Chinese YLLs' L2 motivation relative to standardized language assessment.

Research on YLLs is not a simple task, and motivation is also one of the most elusive concepts in second-language acquisition research due to its complex and multi-faceted nature (Papi, 2010). Consequently, it can be quite challenging for researchers to examine how standardized language assessment affects Chinese YLLs' motivation for L2 learning. Based on what I studied in this field of inquiry, I identified three critical implications for future studies. Firstly, in agreement with Cheng et al. (2014), future survey-based research should develop questionnaire items with high quality to elicit accurate and reflective data from stakeholders in the standardized testing system. While YLLs' voices deserve to be heard, collecting data from them is susceptible to their abilities to process given information, articulate their thoughts, and stay focused on responding to researchers' questions (Dörnyei & Taguchi, 2009). Since Chinese YLLs' motivation are greatly influenced by teachers and parents, it is crucial to gain parental support and teachers' collaboration when it is needed to obtain useful data indirectly. Secondly, researchers should not only be aware of the national education and testing policies, but also keep updating information from local institutions. Since Chinese local private schools are frequently seen to play an active role in developing and administering large-scale and high-stakes entrance tests among YLLs, those institutions function as de facto policy-makers by interpreting, and implementing education and testing policies in divergent ways (Menken, 2008; Shohamy, 1998). Thus, local institutions can powerfully affect Chinese YLLs' motivation for continued learning. Finally, on the ideological level, standardized language assessment should not be simply viewed as a tool for educational measurement. Rather, its implicit social and political characters can ultimately result in "explicit policies of social engineering," (McNamara & Roever, 2006, p. 244) thereby drastically affecting the social life of both individuals and communities. While all these technical, contextual, and sociopolitical factors may increase the difficulty of exploring test motivational effects on YLLs, it will be exciting for both researchers and educators to gain new insights about how to maintain and strengthen children's learning motivation.

Acknowledgment

I sincerely thank Professor Yuko Goto Butler and Professor Jonathan Andrew Supovitz for inspiring and supporting me to explore the intriguing yet complex sociopolitical dimensions of standardized language assessment. I am also grateful to Professor Nancy H. Hornberger for providing me with insightful feedback and resources while I was drafting the paper.

Funding

The author declare no direct funding for this research.

Author details

Chujiao Zhao¹
E-mail: chujiao@se.upenn.edu

¹ Graduate School of Education, University of Pennsylvania, Philadelphia, PA, USA.

Citation information

Cite this article as: Motivational effects of standardized language assessment on Chinese young learners, Chujiao Zhao, *Cogent Education* (2016), 3: 1227020.

Notes

1. In line with Alderson and Banerjee's (2001) definition, I henceforth refer to YLLs as children from age 5 to 12. While the age range is relatively wide, I mainly focus on discussing 5th-6th graders in this paper.
2. *Washback* is originally used as a term in applied linguistics. In the field of language testing, it refers to the influence of tests on teaching and learning (Alderson & Wall, 1993; Brown & Abeywickrama, 2010; Cheng & Watanabe, 2004). In this paper, I focus on discussing the test impact on children' motivation for L2 learning.

References

- Airasian, P. W. (1988). Symbolic validation: The case of state-mandated, high-stakes testing. *Educational Evaluation and Policy Analysis*, 10, 301–313.
<http://dx.doi.org/10.3102/01623737010004301>
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part I). *Language Teaching*, 34, 213–236.
<http://dx.doi.org/10.1017/S0261444800014464>
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115–129.
<http://dx.doi.org/10.1093/applin/14.2.115>
- Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership*, 32–38.
- Black, P., Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M. ... William, D. (2002). *Testing, motivation and learning*. Cambridge: The Assessment Reform Group.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). New York, NY: Pearson.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32, 653–675.
<http://dx.doi.org/10.2307/3587999>
- Butler, Y. G. (2014). Current issues in English education for young learners in East Asia. *English Teaching*, 69, 3–25.
<http://dx.doi.org/10.15858/engtea>
- Butler, Y. G. (2015). English language education among young learners in East Asia: A review of current research (2004–2014). *Language Teaching*, 48, 303–342.
<http://dx.doi.org/10.1017/S0261444815000105>
- Butler, Y. G. (2016). Assessing young learners. In D. Tsagari (Eds.), *Handbook of second language assessment* (pp. 359–376). Berlin: Mouton de Gruyter.
- Chen, Q. (2011). *Assessment policy change in relation to English language teaching and learning in China: A study of perspectives from two universities (doctoral dissertation)*. Brisbane: Queensland University of Technology.
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 147–170). Mahwah, NJ: Lawrence Erlbaum.
- Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, 25, 15–37.
<http://dx.doi.org/10.1177/0265532207083743>
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3–17). Mahwah, NJ: Lawrence Erlbaum.
- Cheng, L., & Qi, L. (2006). Description and examination of the National Matriculation English Test. *Language Assessment Quarterly*, 3, 53–70.
http://dx.doi.org/10.1207/s15434311laq0301_4
- Cheng, L., & Watanabe, Y. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum.
- Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *TESOL Quarterly*, 48, 300–330.
<http://dx.doi.org/10.1002/tesq.105>
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438–481.
<http://dx.doi.org/10.3102/00346543058004438>
- Dörnyei, Z. (1994). Understanding second language motivation: On with the challenge! *Modern Language Journal*, 79, 505–518.
- Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, 31, 117–135.
<http://dx.doi.org/10.1017/S026144480001315X>
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Bristol: Multilingual Matters.
- Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). New York, NY: Routledge.
- Eccles, J. S. (1999). The development of children ages 6 to 14. *The Future of Children*, 9, 30–44.
<http://dx.doi.org/10.2307/1602703>
- Flores, N., & Schissel, J. L. (2014). Dynamic Bilingualism as the norm: Envisioning a heteroglossic approach to standards-based reform. *TESOL Quarterly*, 48, 454–479.
<http://dx.doi.org/10.1002/tesq.182>
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.
- Haggerty, J. F. (2011). An analysis of L2 motivation, test validity and language proficiency identity (LPID): A Vygotskian approach. *Asian EFL Journal*, 13, 198–227.
- Haggerty, J. F., & Fox, J. (2015). Raising the bar: Language testing experience and second language motivation among South Korean young adolescents. *Language Testing in Asia*, 5, 1–16.
- Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Making sense of test-based accountability in education*. Santa Monica, CA: RAND.
- Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, 10, 169–207.
<http://dx.doi.org/10.1080/0969594032000121270>
- Hesketh, T., Zhen, Y., Lu, L., Dong, Z. X., Jun, Y. X., & Xing, Z. W. (2010). Stress and psychosomatic symptoms in Chinese school children: Cross-sectional survey. *Archives of Disease in Childhood*, 95, 136–140.
<http://dx.doi.org/10.1136/adc.2009.171660>
- Hu, G. (2003). English language teaching in China: Regional differences and contributing factors. *Journal of Multilingual and Multicultural Development*, 24, 290–318.
<http://dx.doi.org/10.1080/01434630308666503>
- Kim, T., & Seo, H. (2012). Elementary school students' foreign language learning demotivation: A mixed methods study of Korean EFL context. *The Asia-Pacific Education Researcher*, 21, 160–171.
- Liu, Y., & Dunne, M. (2009). Educational reform in China: Tensions in national policy and local practice. *Comparative Education*, 45, 461–476.
<http://dx.doi.org/10.1080/03050060903391594>
- Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research. *Language Testing*, 28, 367–382.
<http://dx.doi.org/10.1177/0265532211404188>
- Lockette, K. F. (2012). Creativity and Chinese education reform. *International Journal of Global Education*, 1, 34–39.
- Madaus, G., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 85–106). New York, NY: Century Foundation.

- Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18, 351–372.
<http://dx.doi.org/10.1177/026553220101800403>
- McDonald, A. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, 21, 89–101. <http://dx.doi.org/10.1080/01443410020019867>
- McDonnell, L. (2005). Assessment and accountability from the policymaker's perspective. In J. Herman & E. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement* (pp. 35–54). Chicago, IL: National Society for the Study of Education.
- McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.
- McNamara, T. (1997). Theorizing social identity: What do we mean by social identity? competing frameworks, competing discourses. *TESOL Quarterly*, 31, 561–567.
<http://dx.doi.org/10.2307/3587838>
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18, 333–349.
<http://dx.doi.org/10.1177/026553220101800402>
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell Publishing.
- Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Clevedon: Multilingual Matters.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241–256.
<http://dx.doi.org/10.1177/026553229601300302>
- Mihaljević Djigunović, J. (2012). Attitudes and motivation in early foreign language learning. *CEPS Journal*, 2, 55–74.
- Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System*, 38, 467–479.
<http://dx.doi.org/10.1016/j.system.2010.06.011>
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171–190). Mahwah, NJ: Lawrence Erlbaum.
- Schwarzer, D. (2001). *Noa's Ark: One child's voyage into multiliteracy*. Westport, CT: Heinemann.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24, 331–345.
[http://dx.doi.org/10.1016/S0191-491X\(98\)00020-0](http://dx.doi.org/10.1016/S0191-491X(98)00020-0)
- Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, 10, 211–227.
<http://dx.doi.org/10.1007/s10833-009-9105-2>
- Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Educational Research Journal*, 31, 231–262.
<http://dx.doi.org/10.3102/00028312031002231>
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19–36). Mahwah, NJ: Lawrence Erlbaum.
- Wortham, S. C. (2005). *Assessment in early childhood education* (4th ed.). Upper Saddle River, NJ: Pearson.
- Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress, and Coping*, 17, 227–240.
<http://dx.doi.org/10.1080/10615800412331292606>
- Zhao, X., Selman, R. L., & Haste, H. (2015). Academic stress in Chinese schools and a proposed preventive intervention program. *Cogent Education*, 2, 1–14.



© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format
Adapt — remix, transform, and build upon the material for any purpose, even commercially.
The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



Cogent Education (ISSN: 2331-186X) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

