

Benchmarking Students' Attainment in Australian Tertiary Chinese Programs Using the New HSK Tests

Wenyang Jiang 

School of Languages and Cultures, the University of Queensland, St. Lucia, Australia

Email: w.jiang2@uq.edu.au

How to cite this paper: Jiang, W. Y. (2020). Benchmarking Students' Attainment in Australian Tertiary Chinese Programs Using the New HSK Tests. *Creative Education, 11*, 624-638.
<https://doi.org/10.4236/ce.2020.115047>

Received: March 3, 2020

Accepted: May 5, 2020

Published: May 8, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).
<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

With the increase of China's global influence, the Chinese as a foreign language (CFL)¹ programs in higher education institutions have expanded worldwide and reached a stage that benchmarking must be considered to make sure that these CFL programs achieve their expected education quality. After reviewing and comparing a few possible options like the Common European Framework of Reference for languages (CEFR), this study argues that the new HSK, namely the standardized Chinese Proficiency Test developed in China, should be adopted for benchmarking students' attainment in Australian tertiary CFL programs. Through supporting empirical data, it is proposed that the new HSK level 4 be established as the Chinese BA language proficiency threshold in Australian tertiary CFL programs. Benefits of implementing such outcome-based assessment in tertiary language programs have also been discussed.

Keywords

Chinese as a Foreign Language (CFL), Benchmarking, Higher Education, HSK

1. Introduction

Bellassen (2017) compares modern languages as “moving celestial bodies”, which are being driven by a number of forces including communication, culture and identity.

¹The distinction between Chinese as a second language (CSL) and Chinese as a foreign language (CFL) comes from the native language of the country in which instruction is conducted. An CSL learner learns Chinese in the country where Chinese is the primary national language like mainland China, Taiwan or Singapore while an CFL learner learns Chinese in a country where Chinese is not the primary national language like the US, the UK and Australia. Sometimes CSL or L2 Chinese is used as a broad term covering both CSL and CFL.

He further describes that:

Some of them are near and others are far away; some appear to cluster together as a group and others seem to stand alone. Chinese, which was once hardly visible within the educational galaxy, has since become much brighter and its attraction is such that it could now be ranked the second foreign language studied in various Asian countries just after English (p. x).

This metaphor is not only vivid in describing modern languages, but also indicating the fact that Chinese as a foreign or a second language (CFL, CSL, or L2 Chinese) has become increasingly popular nowadays. In Bellassen's words, "A historic page in the history of teaching Chinese as a foreign language is writing itself under our very eyes: the advent of the teaching of Chinese as a structured programme and not just as a specialist subject" (p. x). This summarizes the phenomenon that along with the increase of China's global influence, the Chinese language, a part of the country's soft power, has also gained unprecedented popularity around the world. In other words, the CFL programs have been expanded worldwide and reached a stage that benchmarking must be considered to make sure that these CFL programs achieve their expected education quality.

Being an integral part of the Asia-Pacific region, Australia has realized the importance of seeking to strengthen its ties with Asia, particularly China. Hence the Chinese language is considered of strategic importance to the socioeconomic development of contemporary Australia. That is why the *Australia in the Asian Century* White Paper sets out an ambitious roadmap to make sure Australia achieving an Asia capable skill set by 2025, particularly the Chinese language skill. Among the 43 higher education institutions in Australia, 32 universities offer CFL programs (Wang & Niu, 2014). However, the expansion of tertiary CFL programs has not been without its challenges and demands, particularly in relation to the standardisation of L2 Chinese competence. One of the challenges is the provision of high-quality teaching for learning opportunities in classrooms and the development of appropriate methods of assessment that reflect the students' actual competence in the language.

In order to know whether university programs deliver what they claim to, benchmarking is called upon to demonstrate that graduates achieve the targeted learning outcomes at required performance standards. In fact, demonstrated outcomes benchmarking is an explicit requirement in the Australian Tertiary Education Quality Standards Agency (TEQSA) threshold standards framework. As King and Hoffmann (2013) argue, "Published benchmarking criteria and outcomes provide stakeholders (governments, employers, professions, providers, parents, graduates and students) with assurance that universities are delivering graduates with the capabilities they claim. Higher education providers also use confidential benchmarking for improvement of their programs and supporting processes" (p. 2). It is important to keep firmly in mind that one of the main purposes of standards, benchmarks, etc. is to provide a way of informing and improving classroom learning. This study aims to propose an outcomes-based proficiency criterion for graduates of tertiary CFL programs in Australia.

2. Literature Review

For benchmarking BA graduates' attainment in Australian tertiary CFL programs, it is necessary to review what have been used so far and what would be the most promising standards in benchmarking foreign language progress. In this section, measures have been employed in benchmarking foreign language development, particularly CFL development, in Europe, USA, Taiwan and Mainland China will be reviewed in order to identify the most promising candidate for benchmarking graduates' attainment in tertiary CFL programs in Australia. This section concludes with the best option identified.

2.1. The Common European Framework of Reference for Languages (CEFR) in Europe

The Common European Framework of Reference for languages (CEFR) is a guideline used to describe achievements of learners of foreign languages across Europe and, increasingly, in other countries as well. It was developed by the Council of Europe as the main part of the project "Language Learning for European Citizenship" between 1989 and 1996. Its main aim is to provide a "transparent, coherent and comprehensive basis for the elaboration of language syllabuses and curriculum guidelines, the design of teaching and learning materials, and the assessment of foreign language proficiency" (CEFR, 2019). Its standards have incorporated descriptors for more than 40 modern languages. These descriptors feature six levels (A1, A2, B1, B2, C1 and C2) that are divided into the three broad competence levels of Basic User, Independent User and Proficient User (Lu & Song, 2017). As the Council of Europe claims, the development of the descriptors was based on extensive research and widespread consultation and they are increasingly shaping the reform of foreign language curricula and the development of teaching materials and of foreign-language proficiency tests.

As is commonly known, the Chinese language is a non-alphabetic language with a writing system usually termed as logography. Learners of CFL, consequently, have very different learning experiences and difficulties from those learning alphabetic languages. Therefore, concerns have been raised about adjusting the CEFR standards for the teaching and learning of the Chinese language in European educational settings. For example, Bellassen and Zhang (2008) warn that the unique linguistic features of the Chinese language should not be overlooked when aligning with CEFR standards. When the effort of applying the CEFR proficiency descriptors to benchmark the Chinese language was made in the European context, it was found that no specific descriptors for recognition and production of the Chinese characters have been included, which places greater cognitive demands on visual-spatial analytic skills in both cognitive processing and reading acquisition (Li, Shu, & Liu, 2014; Lu & Song, 2017). As Lu and Song (2017) state,

When learning a word in Chinese, either a single-character or a two character word, the pronunciation is usually learned through the Romanised

system of *Pinyin*; the logographical characters consisting of strokes or radicals with certain structures representing the meanings must then be learned, which in most cases contain little or rather unreliable phonetic cues. While *Pinyin* might be easier to learn, as some of the syllables resemble pronunciation in English or in other alphabetical languages, to memorise the components of a character that may consist of more than 15 strokes, and then to be able to reproduce these from memory in writing is the most challenging task for learners at all levels (p. 14).

Due to the learning difficulties in recognizing and producing Chinese characters from memory, mastering a certain number of vocabulary, namely recognition and production, has become one of the major criteria in a certain level of CFL proficiency. This is also the major difference in learning CFL from learning other alphabetic foreign languages. Therefore, efforts were made through the EBCL project (EBCL, 2019) to expand and modify the existing CEFR descriptors to incorporate the specific CFL characteristics, namely the number of characters a learner can recognize and produce from memory. However, the outcome of the EBCL project was that some specific new descriptors at the Basic User level, namely A1 and A2, have been added, leaving the remaining two broad levels of Independent User and Proficient User untouched. More research is required to be able to figure out how many Chinese characters should be mastered by Independent User and Proficient User, which was beyond the commitments of the EBCL project.

In sum, the main problem with CEFR in benchmarking CFL development lies in the fact that the ability to recognize and produce Chinese characters is not incorporated in the assessment descriptors. Li and Zhang (2009) have also argued that the CEFR should not be applied to benchmarking Chinese proficiency for three reasons. First, the political agenda of the CEFR is to achieve greater unity among Council of Europe member states, which does not include China; second, the CEFR is primarily for European languages that use alphabetic writing systems, whereas Chinese is a non-alphabetic language; and third, the socio-cultural differences between Chinese and European languages lie beyond the scope of CEFR. Therefore, it is unlikely that the CEFR will be a promising option to be used to benchmark graduates' attainment in Australian tertiary Chinese programs although it is well known and most influential in benchmarking foreign language learners' development of alphabetic languages.

2.2. Proficiency Tests Developed on the Basis of ACTFL Proficiency Guidelines in the USA

Aside from the CEFR discussed above, the American Council on the Teaching of Foreign Languages (ACTFL) has been another influential organisation with regard to language standards and guidelines. The ACTFL Proficiency Guidelines (hereafter, the Guidelines) were created by the American Council on the Teaching of Foreign Languages in order to provide a means of assessing the proficiency of a foreign language speaker. The Guidelines are broken up into different profi-

ciency levels: novice, intermediate, advanced, superior, and distinguished. Additionally, each of these (except superior and distinguished) levels is further subdivided into low, mid and high. These proficiency levels are defined separately for ability to listen, speak, read and write. The Guidelines were first published in 1986, and subsequent revisions were made in 1999 and 2012 describing what individuals can do with a language in terms of speaking, writing, listening, and reading in a spontaneous context. Under the Guidelines, the ACTFL Oral Proficiency Interview (OPI), the ACTFL Oral Proficiency Interview Computer Test (OPIc), the Stimulated Oral Proficiency Interview (SOPI), the Computerized Oral Instrument (COPI), and the ACTFL Writing Proficiency Test (WPT) have been developed one after another. According to Liu (2017), Chinese enrolments in American institutions of higher education were 61,055 students in 2013. This number has more than tripled compared with 19,427 in 1990. The growing popularity of Chinese education in the USA has greatly increased the demands for assessments.

The OPI, being developed in the late 1980s, is a face-to-face interview or a phone interview between a certified tester and an examinee. It lasts for 10-30 minutes depending on the examinee's oral proficiency aiming to assess an examinee's oral proficiency through a natural conversation with the tester. In response to the demand of large-scale oral proficiency testing, ACTFL also developed a computer-delivered version of the OPI, namely OPIc in 2007. Compared with the OPI, the OPIc is more flexible in terms of test time and location. Test-takers can take it whenever and wherever there is a computer with internet connection. Test takers' responses can be automatically saved on the computer, and test raters can access the responses at anytime and anywhere to assess and provide a score. However, OPIc is not as natural as talking to a person in the case of the OPI and sometimes examinees may experience technical problems during the test. Later on SOPI was developed in order to remove the constraint of using certified interviewers and testing only one individual each time. In other words, SOPI can be administered by anyone to a group of examinees in a language laboratory with two-tape recorders. A master tape plays the instructions of tasks and a blank tape records each examinee's responses. With the tape recorders being gradually replaced by computers, it is only natural that SOPI has been replaced by COPI. Compared with the SOPI, the COPI has a larger pool of tasks, which gives examinees more task selection choices. In addition, the COPI has no time limits for examinees to think about or respond to a task, which gives them more control of the test. It also has a self-assessment to help examinees select tasks at appropriate difficult levels.

For the written skill assessment, the ACTFL Writing Proficiency Test (WPT) for Chinese has also been developed to assess CFL functional writing ability measuring how well a Chinese L2 learner spontaneously writes in Chinese. This test is available in both a paper-and-pencil format and a computerized format beginning with an introduction of the test followed by a warm-up activity at the novice level.

All these tests consist of four ranges of proficiency levels (novice, intermediate, advanced, and superior) in speaking and writing, and each higher level subsumes all lower levels. However, these tests do not suggest the number of characters or lexical items that learners should master at each level for assessing performance in the written language. This suggests that the ACTFL might not have addressed issues specifically related to the Chinese written language in relation to receptive and productive vocabulary competencies, which are crucial in benchmarking the Chinese language development (Liu, 2017). In addition, these tests are used in the USA, so they obviously have an American flavour in terms of the standards and expectations. Particularly, little research has been published so far to provide validation evidence for these tests (Liu, 2017). Hence it is unlikely that these tests will be used to achieve the purpose in benchmarking graduates' attainment in Australian Chinese tertiary programs. Now that the Chinese proficiency benchmarks used in non-Chinese speaking countries, namely CEFR in Europe and ACTFL tests in the USA, are not the best candidates for benchmarking students' attainment in Australian tertiary Chinese programs. It is only natural that Chinese proficiency tests developed in Chinese speaking countries or regions like Taiwan and mainland China should be considered.

2.3. Test of Chinese as a Foreign Language (TOCFL) in Taiwan

According to Chang (2017), the Test of Chinese as a Foreign Language (TOCFL) (in Chinese 华语文能力测验, pinyin: *Huáyǔwén Nénglì Cèyàn*) is Taiwan's Mandarin Chinese proficiency test designed for non-native speakers of Chinese. It is administered by the Steering Committee for the Test of Proficiency-Huayu (SC-TOP), which was established by the Taiwanese Ministry of Education in 2005. After some efforts from the research team of TOCFL trying to map the test to the CEFR scales, the new version of TOCFL has been developed and become available in 2013 with three proficiency bands: Band A, Band B, and Band C. Each of the bands has two levels. Therefore, there are a total of six levels: Levels 1 to 6 corresponding to the levels described by CEFR. Three types of TOCFL tests have been developed: 1) TOCFL Listening & Reading; 2) TOCFL Speaking; and 3) TOCFL Writing. Among the three tests, TOCFL Listening & Reading is the most popular one. The items on the test of each level are 50 multiple choice items for listening and 50 multiple choice items for reading. The test time is 2 hours. Test takers can choose the test levels best suited to them based on their Chinese language proficiency and learning background.

Despite the fact that TOCFL is a well-developed standardized Chinese proficiency test for non-native speakers of Chinese, usually only those international students who wish to apply for the Taiwan scholarship to be able to study in Taiwan will take TOCFL. Compared with the mainland China's Mandarin Chinese proficiency test HSK (see the section below), the number of test locations and test takers are quite limited. Although TOCFL tests are available in both traditional and simplified character versions, what "the tests have to deal with is the discrepancy in the use of Mandarin Chinese between mainland China and Taiwan,

which occurred due to the political separation in 1949” (Chang, 2017: p. 35). For example, some vocabularies are used only in Taiwan. In addition, the market value of TOCFL is much lower than that of HSK. Another reason that HSK is preferred lies in the fact that TOCFL tests are much more difficult to pass. For instance, the number of Chinese words are required for the new HSK level 1 is 150 while that of A1 level for TOCFL is 500 (Chang, 2017). It seems that CFL learners prefer HSK tests to TOCFL tests due to the reason that they feel more encouraged in taking HSK tests, particularly lower level HSK tests. In addition, there are more work opportunities in mainland China than in Taiwan. Therefore, it is reasonable to believe that TOCFL tests are unlikely to be used as an option for benchmarking students’ attainment in Australian tertiary Chinese programs.

2.4. The New Hanyu Shuiping Kaoshi (HSK) Developed in Mainland China

The new *Hanyu Shuiping Kaoshi* (HSK) test, namely the Chinese Proficiency Test for non-native speakers, was launched by Hanban² in an effort to better serve Chinese language learners internationally. The test is the result of coordinated efforts by domestic and international experts from different disciplines including Chinese language teaching, linguistics, psychology and educational measurement. The new HSK test incorporates the advantages of the original HSK while taking into consideration recent trends in Chinese language training by conducting surveys and making use of the latest findings in international testing. It is an international standardized exam that tests and rates Chinese language proficiency for Chinese L2 learners. It assesses non-native Chinese speakers’ abilities in using the Chinese language in their daily, academic and professional lives. It consists of six levels from level I to level VI. Listening and reading are assessed in all six levels while writing is assessed only from Level 3 and above. Lengths of the tests vary from 40 to 130 minutes, with a total mark range from 200 to 300. The corresponding oral exam, Hanyu Shuiping Kouyu Kaoshi (HSKK), consists of three levels: Elementary, Intermediate and Advanced. Both the new HSK and HSCK tests set explicit study objectives, allowing examinees to more effectively improve their Chinese proficiency with defined study plans and goals.

According to the official HSK website, its main principle for the test is to establish a “test-teaching correlation”, underpinning its aim to “promote learning and teaching through testing”. Furthermore, it is stated that the new HSK tests are standardised against the CEFR descriptions of proficiency levels: namely, Levels 1 to 6 corresponding to CEFR A1, A2, B1, B2, C1 and C2 levels. The alignment details are listed in **Table 1** below:

²Hanban is the colloquial abbreviation for the Chinese National Office for Teaching Chinese as a Foreign Language. It is a non-government and non-profit organization affiliated with the Ministry of Education of the People’s Republic of China. It is also known as the Confucius Institute Headquarters sponsoring the Chinese Bridge competition, which is a competition in Chinese proficiency for non-native speakers. According to the mission statement: Hanban is committed to developing Chinese language and culture teaching resources and making its services available worldwide, meeting the demands of overseas Chinese learners to the utmost degree, and to contributing to global cultural diversity and harmony.

Table 1. Alignment between new HSK and CEFR levels.

New HSK	Number of Characters	CEFR
HSK-Level 6	Over 5000	C2
HSK-Level 5	2500	C1
HSK-Level 4	1200	B2
HSK-Level 3	600	B1
HSK-Level 2	300	A2
HSK-Level 1	150	A1

Source: Hanban website: <http://www.chinesetest.cn/gosign.do?id=1&lid=0#>. Retrieved on the 20th January 2019.

Having incorporated the influential CEFR benchmarking principles and scales, the new HSK has been designed to tackle the unique linguistic features of the Chinese language including the testing of recognition and production of the Chinese characters. It explicitly sets the objectives of mastering a certain number of vocabularies at each level, as shown in **Table 1**, which allows the test takers to be able to improve their Chinese abilities in a systematic and efficient way. With these clear objectives, the new HSK can certainly serve as the benchmarking guideline for graduates' attainment in Australian tertiary Chinese programs. According to **Lu (2017)**, some CFL programs and professionals in the UK or European higher educations have already regarded the new HSK proficiency levels and standards as benchmarks for their CFL courses at different levels. Likewise, many students learning Chinese in such CFL programs have considered the HSK proficiency levels as their learning objectives to achieve the expected competence in the language, as the new HSK is a widely recognised standardised Chinese proficiency test. In fact, dependence on the new HSK tests for benchmarking is prevalent among universities worldwide. L2 learners of Chinese regard HSK as the Chinese version of TOEFL (Test of English as a foreign language) or the Chinese version of IELTS (International English Language Testing System) as for L2 English learners. It is a common knowledge that L2 English learners take TOEFL if they would like to study in America while they take IELTS if they would like to study in the UK or its Commonwealth countries like Australia and New Zealand. Nowadays L2 Chinese learners will take HSK if they would like to study or work in China. As a result, the new HSK tests have been greatly demanded by students and employers in order to assess jobseekers' competence in Chinese.

So far four options used in benchmarking foreign language development have been reviewed. CEFR is good for benchmarking European languages, but not yet suitable for L2 Chinese. The ACTFL tests used in America and the TOCFL used in Taiwan each has its own limitations and not used worldwide. Only the new HSK developed in Mainland China is the most influential test used worldwide in benchmarking CFL learners' Chinese proficiency. Therefore, a review of the literature indicates that the new HSK is the best option to be used in benchmarking graduates' attainment in Australian tertiary Chinese programs. Now the question is which HSK level the Chinese BA holders in Australian tertiary CFL programs should achieve, which is the research question of this paper:

Which HSK level should the Chinese BA holders in Australian tertiary CFL programs achieve?

3. The Study

In discussing the rationale for holding a benchmarking approach to higher education systems, the *OECD (2017)* report states clearly that “Benchmarking higher education system performance will contribute towards improvement across different higher education systems” (p. 55). This approach applies to the Australian tertiary Chinese programs. In order to identify which HSK level the Chinese BA holders in CFL programs in Australian universities should achieve, the author designed the study and empirically investigated this issue among CFL learners in a well-known university in Queensland Australia.

3.1. Design of the Study

The academic year at the university where data were collected, for the majority of undergraduate programs, is divided into two 13-week semesters. Semester 1 (S1) usually lasts from February to June while Semester 2 (S2) usually lasts from July to November. During each of the 13-week teaching semester, there is usually one-week break in between. Each semester is followed by a formal examination period of two and half weeks. The standard Bachelor of Arts (BA) program consists of three years of study, so does the BA in Chinese in this university. In the past few years, namely, 2016, 2017 and 2018, the new HSK test was incorporated into the CFL curricula. In particular, at the teaching weeks 1 and 13 of each semester, a new HSK mock test at different levels was implemented in all the Chinese language courses. This practice was designed to achieve four purposes: 1) explore for an appropriate Chinese BA language proficiency threshold in Australian tertiary CFL programs against the HSK benchmarks; 2) provide students with the opportunity to get familiar to the format and difficulty level of the HSK tests in case they would like to take an official test in the future; 3) let students roughly know their own Chinese proficiency level against the HSK benchmarks and chart progresses along their study; and 4) use the test outcome to assist the screening and placement process and support any adjustment when it is needed. Given that the Chinese BA program consists of Year I, Year II and Year III studies, the HSK level that Year III students attain will be considered as the benchmarking level for BA graduates.

3.2. Participants

All the students of Chinese in Year I, II and III enrolled in the years 2016, 2017 and 2018 were involved in the study. The outcome reported in the article was based on the results of the mock HSK tests conducted in Semester 2, 2018 with Year II and Year III students. Year II students consisted of 50 young adults including 24 males and 26 females with an average age of 20.3. Year III students consisted of 27 young adults including 13 males and 14 females with an average age of 23.1. The learner cohort was diverse, reflecting the contemporary Australian tertiary community demographic. A broad range of ethnicity and age was represented.

3.3. Data Collection Procedure

Based on the testing outcomes in 2016 and 2017, the author was made aware that the majority of Year III students could pass HSK level 4 and the majority of Year II students could pass HSK level 3. Therefore, in teaching week 13 Semester 2 of 2018, the last week of teaching in the semester, an HSK level 4 mock test was implemented with the Year III students on Monday 22 October 2018 while an HSK level 3 mock test was implemented with the Year II students on Tuesday 23 October 2018. The language teacher printed the HSK mock test papers and handed out to students. 10 minutes perusal was given to students before the listening audio file was played. The goal of the HSK mock tests was to see the passing rate of HSK level 3 test by Year II students and the passing rate of HSK level 4 test by Year III students. The author herself handed out the HSK mock test paper labelled as H31110 to Year II students and played the audio sound file with the computer in class for the listening part after 10 minutes perusal. Fifty three students participated in the test. After checking all the test papers handed in, 50 Year II students completed the HSK level 3 test. The three incomplete test papers were excluded from this study. A colleague of the author who taught Year III students handed out the HSK mock test paper labelled as H41001 to Year III students and played the audio sound file with the computer in class for the listening part after 10 minutes perusal. Twenty nine students participated in the test, but only twenty seven Year III students completed HSK level 4 test. The two incomplete test papers were excluded from this study. In sum, the participant number is 50 for Year II level and 27 for Year III level.

3.4. Test Results

The test results in the past few years indicate a general alignment of the Chinese proficiency attainment of the BA Chinese courses with the new HSK levels as shown in **Table 2**.

Table 2. Alignment of the BA Chinese courses with the new HSK levels.

Year/Semester	Teaching hours	Minimum number of vocabulary to master	New HSK levels Average
Year I/S1	6 hours × 13 weeks*	150	Level 1
Year I/S2	6 hours × 13 weeks	300	Level 2
Year II/S1	6 hours × 13 weeks	450	Level 2.5
Year II/S2	6 hours × 13 weeks	600	Level 3
Year III/S1	6 hours × 13 weeks	1000	Level 3.5
Year III/S2	6 hours × 13 weeks	1500	Level 4

*Note: The 6 hours teaching consists of two courses, namely a Spoken course and a Written course. Allocation of time in the two courses: 2 hours combined lecture + 2 hours Spoken tutorial + 2 hours Written tutorial. Each course is worth 2 credit points and a Chinese single major consists of 16 credit points. By using HSK level 2.5 to align with Year II/S1 courses, it does not mean that there is HSK level 2.5 test. It means that roughly half of the students can pass HSK level 2 test and half can pass HSK level 3 test. The average level is between level 2 and level 3.

The new HSK level 3 test has 80 test items in total including listening 40 items, reading 30 items and writing 10 items. The three skills were marked separately with each full score being a 100, as required. A passing score for HSK level 3 is 180. Among the 50 Year II students who did HSK level 3 mock test, 4 students failed the test. Therefore the passing rate was $46/50 = 92\%$. The average scores in listening, reading and writing achieved respectively are shown in **Figure 1**.

As shown in **Figure 1**, the average scores of listening, reading and writing are 87.3, 76.12 and 74.6 respectively. Listening has achieved the highest score. Scores for reading and writing are quite similar although reading is slightly higher. Lu's (2017) study has concluded that the writing items in the new HSK level 3 and level 4 are regarded too easy in comparison with listening and reading items. However, the writing score for the current study is still the lowest, which might indicate that writing is the most difficult skill to be achieved for the Australian CFL learners. Given that 92% Year II students passed the new HSK level 3 test, in the future, HSK level 3 can be used as a benchmarking yardstick to measure Year II students' performance.

The new HSK level 4 test consists of 100 test items in total including listening 45 items, reading 40 items and writing 15 items. The three skills were marked separately with each skill full score being a 100. Similar to level 3 test, the passing score for level 4 is also 180. Among the 27 Year III students who did HSK level 4 mock test, 3 students failed the test, that is to say, their total scores were below 180, achieving 116, 151 and 162 respectively. Therefore the passing rate was $24/27 = 89\%$. The average scores in listening, reading and writing achieved respectively are shown in **Figure 2**.

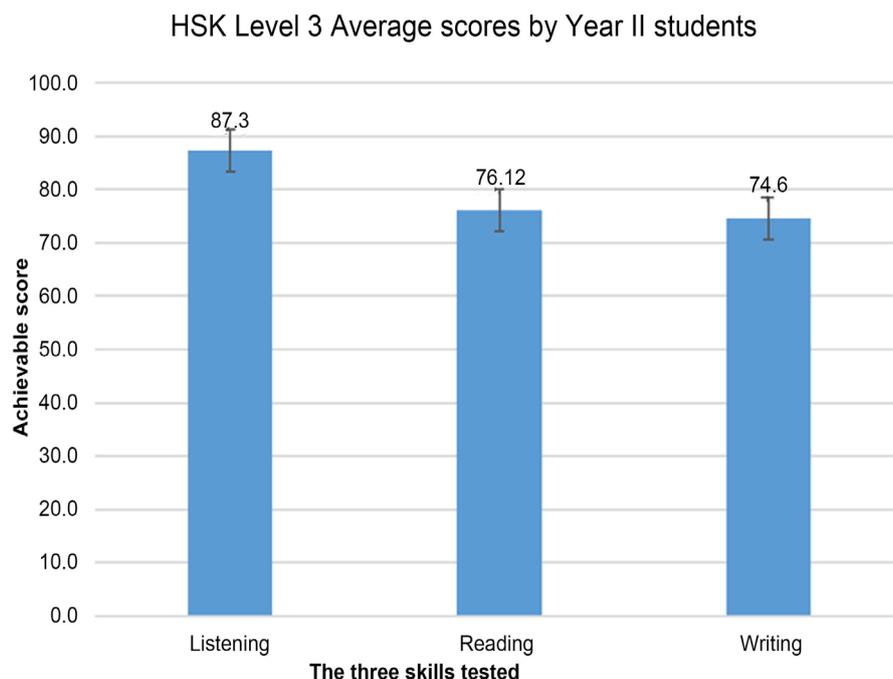


Figure 1. HSK level 3 average scores achieved by Year II students.

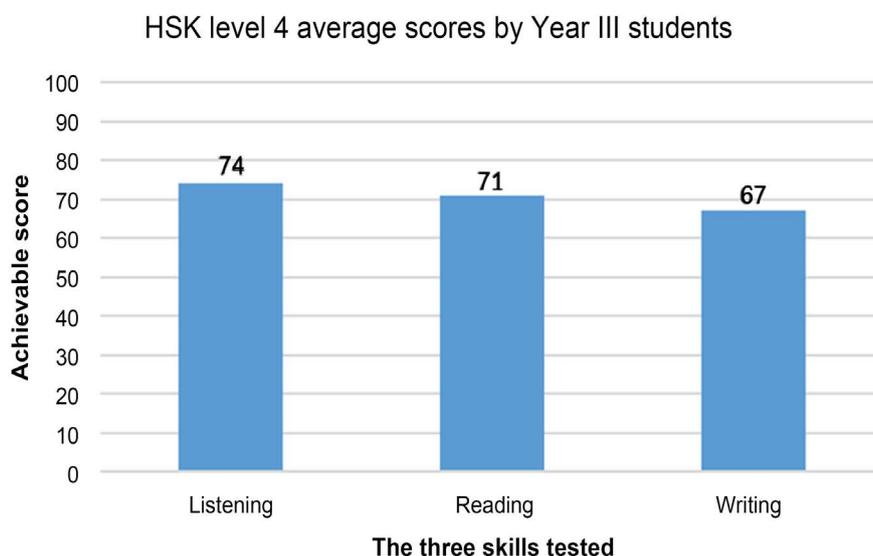


Figure 2. HSK level 4 average scores achieved by Year III students.

As shown in **Figure 2**, the average scores of listening, reading and writing are 74, 71 and 67 respectively. Similar to the HSK level 3 test results, listening has achieved the highest score, which is followed by reading. Writing, again, is the weakest among this cohort of CFL learners.

Based on the above test results, 89% of the Year III students, namely the BA graduates in Chinese language major at the Australian university where data were collected, passed HSK level 4 when they completed all the Chinese courses for their study. It is reasonable for this study to propose that HSK Level 4 be established as Chinese BA language proficiency threshold in tertiary CFL programs in Australia.

4. Discussion

Lu (2017) suggests that the level of difficulty of the HSK level 3 and 4 writing items should be increased in accordance with the CEFR standards if the testing organisation claims to align with them. She further explains that “By assigning less importance to the Chinese written language, the HSK exams could give the CFL learners the impression that competence in writing is not as important as other language skills” (p. 52). However, the findings of the current study seem to indicate that the difficulty levels of the writing items of the HSK level 3 and 4 papers should not be increased because the average score of writing was the lowest for both HSK level 3 and level 4 tests. The reasons for the low scores of writing, in comparison with listening and reading, could be that the productive skill is acquired slower or at a later stage of learning. In particular for the logographic writing system in Chinese, writing items should not be given too much weighting and should not be designed with too much difficulty. It is justified that the writing section of the HSK level 3 and level 4 tests is kept at a comparatively lower level with lower weighting compared with listening and reading.

If HSK level 4 is established as the Chinese language proficiency threshold for the attainment of BA graduates from Australian tertiary Chinese programs, it means that we are proposing an outcomes-based assessment and benchmarking approach in language teaching and learning. According to Brindley (2001), “Outcomes-based assessment appears to offer a number of advantages to the key stakeholders in educational programs, including transparency of reporting, alignment of teaching and curriculum goals, and sensitivity to individual needs” (p. 394).

In other words, a language learning quality management system should be developed in Australia. When there is a common yardstick, or when national standards are being developed and when the quality of learning outcomes becomes an objective to be measured and controlled, outcomes-based assessment, namely the assessment of teachers, learners, and institutions based on the results a particular program is able to deliver according to predefined criteria, becomes a distinct possibility (Schalock, 2001).

An objective will become much more achievable when it is explicitly set and required. There are a number of benefits for the language programs when an outcomes-based assessment is implemented. Bärenfänger and Tschirner (2008) state,

Focusing on the outcome of a program emphasizes the effects a program has on the life of a particular learner or a particular social group; creates transparency and raises the accountability of the people responsible for the success of the program; encourages the responsible use of resources; and helps teachers, learners, institutions, and politicians make informed decisions (p. 83).

There are so many advantages in employing outcomes-based assessment such as using the new HSK test to benchmark students’ language attainment in their CFL development. There is no reason for us not to use this approach in developing a quality management system for Australian tertiary CFL programs.

5. Conclusion

Benchmarking of learning outcomes in higher education has become a matter of increasing interest and importance. A review of the options used in benchmarking foreign language development such as CEFR indicates that the new HSK should be adopted for benchmarking students’ attainment in Australian tertiary CFL programs. In particular, this study aimed to find out which HSK level the BA holders in Australian tertiary CFL programs should achieve. Based on the empirical evidences reported in this study, it is proposed that HSK level 4 be established as the Chinese language proficiency threshold for graduates of Australian tertiary Chinese programs. The limitation of the study lies in that the participant cohort was small and the experiment was conducted only in one university. The findings call for more replications. Having this said, with such outcomes-based assessment and benchmarking criterion being proposed, the first step has been taken for setting up quality management system for Australian tertiary CFL programs.

Acknowledgements

This study was funded by the Chinese Ministry of Education (grant number 2014). The assistance of the author's colleagues Dr. Shenghsun Lee and Dr. Melody Chang in helping implement the new HSK tests to all the Chinese language courses is acknowledged with thanks.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Bärenfänger, O., & Tschirner, E. (2008). Language Education Policy and Language Learning Quality Management: The Common European Framework of Reference. *Foreign Language Annals*, 41, 81-101. <https://doi.org/10.1111/j.1944-9720.2008.tb03280.x>
- Bellassen, J. (2017). Preface. In L. Yang (Ed.), *Teaching and Learning Chinese in Higher Education: Theoretical and Practical Issues* (pp. 10-11). Florence: Routledge.
- Bellassen, J., & Zhang, L. (2008). The CEFR: New Concept and Its Implications and Impetus—Chinese Language Teaching at Its Critical Moment. *Chinese Teaching in the World*, No. 3, 58-73.
- Brindley, G. (2001). Outcomes-Based Assessment in Practice: Some Examples and Emerging Insights. *Language Testing*, 18, 393-407. <https://doi.org/10.1177/026553220101800405>
- CEFR, Council of Europe (2019). <https://www.coe.int/en/web/common-european-framework-reference-languages>
- Chang, L. (2017). The Development of the Test of Chinese as a Foreign Language (TOCFL). In D. Zhang, & C.-H. Lin (Eds.), *Chinese as a Second Language Assessment* (pp. 21-42). Chinese Language Learning Sciences, Singapore: Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-10-4089-4_2
- EBCL (2019). *European Benchmarking Chinese Language (EBCL) Project and Supporting Documents for Can-Do Statements*. <http://ebcl.eu.com>
- King, R., & Hoffmann, P. (2013). Benchmarking the Learning Outcomes of Formative Australian Engineering Degrees: National and International Perspectives. In *Proceedings of the 2013 AAEE Conference, Gold Coast* (pp. 1-9). Canberra: AAEE.
- Li, P., Shu, H., & Liu, Y. (2014). Neurocognitive Approaches to the Processing of Chinese. In J. Huang, A. Li, & A. Simpson (Eds.), *Handbook of Chinese Linguistics* (pp. 511-533). Wiley Online Library. <https://doi.org/10.1002/9781118584552.ch20>
<http://onlinelibrary.wiley.com/doi/10.1002/9781118584552.ch20/summary>
- Li, M. F., & Zhang, X. S. (2009). A Common Framework of Standard for Non Major Chinese Courses in British Universities. In X. S. Zhang (Ed.), *Chinese Language Teaching and Learning: Theories and Practices, Applied Language Studies II* (pp. 136-143). London: Sinolingua London Ltd.
- Liu, Y. (2017). Assessing Chinese in the USA: An Overview of Major Tests. In D. Zhang, & C.-H. Lin (Eds.), *Chinese as a Second Language Assessment* (pp. 43-66). Chinese Language Learning Sciences, Singapore: Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-10-4089-4_3
- Lu, Y. (2017). Exploring the Criterion-Validity of HSK Levels 3 and 4: Are Assessments and CEFR Standards Related? In L. Yang (Ed.), *Teaching and Learning Chinese in*

Higher Education: Theoretical and Practical Issues (pp. 35-56). Florence: Routledge.

<https://doi.org/10.4324/9781315520810-3>

Lu, Y., & Song, L. (2017). European Benchmarking Chinese Language: Defining the Competences in Written Language. In L. Yang (Ed.), *Teaching and Learning Chinese in Higher Education: Theoretical and Practical Issues* (pp. 13-34). Florence: Routledge.

<https://doi.org/10.4324/9781315520810-2>

OECD (2017). *Benchmarking Higher Education System Performance: Conceptual Framework and Data, Enhancing Higher Education System Performance*. Paris: OECD.

Schalock, R. L. (2001). *Outcome-Based Evaluation* (2nd ed.). New York and London: Kluwer Academic.

Wang, Z., & Niu, D. (2014). Australian Tertiary Chinese as a Foreign Language Teaching: Current Status and Prospects. *Overseas Chinese Education, No. 2*, 153-159.