

**INVESTIGATING AUTHENTICITY IN THE MEDICAL ENGLISH TEST SYSTEM FOR HEALTH PROFESSIONALS IN CHINA****^{1,*} Linwei Ma and ²Mengyang Zhang**¹School of Foreign Language, Southern Medical University, China²School of Basic Medical Sciences, Zhejiang Chinese Medical University**Received 20th July 2023; Accepted 15th August 2023; Published online 30th September 2023**

Abstract

Authenticity is an essential component of test validity and usefulness, and its importance role in language for specific purpose (LSP) test is more prominent since LSP test is practical and assesses test takers' language ability in real-world situation. But how to assure the appropriate authenticity is still unclear and rare studies have ever examined the authenticity in the current LSP test. Herein, this study attempted at exploring the health professionals' perception of authenticity in a LSP test Medical English Test System (METS) in China context using questionnaires and interview. The results demonstrated that a majority of participants admitted the importance of speaking task and the METS. Though the needs for English literacy activities are relatively low, preparing for the METS is still considered meaningful and beneficial. Compared with medical students, health professionals hold a more positive attitude towards situational authenticity and rated higher authenticity in the METS test. With sufficient clinical experience, the health professionals stressed the positive impact of METS on English learning and clinical communication. In conclusion, the situational, interactive and textual authenticity of the METS is recognized by the health professionals and medical students but still needs improvement. The METS fails to appropriately meet the demand of health professionals in the healthcare sector in the China context.

Keywords: Authenticity, Medical English Test System, Language for specific purpose

INTRODUCTION

It is acknowledged that Chinese English proficiency has been largely improved after decades of reforming and opening up. Though China has the largest amount of English learners and most students have to learn English at least 10 years before graduation from the university, there are still few language professionals in specific domains. The government has also addressed the development of English for specific purposes (ESP) courses among college students (Luo & Garner, 2017) and ESP courses are gradually replacing general college English in many universities. Assessment and testing is central to language education as it is able to reflect the effectiveness of teaching approach and thereby enhances language learning and teaching. However, most researchers focused on the ESP courses and materials but ignored the necessity of language for specific purpose (LSP) test and professional communication which now is relatively lagging behind (Jin & Hamp-Lyons, 2015). Unlike general tests, LSP testing examines test taker's practical communicative ability in a specific domain such as healthcare and aviation, commerce, whilst the examination manages to mirror authentic language for communicative situations with both their receptive and productive skills (Orozco & Shin, 2019).

LSP testing shall establish what communication entails in the particular context of concern such as in healthcare setting (Elder, 2016). Of note, effective English communication is a cornerstone of human interaction and successful business as well as any patients' and clients' safe. Any communication breakdown probably results inexpensive mistakes and loss of good will. Tests of LSP develop out of social demand to ensure a national healthcare workforce can meet the demand of

patient safety and workplace efficiency (Susy Macqueen *et al.*, 2016). LSP testing of professionals is often considered as part of a formal process of reducing language-associated risk to the public by assessing person's capacity to communicate in professional activities (Knoch & Macqueen, 2019). The characteristics of LSP testing impart the particular importance of authenticity to the test. Authenticity is regarded as a main component of test usefulness and validity (Bachman & Palmer, 1996) and McNamara (1996) had highlighted the crucial role of authenticity in LSP testing. But how to determine and assess authenticity in LSP testing in English-speaking countries has not reached agreement (Burton, 2020; Davidson, 2022; Knoch & Macqueen, 2019) and it seems more difficult in the places outside the English countries like China where English is not the dominant language. Authenticity is highlighted as a criterion of test validity and a key component of language proficiency validity. Extensive research on authenticity in ESP testing and standard setting for clinical communication using interdisciplinary approaches has been conducted in the Australia (Davidson, 2018, 2022; Elder, 2016). However, rare studies have been conducted in Chinese-speaking settings (Luo, 2019).

Medical English Test System (METS) is the first national LSP test for health sector in China available in four levels: METS-1 through METS-4 ("An Introduction to Medical English Test System," 2023) and has been carried out across the country over a decade. The presence of the METS has importantly enhanced the language education in medical universities and improved language proficiency of health professionals. However, it seems that the METS fails to achieve the expected target to become a popular and well-recognized LSP test, since still many institutions and health professionals have not fully acquainted this test. There are possibly plenty of complex factors responsible for such weird phenomenon which deserves

***Corresponding Author: Linwei Ma**

School of Foreign Language, Southern Medical University, China

more attentions and investigations. Herein, this study put the focus on the authenticity in the METS and intended to investigate the health professionals' perceived authenticity using quantitative data and qualitative methods to provide an insight into ESP teaching and LSP development, and shed light on to the further development of LSP testing in China.

LITERATURE REVIEW

Authenticity in LSP testing

Authenticity has been highlighted over decades since Close (1965) first stressed the importance of authentic texts for language learning in 1965. Widdowson (1979) further pointed up that the interaction between the audience and the text determines the authenticity and the role of context cannot be ignored when analyzing authenticity. Authentic test tasks later are noted to promote the development of test takers' communicative language abilities (Burton, 2020). Authenticity was reconceptualized as situational authenticity and interactional authenticity by Bachman (1991): the match between the test tasks to target language use (TLU) tasks and the interaction between the test taker and the test tasks. Authenticity is considered as the resemblance between assessment and real-life tasks. Douglas (2000) demonstrated authenticity of task and unlike general language test, interaction between language knowledge and background knowledge are two key component of LSP tests. Establishing authenticity is essential in measuring the language abilities and cognitive processes of language learners. However, the absence of authenticity imposes a damage on test validity and usefulness, on the contrary. Less authenticity is a potential validity problem as simulations in testing allow researchers to assess essential abilities of test takers under authentic condition (Stadler, Iliescu, & Greiff, 2021). Since the degree of authenticity is determined by several factors such as test construct, need analysis, and policy, we shall not directly identify one test task authentic or inauthentic (Dammann, Friederichs, Lebedinski, & Liesenfeld, 2020).

For perceptions of authenticity, stakeholders' perceptions of test authenticity differ across groups of stakeholders (Bachman & Palmer, 1996). The perceived importance of authenticity in tests may change test-takers orientations and it is thus necessary to investigate the perceptions of a range of stakeholders (Lewkowicz, 2000). Especially in LSP tests involving language knowledge and specific purpose background knowledge, opinions of domain experts are more central to development of authentic test tasks and appropriate perception of authenticity in a specific LSP test (Davidson, 2022), which will be discussed in following section. Language knowledge intertwines with specific purpose background knowledge. Interestingly, Lewkowicz (2000) found that authenticity was not a priority for most student participants, although its perceived importance varied as a function of students' levels of proficiency. Domain expert involvement is key to authenticity concerns in LSP testing, covering authentic test and assessment criteria (Davidson, 2022). The experts should be included more regularly in all aspects of LSP test development. Continuous feedback from practitioners in the field and TLU specialists will help test development, adhering to the original specifications. Collectively, there is general agreement of significance of certain degree of authenticity to testing, but there is disagreement on how it is operationalized and how to determine what level of performance on the

specific-purpose language test so that the qualified test takers can participate safely and effectively in the target language use situation. However, little is known the perceived authenticity in the LSP testing in the eye of expert domain and even none of studies have ever explored the authenticity in the language tests in China context. Considering the particular importance of authenticity to the LSP testing, the aim of this paper is to show how test-takers and health professionals perceive authenticity in the METS using qualitative and quantitative data. The study address two following questions:

1. Is the METS an authentic test from the perspective of test-takers?
2. How do participants in the healthcare domain perceive the authenticity in the test of LSP in China?

METHODOLOGY

Participants

To acquire the perceived authenticity in the METS, opinions from healthcare doctors, nurses, technical physicians and medical students were addressed herein. These stakeholders were considered first since they are the major populations of healthcare and potential candidate of the METS. In terms of their clinical experience and identity, a total of 47 participants was divided into healthcare professional group (n = 12) and medical student group (n = 35). Two stakeholder groups were comprised of nurses (n = 3), doctors (n = 2) and other healthcare professionals (n = 7) from key public hospitals in Guangzhou and medical students (n = 35) majoring in nurse, clinical medicine, traditional Chinese medicine, pharmacy, rehabilitation medicine from a key medical university in southeast of China. For the first group (healthcare professionals), they all had more than three years work experience in Chinese healthcare contexts so that they might comment on workplace communication practices based on their authentic experience. For the medical students, they are studying the subjects for METS but only few have passed the test. Despite of the lack of experience of workplace communication or insufficient clinical practice, they are the major participants of METS and their views hence should be valued. In the study, they assessed the authenticity in the METS based on their experience of internship in the hospitals or their assumption and learning experience. Afterwards, four participants from both groups were volunteered to participate in the following semi-structure interview based on their questionnaire results.

Instruments

The questionnaire was comprised of 19 items including 16 5-point Likert scale items as previously described by Dorman and Knightley (2006) (Table 1). The items can be briefly classified into person information, test usefulness textual authenticity, situational authenticity, and test necessity (Table 2). Two nurses and two university medical students volunteered to participate a semi-structure interview. Through the qualitative analysis of interview script, the participants' perceived authenticity in the METS was further explored and unveiled.

Statistical analysis

Data of the questionnaire were processed by SPSS 26.0 and the descriptive statistic were presented in Table 4.

Table 1. Questionnaire of authenticity in METS

Item	Question
1	I know the METS.
2	Have you taken the METS?
3	What is your profession?
4	My English language skills improved/have improved considerably since I started studying for METS
5	Which of the four language skills I have developed the most through studying for OET?
6	I am asked to apply my learning to real-life scenarios.
7	I find/found METS tasks are meaningful.
8	I find/found the subjects I study/studied for METS to be beneficial
9	I found the subjects I studied/am studying for METS to be related to my future career
10	I think METS tasks check my understanding of topics.
11	I find/found METS mainly assesses test taker's English skills.
12	I find that writing process in the METS test is similar to writing process in daily life.
13	I often use oral English in my daily life.
14	I think it is necessary to add oral English exam in METS.
15	I find/found topics in the METS are similar to the topics in daily work.
16	I find/found the difficulty of the METS test is similar to the difficulty of English tasks in real work.
17	I believe that qualified test takers can communicate effectively in English in the workplace of health sector.
18	Overall, I think the experience of studying for METS is excellent.
19	I find that METS is not reliable or authentic.

Table 3. Distribution of items in the questionnaire

Sections	Items
Personal information	1, 2, 3
Usefulness	4, 5, 7, 8, 18
Textual authenticity	10, 11
Necessity of oral English	13, 14
Situational authenticity	6, 9, 12, 15, 16, 17, 19

Table 3. KMO and Bartlett's test of the questionnaire

KMO and Bartlett's test		
KMO value		0.785
Bartlett's test	Sig.	0.000
	df	120

Chi-square test was carried out to explore the potential difference in perceived authenticity between medical students and health professionals. KMO and Bartlett's Tests proved satisfactory with the KMO index being 0.785 (Table 3).

RESULTS

Evaluation of usefulness of METS

High-stake language tests often serve as gates into socioeconomically desirable domains. Usefulness is an important element in designing and developing a language test and authenticity is closely related to functions of a test. In the study, health professionals showed positive attitude towards usefulness of the METS (Mean \leq 2.00) but university students still doubted its function (Mean $>$ 2.00). For item 4 (My English language skills improved/have improved considerably since I started studying for METS), major medical students held neutral attitude (24/35, 68.6%) (Table 6), while the healthcare professionals recognized the function of the METS on language learning (Mean = 2.00) ($p = 0.005$) (Table 7). Both groups acknowledged the role of preparing for the METS in improving their English listening (14.90%) and reading ability (68.60%) (Table 8). One student described that his vocabulary got richer when preparing for the METS and had learned more prefix, root and suffix of medical terms, but he also showed his concern on test usefulness from the perspective of employment and policy:

Many employers and hospitals don't pay attention to and value this test, and the certificate of METS could not give me a hand indeed, though studying for the test does improve my English for medical purpose. English is seldom used in the hospital and I used mainly for reading references during my internship in the hospital. Though rare, once foreign patients come to hospital, most nurses and some doctors still can't communicate effectively with them. In conclusion, the above data demonstrated preparing for METS is instrumental to test-takers and METS is a useful test.

Assessment of textual authenticity and situational authenticity

The study intended to evaluate the authenticity based on test-takers' assessment of authentic test and situational authenticity using item 10 (I think METS tasks check my understanding of topics) and item 11 (I find/found METS mainly assesses test taker's English application ability). Both health professionals and university students agreed the textual authenticity in METS (Table 9) ($p = 0.093$, $p = 0.104$). Health professionals hold positive attitude towards situational authenticity in the METS with means less than 2.00 among items 6, 9, 12, 15 and 17 (Table 6). In their views, the tasks in the METS are similar to the real-world English literacy activities they encounter in the work place with similar topics and type, but the difficulty of simulated task in the test did not appropriately meet the condition of real-life task varies.

Table 4. Descriptive analysis of questionnaire

Item	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
1	0	3(6.4%)	14(29.8%)	18(38.3%)	12(25.5%)
4	6(12.8%)	9(19.1%)	29(61.7%)	3(6.4%)	0
6	9(19.1%)	26(55.3%)	9(19.1%)	2(4.3%)	1(2.1%)
7	11(23.4%)	26(55.3%)	7(14.9%)	3(6.4%)	0
8	14(29.8%)	24(51.1%)	7(14.9%)	3(6.4%)	0
9	13(27.7%)	23(48.9%)	11(23.4%)	0	0
10	14(29.8%)	25(53.2%)	7(14.9%)	1(2.1%)	0
11	15(31.9%)	23(48.9%)	7(14.9%)	2(4.3%)	0
12	7(14.9%)	21(44.7%)	14(29.8%)	5(10.6%)	0
13	7(14.9%)	12(25.5%)	21(44.7%)	5(10.6%)	0
14	10(21.3%)	17(36.2%)	15(31.9%)	5(10.6%)	0
15	8(17%)	18(38.3%)	17(36.2%)	4(8.5%)	0
16	6(12.8%)	16(34%)	19(40.4%)	6(12.8%)	0
17	8(17%)	25(53.2%)	11(23.4%)	3(6.4%)	0
18	9(19.1%)	25(53.2%)	11(23.4%)	2(4.3%)	0
19	4(8.5%)	5(10.6%)	15(31.9%)	19(40.4%)	4(8.5%)

Table 5. Means of questions on usefulness of METS

Section	Item	N	Mean
Usefulness	4	47	2.62
	7	47	2.04
	8	47	1.94
	18	47	2.13

Notes: 1=strongly agree, 5= strongly disagree

Table 6. Comparison of mean value between two groups

Groups	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item	
	1	4	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Healthcare professionals	Mean	3.42	2	1.92	1.83	1.67	1.58	1.5	1.67	1.75	2.17	1.92	2.08	2.42	2	1.67	3.25
	N	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
	SD	1.165	0.953	1.084	1.115	0.985	0.793	0.674	0.985	0.965	1.115	1.165	1.165	1.311	1.128	0.651	1.485
University students	Mean	3.97	2.83	2.23	2.11	2.03	2.09	2.03	2	2.57	2.8	2.46	2.46	2.57	2.26	2.29	3.31
	N	35	35	35	35	35	35	35	34	35	35	35	35	35	35	35	35
	SD	0.747	0.618	0.77	0.676	0.707	0.658	0.707	0.739	0.739	0.933	0.817	0.741	0.698	0.657	0.75	0.9
Total	Mean	3.83	2.62	2.15	2.04	1.94	1.96	1.89	1.91	2.36	2.64	2.32	2.36	2.53	2.19	2.13	3.3
	N	47	47	47	47	47	47	47	46	47	47	47	47	47	47	47	47
	SD	0.892	0.795	0.859	0.806	0.791	0.721	0.729	0.812	0.87	1.009	0.935	0.87	0.881	0.798	0.769	1.061

Table 7. Chi-square test of authenticity among two groups stakeholders

Item	Stakeholder	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree	Sig.
1	Health professionals		3(25.0%)	4(33.3%)	2(16.7%)	3(25.0%)	0.013
	University Students			10(28.6%)	16(45.7%)	9(25.7%)	
4	Health professionals	5(41.7%)	2(16.7%)	5(41.7%)	0	0	0.005
	University Students	1(2.9%)	7(20.0%)	24(68.6%)	3(8.6%)		
6	Health professionals	5(41.7%)	5(41.7%)	0	2(16.7%)	0	0.007
	University Students	4(11.4%)	21(60.0%)	9(25.7%)		1(2.9%)	
7	Health professionals	6(50.0%)	4(33.3%)	0	2(16.7%)	0	0.01
	University Students	5(14.3%)	22(62.9%)	7(20.0%)	1(2.9%)		
8	Health professionals	7(58.3%)	3(25.0%)	1(8.3%)	1(8.3%)	0	0.054
	University Students	7(20.0%)	21(60.0%)	6(17.1%)	1(2.9%)		
9	Health professionals	7(58.3%)	3(25.0%)	2(16.7%)	0	0	0.021
	University Students	6(17.1%)	20(57.1%)	9(25.7%)	0	0	
10	Health professionals	7(58.3%)	4(33.3%)	1(8.3%)	0	0	0.093
	University Students	7(20.0%)	21(60.0%)	6(17.1%)	1(2.9%)		
11	Health professionals	7(58.3%)	3(25.0%)	1(8.3%)	1(8.3%)	0	0.104
	University Students	8(23.5%)	19(55.9%)	6(17.6%)	1(2.9%)		
12	Health professionals	6(50.0%)	4(33.3%)	1(8.3%)	1(8.3%)	0	0.001
	University Students	1(2.9%)	17(48.6%)	13(37.1%)	4(11.4%)		
13	Health professionals	4(33.3%)	4(33.3%)	2(16.7%)	2(16.7%)	0	0.087
	University Students	3(8.6%)	8(22.9%)	19(54.3%)	3(8.6%)		
14	Health professionals	6(50.0%)	3(25.0%)	1(8.3%)	2(16.7%)	0	0.017
	University Students	4(11.4%)	14(40.0%)	14(40.0%)	3(8.6%)		
15	Health professionals	5(41.7%)	3(25.0%)	2(16.7%)	2(16.7%)	0	0.024
	University Students	3(8.6%)	15(42.9%)	15(42.9%)	2(5.7%)		
16	Health professionals	4(33.3%)	3(25.0%)	1(8.3%)	4(33.3%)	0	0.002
	University Students	2(5.7%)	13(37.1%)	18(51.4%)	2(5.7%)	0	
17	Health professionals	5(41.7%)	4(33.3%)	1(8.3%)	2(16.7%)		0.011
	University Students	3(8.6%)	21(60.0%)	10(28.6%)	1(2.9%)	0	
18	Health professionals	5(41.7%)	6(50.0%)	1(8.3%)	0	0	0.088
	University Students	4(11.4%)	19(54.3%)	10(28.6%)	2(5.7%)	0	
19	Health professionals	3(25.0%)	0	2(16.7%)	5(41.7%)	2(16.7%)	0.054
	University Students	1(2.9%)	5(14.3%)	13(37.1%)	14(40.0%)	2(5.7%)	

Table 8. Frequency of English activities in the hospital

Language skill		Group			Sig.
		Healthcare professionals	University students	Total	
Listening	N	3	4	7	0.563
	Percentage	25.00%	11.40%	14.90%	
Speaking	N	1	5	6	
	Percentage	8.30%	14.30%	12.80%	
Reading	N	8	24	32	
	Percentage	66.70%	68.60%	68.10%	
Writing	N	0	2	2	
	Percentage	0.00%	5.70%	4.30%	

Table 9. Textual and situational authenticity in the METS

Section	Item	N	Mean
Textual Authenticity	10	47	1.89
	11	47	1.91
Situational Authenticity	6	47	2.15
	9	47	1.96
	12	47	2.36
	15	47	2.36
	16	47	2.53
	17	47	2.19
Authenticity	19	47	3.3

Note: 1 = strongly agree, 5 = strongly disagree

A nurse explained that the test task seems easier compared with the complex English tasks in the hospital, which rarely take place actually. The actual task possibly integrates several language abilities at the same time such as listening and speaking, writing and reading. As for the difficulty of the test, many students (18/35, 51.4%) also remain skeptical attitude and few students (5.7%) even disagreed with the present threshold. The difference between health professional and university students was statically significant ($p < 0.002$). Apart from item 16, the difference in item 6, 9, 12, 15, and 17 was also significant ($p < 0.05$). Both health professionals and students noted the similarity between the content of the test and actual task (Mean < 2.00), affirming the authenticity and reliability in general (Table 7). Taken altogether, both groups of stakeholders confirmed the authenticity in the METS, though there are still some doubts or questions exist which deserves further investigation.

DISCUSSION

This paper has sought answers to two research questions. In response to the first question, both professionals and medical students largely ascertained the similarities between test task and real-life activities in health sector and test usefulness. The participants noted that preparing for the METS is meaningful and useful since it might improve their language skills especially reading ability. Previously, Pill and Knoch (2013) also found that preparing for occupation English test (OET) might increase test candidates' understanding and awareness of patient-centered care. The OET is an international LSP test for healthcare professionals evaluating four skills (speaking, writing, listening and reading), developed by Australia government. In line with Macqueen's finding, our work demonstrated that for medical students, preparing for the METS might help set test candidates' expectations of

professional life and the workplace. Though the health professionals indicated the infrequency of English literacy activities in their daily routine in hospital, the necessity is still stressed in the healthcare sector and the certificate of college English test 4 (CET-4) and CET-6 is usually required in job employment. The needs of preparing for the METS are closely associated with the policy and the real contexts. As China continues to further open up and develops, more foreigners and immigrants will come to our country and at that time, international hospitals are required. In fact, the major hospitals are progressing to become international hospitals, so apart from advanced instruments and technology, international-level service and English communication ability is needed, which suggests the necessity of LSP testing. But how to meet the potential demand of huge health professionals shall get more attention and authenticity should be one of the key breakthroughs. The answer of second question is thus meaningful.

In terms of second question (How do participants in the healthcare domain perceive the authenticity in the test of LSP in China?), METS is believed to successfully function as gatekeeper of health sector. Both medical students and health professionals agreed with the similarities of topics, subjects and genre between test tasks and real activities. The authentic tasks and texts in the test make METS more practical and reliable. But the participants noted a gap in task difficulty between test task and real-world activities. Some deemed the test task appears relatively easy, compared with the complex task in the workplace and they thus doubted the effectiveness of the test. Task difficulty importantly affects the authenticity and validity of a test. A too easy test might pose a potential danger to the test takers, patients and even the society. The past candidates are expected to effectively deal with the English tasks in the work place but they might hardly meet the expectation in fact. Such a huge gap might increase the possibilities of miscommunication. On the one hand, the participants perceived the METS easy probably due to their disciplinary knowledge and academic background. However, some past candidates failed to communicate with the international patients due to the absence of speaking test task. But apart from essay, there are other types of writing activities in healthcare not included, such as note-taking and writing report, which are more frequently-encountered tasks in daily life. McNamara and Pill (2016) set a new minimum standard of professionally relevant oral competence based on opinions of medicine, nursing, and physiotherapy. Such a new standard better meets test takers' demand from a new perspective. An analysis of the communication problems encountered by potential test-takers once in the TLU is a key to design of LSP test as this can help narrow down the potential tasks that could be identified in a domain analysis. More importantly, design of test task should be based on need analysis, following the needs of learner, society, policy or other factors. Among various profession, lawyers and doctors may use English more frequently in the Chinese context (Jin & Hamp-Lyons, 2015). Wherever possible, researchers should examine actual or potential language and communication problems future test-takers may have when entering the domain. Communication with patients is relatively infrequent activity for most health professionals up to now, but it is still of crucial importance. There are still several limitations in the study, which will be the direction of further investigation. For instance, the questionnaire should be administered to larger sample sizes to ensure the validity and generalizability of the data collected.

The incomplete list of tasks and thereby collected information cannot fully represent communication in healthcare. A large-scale investigation should be carried out in the following study. Opinions from test developer, educators and medical experts will be more convincing and valuable, which are crucial to the design and development of METS. METS is an important test for construction of talent team of healthcare professionals and international hospitals in China, deserving more investigation and support. Meanwhile, the METS is a LSP test in the Chinese-speaking regions and conceptualization of authenticity in such a LSP test might provide valuable insights into test development and ESP education

Conclusion

Collectively, this study has noted the authenticity and usefulness of a LSP test for healthcare sector in China, METS, from the perspective of test-takers (i.e., health professionals and medical students). Investigating test wash back and paying attention to stake holder perception play a large role in language testing. This empirical research first investigated the authenticity of METS and provided test-taker's views towards the construct of the test, helping to gain a better understanding of the authenticity in METS and laying a foundation for LSP development in China. The authenticity and interactiveness between test and real-world activities requires further improvement. Therefore, it is necessary to further explore the process of standard setting of the METS based on the need analysis and stakeholders' perception.

Acknowledgement: We would like to express our deepest gratitude to the participants who honestly shared their opinions on LSP in healthcare sector and appreciated the suggestion from Dr. Zirong Li.

Statement of Competing Interests: The authors have no competing interests.

REFERENCES

- Bachman, L. F. (1991). What Does Language Testing Have to Offer? *TESOL quarterly*, 25(4), 671-704. doi: <https://doi.org/10.2307/3587082>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
- Burton, J. D. (2020). Raters' measurement of test-task authentic engagement in L2 oral-performance assessment: An exploration of scale development. *System*, 90, 102233-102212. doi:10.1016/j.system.2020.102233
- Close, R. A. (1965). *The English we use for science: a selection of texts, with exercises for language practice*: Longman ELT.
- Dammann, O., Friederichs, K. M., Lebedinski, S., & Liesenfeld, K. M. (2020). The Essence of Authenticity. *Front Psychol*, 11, 629654. doi:10.3389/fpsyg.2020.629654
- Davidson, S. (2018). How valid are domain experts' judgements of workplace communication? Implications for setting standards on the Occupational English Test (OET) Writing sub-test.
- Davidson, S. (2022). The domain expert perspective: A qualitative study into the views expressed in a standard-setting exercise on a language for specific purposes (LSP) test for health professionals. *Language testing*, 39(1), 117-141. doi:10.1177/02655322211010737
- Dorman, J. P., & Knightley, W. M. (2006). Development and validation of an instrument to assess secondary school students' perceptions of assessment tasks. *Educational Studies*, 32(1), 47-58. doi:10.1080/03055690500415951
- Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press.
- Elder, C. (2016). Exploring the limits of authenticity in LSP testing: The case of a specific-purpose language test for health professionals. *Language testing*, 33(2), 147-152. doi:10.1177/0265532215607397
- An Introduction to Medicial English Test System. (2023).
- Jin, Y., & Hamp-Lyons, L. (2015). A new test for China? Stages in the development of an assessment for professional purposes. *Assessment in education : principles, policy & practice*, 22(4), 397-426. doi:10.1080/0969594X.2014.1003794
- Knoch, U., & Macqueen, S. (2019). *Assessing English for Professional Purposes*: Routledge.
- Lewkowicz, J. A. (2000). Authenticity in language testing: some outstanding questions. *Language testing*, 17(1), 43-64. doi:10.1177/026553220001700102
- Luo, J., & Garner, M. (2017). The Challenges and Opportunities for English Teachers in Teaching ESP in China. *Journal of Language Teaching Research*, 8, 81-86.
- Luo, L. (2019). Pursuing authenticity in ESP testing-the need for interdisciplinary collaboration. *Journal of Teaching English for Specific and Academic Purposes*, 159. doi:10.22190/JTESAP1902159L
- Macqueen, S., Pill, J., & Elder, C. K., U. . (2013). *Investigating the test impact of the OET: A qualitative study of stakeholder perceptions of test relevance and efficacy*. Retrieved from University of Melbourne, Language Testing Research Centre, Melbourne, Australia:
- Macqueen, S., Pill, J., & Knoch, U. (2016). Language test as boundary object : Perspectives from test users in the healthcare domain. *Language testing*, 33(2), 271-288. doi:10.1177/0265532215607401
- McNamara, T. (1996). *Measuring Second Language Performance*: Longman Pub Group.
- Orozco, R. A. Z., & Shin, S. Y. (2019). Developing and validating an English proficiency Test. *MEXTESOL Journal*, 43.
- Stadler, M., Iliescu, D., & Greiff, S. (2021). Validly authentic: Some recommendations to researchers using simulations in psychological assessment. *European Journal of Psychological Assessment*, 37(6), 419-422. doi:10.1027/1015-5759/a000686
- Widdowson, H. G. (1979). *Teaching Language as Communication*. Oxford University Press. doi: [https://doi.org/10.1016/0346-251X\(79\)90008-3](https://doi.org/10.1016/0346-251X(79)90008-3)