The Development of a Local Online English Proficiency Test for Young Learners: Key Considerations and Challenges

Jirada Wudthayagorn[a], Chatraporn Piamsai[b*], and Phasphan Thanompongchart[c]

[a]Chulalongkorn University Language Institute, Bangkok, Thailand
[b]Chulalongkorn University Language Institute, Bangkok, Thailand
[c]Chulalongkorn University Demonstration Elementary School, Bangkok, Thailand
*Corresponding author: chatraporn.p@chula.ac.th

Article information	
Abstract	In Thailand, a series of policies have been implemented at all levels of Thai education. It is hoped that English language ability among Thai citizens will be improved. In primary education, following the English language education policy for primary school students introduced by the Office of Basic Education Commission (2016), English has been introduced to first graders onwards. They study English one hour every day, five days a week. When they finish grade 6, they should demonstrate their language ability at a CEFR A1 level. Various English tests have been used to check if the students are either proficient or able to achieve the learning goals. These tests include teacher-prepared tests at the classroom level and the Ordinary National Tests, or O-NETs, at the national level. Essentially, the O-NETs are administered to the sixth, ninth and twelfth graders. Interestingly, although the CEFR has been implemented in our educational system, O-NETs test scores have not been aligned with CEFR. This research project focused on developing a local English language proficiency test for young learners. In this article, key considerations, and challenges in the process of developing the test are presented and discussed. It is hoped that the presentation and discussion are timely and beneficial for those planning to develop their own tests for use in their own contexts.
Keywords	test development, young learner test, considerations, challenges

APA citation:	Wudthayagorn, J., Piamsai, C., & Thanompongchart, P. (2022). The development of a local online English proficiency test for young learners: Key considerations and challenges. *PASAA, 64,* 126–137.

Introduction

Improving English language ability among Thai citizens has always been a top priority in Thai education reforms. A series of policies have been implemented at all levels of Thai education. In primary education, following the English language education policy for primary school students introduced by the Office of Basic Education Commission (2016), English has been introduced to first graders onwards. They study English one hour every day, five days per week. When they finish grade 6, they should be able to demonstrate their language ability at a CEFR A1 level. These students continue their junior and senior high schooling for another six years. When they finish junior and senior high school, they should be able to demonstrate their language ability at CEFR A2 and CERR B1 levels, respectively. Put differently, when they finish their basic education (i.e., grade 12) from mainstream Thai schools, their English language ability should be at least at B1 level.

Various English tests have been used to check if the students are either proficient or able to achieve the learning goals. At the classroom level, teachers write their own tests. At the national level, the National Institute of Educational Testing Service (NIETS) produces several tests, including large-scale tests called O-NETs (Ordinary National Educational Tests). The O-NETs serve as achievement tests checking how far the students have achieved the learning goals. The O-NETs are administered when students have completed grades 6, 9, and 12. Interestingly, although we have adopted the CEFR into our educational system, the test scores have not been aligned with the CEFR. This can probably be explained by the fact that the major goal of the O-NET is to check the students' achievement; therefore, the test items are written based on the learning objectives stated in the core curriculum. Likewise, some drawbacks of O-NETs can be observed as far as the recent policy for primary school students (Office of Basic Education Commission, 2016) is concerned.

An English language proficiency test that is aligned with the CEFR is needed. This research project, thus, aimed to develop a local English online test for young learners at the upper primary school levels (grades 4-6) or 10-12 years old. The purpose of this test is to check the students' English proficiency using the CEFR descriptors as guidelines for test items. In the first project, Wudthayagorn, Piamsai, and Chairaksak (2018) focused on teaching, learning, and assessing the English language of Thai EFL grade 6 students. The highlight of the first project was the teaching programs that the school created based on parents' needs and tailored-made to students' English language ability. At the end of the school year, the Cambridge language assessment test for young learners (Starter Level) was used to assess Grade 6 students' language proficiency. At the same time, Wudthayagorn, Piamsai, and Chairaksak (2018) also drafted the English test, hoping that it would be a prototype local test appropriate for Thai EFL students. This second research project started in 2019. Initially, the authors aimed to upgrade the first draft of a prototype local test to an on-site computerized test. The process included designing the test, preparing the test blueprint, training item writers, editing and revising test items, contacting schools to pilot the test, and so on. Unexpectedly, the COVID-19 pandemic, coupled with a series of lockdowns and strict preventions, dramatically affected our research design and data collection. The COVID-19 pandemic is undoubtedly a threat. However, beyond this threat, we encountered several challenges in the design and implementation of the test. As governed by the ILTA Code of Ethics (2000, 2018), the authors believe that it is our responsibility to share our experience and knowledge during the time we carried out this research with our students, colleagues, and other language professionals.

In this paper, we will discuss considerations, based on our hands-on experience, when developing the test covering:
- High-quality English language proficiency standards and assessments
- Conceptual framework for test development
- Good practices for the assessment of young learners

Then, toward the end, the challenges that we have encountered regarding the development of a local online English proficiency test for young learners will

be discussed. Our discussion is hoped to benefit various groups of stakeholders, including, but not limited to, test developers, researchers, and practitioners.

High-Quality English Language Proficiency Standards and Assessments

According to the Assessment and Accountability Comprehensive Center (2009), three issues should be taken into consideration when it comes to rigorous English language proficiency standards and assessments. First, the English learner population are, in nature, diverse in their family, education, experience, culture, and economic status backgrounds. The Assessment and Accountability Comprehensive Center (2009) points out that learners may be disabled, affecting their development and proficiency in English. Such diversity in the learner population does not mean that there should be double standards, but only one standard should be employed. The Assessment and Accountability Comprehensive Center (2009) also suggests that "...standards should be conceptualized and designed to appropriately guide implementation — that is, to appropriately guide instruction that can be adapted and differentiated, as necessary..." [emphasis added] (p. 11).

Based on this suggestion, in a Thai context where the CEFR policy has been implemented, it can probably be advised that test items developed based on CEFR descriptors should yield positive washback to a greater extent. Second, the domain of English language use covers language skills and knowledge in listening, speaking, reading, and writing. The Assessment and Accountability Comprehensive Center (2009) urges us to consider how listening, speaking, reading, and writing are interrelated and leads to English language proficiency. Third, context factors – history, policies, and resources – play crucial roles when developing and implementing standards. Although in Thailand, education policies are centralized and top-down, several issues covering administrative systems, learner population, policy implementation and practices, structures, and resources are significantly different across the board. The Assessment and Accountability Comprehensive Center (2009) recommends that key persons such as policymakers, administrators, and teachers should take part in the review and implementation of standards. The following figure summarizes the three interconnected components contributing to high-quality English proficiency

standards and assessments, as proposed by The Assessment and Accountability Comprehensive Center (2009).

Figure 1

The Three Interconnected Components to Ensure High-Quality English Language Proficiency Standards and Assessments
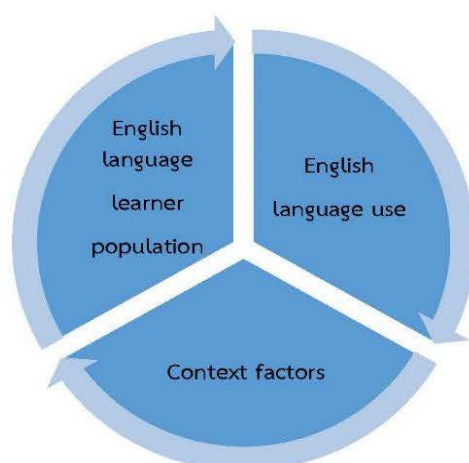
Developing a local online English proficiency test for young learners is a daunting yet essential and timely task. In order to achieve a high-quality language assessment, understanding the nature of young Thai EFL learners in mainstream schools, who make up a majority of the English language learning population, is necessary. From the beginning of students' lives, they participate in routines, activities, and schools that provide examples of cultural behaviors to observe and practice (Goodnow et al., 1995). In a broader view, these students do not bring only cultures but also beliefs and other related socio-psychological backgrounds into their learning. The next component is English language use, which is quite challenging to define and operationalize when doing research. Thus, to operationalize English language use, context factors must be considered, including language use contexts, school contexts, and policy contexts. These three components enable the authors to understand better how to achieve a better way of assessing and evaluating the students.

To accomplish the research goal in developing the test, the conceptual framework for test development is presented below.

Conceptual Framework for Test Development

A test is often used as an important instrument to produce empirical evidence – test scores – to identify how proficient students are. In the end, a collection of empirical evidence can be used as a proxy to identify if high-quality English language proficiency standards and assessments are met.

The body of knowledge related to test development has been substantially increased in the past decade. Detailed information on guidelines for good practices for test development has been made available online by many professional organizations such as the European Association of Language Testing and Assessment (EALTA) (2006), Educational Testing Service (ETS) (2009), and Cambridge English Language Assessment (2016). Researchers such as Bachman and Palmer (2010) also proposed a conceptual framework for test development that has become the foundation for the authors when developing the test in this research project.

Figure 2

The Conceptual Framework by Bachman and Palmer (2010)
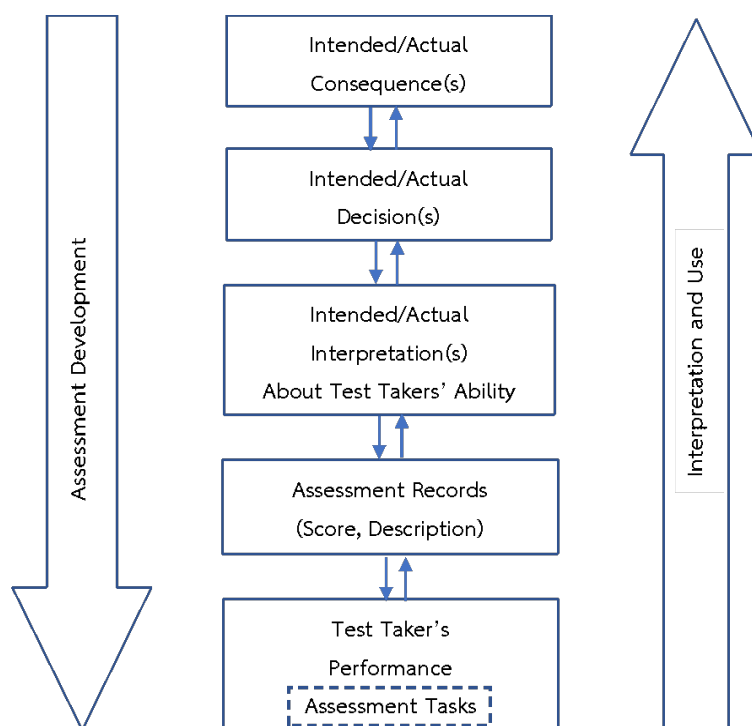
The framework proposed by Bachman and Palmer (2010) shows the assessment development process that portrays inferential links from

consequences to assessment performance. The four claims include intended consequences, intended decisions, intended interpretations of test takers' ability, and assessment records. To illustrate, when an assessment is developed, the first step is to consider the beneficial consequences that the assessment will bring to all stakeholders, such as test developers, test users, decision-makers, and institutions. The next question that can be asked is what kind of decision to make in relation to the intended consequences. Then, what needs to be considered is the aspects of language ability to be included in the assessment to promote accurate decisions about learners' performance. This means the constructs included in the assessment need to be specified. Moreover, to be able to interpret test takers' abilities or performance, evidence needs to be collected. Assessment records that refer to scores or verbal descriptions will have to be determined. This also includes scoring procedures and scoring criteria. Lastly, a decision needs to be made concerning the tasks used to assess test takers' performance as intended. For the interpretation and use of an assessment, the steps are in reverse, starting with test takers' performance on assessment tasks and moving to the assessment record, intended interpretations, intended decisions, and intended consequences. In every step of assessment development and interpretation, and use, assessment justification needs to be made by using an assessment use argument (AUA) and supporting claims with evidence and warrants. This helps increase the accountability of the developed assessment.

Good Practices for Assessment of Young Learners

Although the authors used Bachman and Palmer's conceptual framework as a guideline, we also acknowledge that assessing young learners differs from assessing adults. Papp and Rixon (2018) suggest that an English test should have the following components to achieve good practices for assessing young learners. First, it should not bring any harm but benefit. Papp and Rixon (2018, p. 596) state that the outcome of assessments should be useful. In other words, unlike traditional tests, the tests for young learners should not, for example, evoke fear. Second, the tests should demonstrate the reasoned relationship between the curriculum and the content of the tests themselves. Papp and Rixon (2018, p. 600) called it coherence and compatibility. Third, the outcomes of assessments should be helpful for decision-makers to make an appropriate decision. Also, young learners need to know what they may expect to gain from being assessed (Papp &

Rixon, 2018, p. 597). Fourth, tests should be practical to administer. It is what Papp and Rixon called feasibility instead of practicality. Lastly, the tests should be friendly for young learners.

Papp and Rixon (2018, p. 602) suggest that the number of test items, item types, and test time need to be calibrated to the needs and nature of young learners. Papp and Rixon (2018, p. 600) summarize that, based on these five mentioned components, the tests should produce good results, which can show

- when a learner's past performance or abilities are comparable with the teacher's perceptions of their progress.
- when a learner has met the required criteria along with others in a class.
- when a learner has performed 'better' than others in the same class.

Challenges in the Development of a Local Online English Proficiency Test for Young Learners

Five major challenges based on our experience are discussed below.

The Design of the Test

The target language use (TLU) domain or English language use coined by the Assessment and Accountability Comprehensive Center (2009) was extensively discussed among the research team and the item writers. The authors discussed how test tasks should be designed appropriately so that target language use can be assessed. Because all members of the research team and the item writers are non-native lecturers of English, it is quite unclear for us to operationalize the term TLU. Bachman and Palmer (1996, p. 18) define the TLU domain as the "situation or context in which the test taker will be using the language outside of the test itself." It is challenging to operationalize the term TLU and to narrow down the scope of "the language outside the test itself" because the test we have been developing is for Thai EFL learners who have had an English language classroom as a major place to use the language. Although they may read an English book, listen to music in English, and watch movies in English outside the classroom, chances to use English as frequently as ESL learners are incompatible. Thus, to make the test authentic and fair, the research team and the item writers tried to balance the use of the language, especially the vocabulary, in the text. Put differently, instead of focusing on Western holidays such as Christmas and

Thanksgiving, item writers insert texts centered around Thai holidays such as Songkran and Loy Kratong.

Ability-in Language User-in Context

With the introduction of ability-in language user-in context, Chalhoub-Deville (2003) begins the shift from the more conventional consideration of interaction (interdependent-interaction) to a more progressive, dynamic consideration (reciprocal interaction). In turn, it places a greater emphasis on the role of context regarding both task and person. Ability-in language user-in context revolves around the idea that in a given language assessment situation, the construct of second language ability is in reciprocal interaction with assessment tasks, which produces a specific performance on that assessment. Thus, a test developer is challenged to make a test task authentic. A multiple-choice task may not align with Chalhoub-Deville's (2003) concept. Task types that can elicit ability-in language user-in context need further research that fits the Thai context.

CEFR Descriptors Interpretation

Although the CEFR has been used in Thai education contexts recently, item writers have still interpreted the CEFR descriptors differently, especially at the A1 and A2 levels which are the focus of the test development. For example, at the A1 level, the descriptor reads, "can understand and use familiar everyday expressions..." [emphasis added], and students at the A2 level "can understand sentences and frequently used expressions..." [emphasis added]. The item writers found these two phrases at A1 and A2 levels to be indistinguishable. In order to solve this problem, the authors had to set up a series of meetings with the item writers to review test items and discuss if such test items were aligned appropriately to the CEFR levels. A subjective agreement helped solve this issue.

The Implementation of the Test

The COVID-19 pandemic has forced us to run various activities online. On-site computerized examinations as planned, thus, appeared completely impossible. As a result, the authors needed to pilot the test online. The most serious problem concerns students' accessibility to technology (e.g., internet connection) and equipment (e.g., computer, camera, and microphone). We did not know what basic requirements and specifications of the technology and equipment the students should have. For example, should the students have two devices: one for the test

and the other as CCTV to detect whether there is any cheating? So, to cope with this issue, we piloted the test online by producing the test and formatting the test items through Canva and Microsoft forms. This is because a technician was familiar with these programs, and the students were familiar with taking the test online at home through Microsoft forms.

Social and Political Dimensions of the Test

Teachers assess the students in order to collect related information for various purposes such as placement, diagnosis, achievement checking, or exit decisions. As mentioned earlier, the Assessment and Accountability Comprehensive Center (2009) reminds us to keep integrating all learners, English language use, and context factors in mind. These three components involve the social and political dimensions of the test in the first place. Once the test is finished and used to assess the students, the test itself functions as a "powerful tool in education and society, which may lead to unintended consequences" (Shohamy, 2017, p. 441). For example, the students may be successfully placed in an appropriate level as test scores indicated. However, they may be imposed with unintended ideas about failure versus success or incapability and capability. Such ideas are not only formulated by students, but also by other stakeholders such as teachers, peers, and parents. Thus, the social and political dimensions of the test are one of the important issues in developing the test.

Concluding Thoughts

Although the body of knowledge regarding test development is substantial, based on our hands-on experience in the development of the local test, the authors believed that it is necessary to share key considerations that may be useful for those who plan to develop a local test. The authors also encountered several challenges throughout the research project. As such, we presented and discussed key considerations for the development of the local test, together with the challenges that we have faced. The authors also take Shohamy's (2017) notion into consideration when she says: "It is the power of tests, especially those of high stakes, that causes test takers and educational systems to change their educational behaviors and strategies as they strive to succeed in tests given their detrimental impact" (p. 441).

It is hoped that our discussion will shed some light on how to develop a local test appropriate for our students in our contexts. Results from good tests could, in all probability, reflect the quality of instruction and quality of assessments.

About the Authors

Associate Professor Dr. Jirada Wudthayagorn was a recipient of the Royal Thai Government Scholarship to pursue her Ph.D. studies in Applied Linguistics at the University of Pittsburgh. She is the Director of Chulalongkorn University Language Institute and serves as President of the Asian Association of Language Assessment. She is interested in language assessment, language policy, and quantitative analysis.

Assistant Professor Dr. Chatraporn Piamsai is a lecturer at Chulalongkorn University Language Institute and the English as an International Language Program, Graduate School, Chulalongkorn University. Her research interests are in various aspects of language assessment, including English proficiency and English as a Specific Purposes test development, the use of technology in language assessment, and language assessment in a local context.

Assistant Professor Dr. Phasphan Thanompongchart is an English teacher at Chulalongkorn University Demonstration Elementary School, Faculty of Education. She has experience in teaching young learners. She found a successful key to working with kids in "Fun." She is interested in ESL and EFL teaching and curriculum development.

Acknowledgement

References

Assessment and Accountability Comprehensive Center. (2009). *Framework for High-Quality English Language Proficiency Standards and Assessments*. WestEd. https://files.eric.ed.gov/fulltext/ED524106.pdf

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford University Press.

Cambridge English Language Assessment (2016). *Principles of good practice: Research and innovation in language learning and assessment.* https://www.cambridgeenglish.org/Images/22695-principles-of-good-practice.pdf

Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing, 20*(4), 369–383.

Educational Testing Service. (2009). *Guidelines for the Assessment of English Language Learners.* https://www.ets.org/research/policy_research_reports/publications/publication/2009/hdhw.html

European Association of Language Testing and Assessment (EALTA). (2006). *EALTA Guidelines for Good Practice in Language Testing and Assessment.* http://www.ealta.eu.org/guidelines.htm

Goodnow, J. J., Miller, P. J., & Kessel, F. (1995). *Cultural practices as contexts for development.* Jossey-Bass.

ILTA. (2000, 2007). *ILTA code of ethics.* https://www.iltaonline.com/page/CodeofEthics

Office of Basic Education Commission. (2016). *การบริหารการจัดการเวลาเรียนภาษาอังกฤษชั้นประถมศึกษาปีที่ 1-3 ในสถานศึกษาขั้นพื้นฐาน.* [Time management of English language learning at primary education grade levels 1-3 in basic education institutions]. https://docs.google.com/viewer?a=v&pid=sites&srcid=cGhyYWUyLmdvLlnRofG5pdGVkcGgyfGd4OjU0OTliMjRjZGZjOGFkYjc

Papp, S., & Rixon, S. (2018). *Examining young learners: Research and practice in assessing the English of school-age learners.* Cambridge University Press.

Shohamy, E. (2017). Critical language testing. In E. Shohamy, I. Or & S. May (Eds.), *Language testing and assessment* (pp. 441–454). Springer International Publishing. https://doi.org/10.1007/978-3-319-02261-1_26

Wudthayagorn, J., Piamsai, C., & Chairaksak, P. (2018). *A research and development of CEFR-based English proficiency test for grade 6 students.* Research Report. Learning Innovation Center, Chulalongkorn University.