Performance-Based Assessment on Dispensing Drugs in English of Thai Pharmacy Students.

Miss Sasithorn Limgomolvilas

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in English as an International Language
Inter-Department of English as an International Language
Graduate School
Chulalongkorn University
Academic Year 2018

การประเมินความสามารถนักศึกษาเภสัชศาสตร์ในการใช้ภาษาอังกฤษจ่ายยา

น.ส.ศศิธร ลิมโกมลวิลาศ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต
สาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ สหสาขาวิชาภาษาอังกฤษเป็นภาษานานาชาติ
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561

ACKNOWLEDGEMENTS

I would like to first acknowledge my supervisor, Assistant Professor Jirada Wudthayagorn, Ph.D. Thank you for guiding me through great educational and mental support. Your viewpoints and time were truly valuable and I gratefully appreciate it. I also would like to thank the various committees for their extremely useful comments and questions. In addition, I profoundly appreciate the guidance and support from Dr. Ute Knoch. Her expertise allowed me to expedite my dissertation process with her great insight.

I would not be able to establish the path of this doctoral journey without the scholarships from Chulalongkorn University Language Institute (CULI) and the Royal Golden Jubilee Ph.D. program (RGJ). I also would like to acknowledge the Faculty of Arts, University of Melbourne for providing me with access to the library, online journals, programs, and courses. Last but not least, I would like to express my gratitude to my mom for both financial and mental support. I could not finish this degree without her understanding.

Sasithorn Limgomolvilas

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Chapter 1
Introduction

Background of the study

In Thailand where English is spoken as a foreign language, the duty of teaching English to students, who do not attend international program, is imposed on English instructors. According to the education reform in 2002, various policies were applied to increase the students' language ability before their university graduate (Wiriyachitra, 2002). Among them is to provide two extra English courses to the two English foundation courses for undergraduate students. Many universities including the institute the researcher in responded to this policy by adding English subjects to the requirement courses for bachelor students. Since the purpose of providing this additional English requirement is to equip the graduates with the language needed for their future career, the courses were requested to be English for Specific Purposes (Wiriyachitra, 2002). This means that for undergraduates to fulfill the requirement of English classes, they need to take two fundamental English courses and two English for Specific Purposes courses.

One of the courses the researcher was involved is English for Pharmacy Profession, which is one of two English subjects focusing on the English usage of the students' pharmacy profession. These English for specific purposes are arranged not according to the English institute but the requirement from the students' faculty. The professionals in the field can request the content and the skills they believe are appropriate for guiding their students into their occupation field. For pharmaceutical science at the targeted university, one of the requests is to foster the students to dispense in English. The faculty has in fact received comments from the drug stores, who train students, on equipping students with communication skills while dispensing drugs to patients.

As a response to the demand from the faculty, dispensing skills in English was added to the course. Simple assessment was first adopted to pay attention mainly on language and communication skills. However, dispensing skill is a specialized skill

that cannot be assessed solely on language and communication skills. It is a performance assessment that requires a candidate to integrate language knowledge and their healthcare communication skills in order to successfully perform the dispensing task.

Reasons for conducting this research were woven from many educational requirements. First of all, according to Ministry of Education in Thailand, all undergraduate students are required to study English for at least twelve units, which can be calculated as four subjects when counting three units per subject. At the targeted university, two English courses are provided to students in their first year resulting in two courses of English remaining for their specialized areas. Next, these English for specific purposes are arranged not according to the English institute but the requirement from the students' faculty. The professionals in the field can request the content and the skills they believe are appropriate for guiding their students into their occupation field. For pharmaceutical science at the targeted university, one of the requests is to foster the students to dispense in English. Tasks and rubrics were developed and used to assess pharmaceutical science students. Nonetheless, when comparing the task to the dispensing assessment in Thai conducted in their professional practice courses, the task and rubric were found to be different in a great degree in terms of validity, authenticity, interactiveness, and impact. For example, the dispensing assessment in Thai is rather authentic since it is conducted in a university drug store, where the students are allowed to do their professional practice under the supervision of their advisors. Thus, this research is conducted considering not only the needs of the students, but also the practicality of the assessment.

Assessment is used to accumulate information in order to determine decisions for each consequences (L. Bachman & A. Palmer, 2012). Those consequences can vary from the grade assigned to students to the student's ability to perform such task in real life. Among them exists an assessment that is crucial to our education, learning assessment (Oosterhof, 2003). A language teacher can infer the test takers' performance on a specific situation to the consequences for classroom assessment and stakeholders (L. F. Bachman & A. S. Palmer, 2012). In this

case, the stakeholders to be considered are language instructors, students, content instructors, the faculty, and the employer.

The approaches that might be suited for this classroom assessment are performance-based assessment and authentic assessment. The fact that these two approaches are referred as separated (Frey, 2014) means the performance-based assessment may not always be authentic. This is similar to Butler and McMunn (2006) who believe that authentic assessment can be 'performance driven', but provide no detail about performance assessment being authentic. Performance tests can be divided into three main categories: direct assessment, work sample methods and simulation techniques (T. McNamara, 1996). This classroom assessment research employs performance assessment using the simulation techniques or role-play, since direct assessment and work sample cannot be applied in this English classroom context.

Considering what and how useful the test is to the users should accompany the test based on L. F. Bachman (1990)'s model of proving the test usefulness, he suggests six elements that can attest the quality of the test usefulness, which are validity, reliability, authenticity, interactiveness, impact, and practicality. L. F. Bachman and A. S. Palmer (2012) believe that a language assessment should be drafted following the targeted performance considering what way the language is used and in which situation. When developing a test whether for academic or for specific purpose, we may need to consider the availability of the tests in use as well as the purpose of the test.

In assessing English for Specific Purposes (ESP), three qualities are raised as its characteristics: language use according to the context, precise specific purpose content, and an interaction between specific purpose language and specific purpose background knowledge (Douglas, 2001). With an issue of ignoring professional judgment in ESP testing (Elder et al., 2012), the necessity for collaboration work to include the specific purpose background knowledge has been acknowledged since the early development of Language for Specific Purpose (LSP) (Elder & McNamara, 2016). Although separating the language use from the specific content has long been argued (Brunfaut, 2014; Douglas, 2013), we need to be reminded that no clear

boundary between LSP and General Language Proficiency can be specified and that it is the matter of which section along the continuum the focus of the language is in (Brunfaut, 2014). In fact, a performance test can vary from ‘very general to very specific’(Douglas, 2001).

The issue of specific purpose content in assessment is a vital element to consider in LSP as researchers introduced a couple definitions. Macqueen, Pill, and Knoch (2016) investigated ESP as a concept of boundary objects, which refers to the matter meaningful and relevant to the people in contact with the group. Another definition is the indigenous criteria, which was first used by Jacoby (1998) meaning to capture the matter relevant to the context (Pill, 2016). It was described by Pill (2013) as ‘criteria applied to performance in a particular context by insiders who share a common perspective’. Regardless of different name, the concepts are meant to describe the relationship of the language use and the context it belongs to. In one aspect, the term entails the authenticity, one of the elements of L. F. Bachman (1990) test usefulness.

Following a guidance to how the ESP or LSP assessment can be developed, O'Sullivan (2012a) categorizes it into two ways: the field of its use (such as business and law) and the purpose (work, immigration, and study) of the test. This is somehow similar to T. McNamara (1996)’s concept on language performance assessment that the language is used as a medium of the performance as well as the aim of the assessment. In that sense, the aim of the assessment can be reflected in how similar it is to the context it is being used in of how authentic the test can be conducted. Aiming at assessing the English proficiency to be used in the work situation, this study is hence composed of components leading to language use in the real pharmacy context.

Setting the authenticity of the performance-based test as the aim, studies were conducted on Occupational English Test (OET), which has been used to assess and certify health professionals who wish to work in Australia, New Zealand and Singapore, with its criteria following the law to measure on only linguistic ability and not on professional competence. A recent set of studies (Elder & McNamara, 2016; Macqueen et al., 2016; Woodward-Kron & Elder, 2016) was conducted to improve the

assessment of OET with the qualitative aim on exploring the professional perspective on the criteria and standard they use to judge the performance. To increase the authenticity of the ESP test, recent qualitative studies (O'Hagan, Pill, & Zhang, 2016; Pill, 2016; Woodward-Kron & Elder, 2016) yielded adjustment to the test including the approved and attested indigenous criteria on communication aspects derived from expertise judgment using partial credit scale, which are *clinician engagement* and *management of interaction*. In the OET context, the two new criteria added were to pursue the authenticity of the test as to fill in the criteria that are relevant to the professionals and can be judged by language experts.

For content assessment in a professional community, collaboration work between the assessment and the content experts is essential in order to balance the score weight between the linguistic content and the professional knowledge (Byrnes, 2008). Various classroom research has shown that some collaboration between the fields is essential to reflect indigenous criteria since the focal points of the studies tend to shift according to the specialty of the researchers. Two studies by pharmacy experts (Kimberlin, 2006; Schwartzman, Chung, Sakharkar, & Law, 2013) on overall content matters to teaching communication of pharmacy schools in the U.S., Canada, and Puerto Rico reveal no information on linguistic detail but rather on the general communication skills. In some cases, pharmacy experts (Collett, Rees, Mylrea, & Crowther, 1994) did not include linguistic features in the patient-communication assessment while some (Parkhurst, 1994; Sibbald, 1998) included a slight information in the verbal expression. A possible problem found in these studies is the language background of the participants, as most of them are native speakers of English, including the linguistic detail may not be needed for their group of participants. Similar to what Manias and McNamara (2016) found, the background of the participants as native speakers or passing university language requirement did not bring up the issue in language and communication. Not many studies are found on the EFL students taking the performance-based assessment related to medical or pharmacy students. In other words, the communication skills were commonly inclusive in performance-based assessment while the language use is not much of

the essence even in the case of non-native speakers of English in previous research, which addressed general listening and speaking skills.

The focus of previous classroom assessment research (Graham & Beardsley, 1986; Hyvärinen, Tanskanen, Katajavuori, & Isotalus, 2012; Parkhurst, 1994; Sibbald, 1998) relating to the assessment for students in the pharmacy field depends on the researchers' expertise. As a matter of fact, the experts tend to use their knowledge and professionalism as their baseline while the linguists rely on the language and communication aspects (Douglas & Myers, 2000). Some involvement from the specialists in the field can add a greater range of authenticity to the rubric (O'Hagan et al., 2016) in order to develop the rubric to fit the study of this classroom assessment. A performance-based test thus should not heavily focus on the content but weave in the content in a parallel combination. This study considers the task and rubric development with the concern on the language use in its field of pharmacy area and the purpose on assessing the students' performance after the content is taught in class.

Collaboration between the language instructor and content instructor using quantitative and qualitative method is thus essentially adopted to explore the indigenous criteria of the target language use of the group, which serves the aim of assessing the performance of dispensing drugs in English of Thai pharmacy students. Developing tasks and a rubric specifically for classroom assessment through integrated approach would best respond the researcher's needs.

Research questions

1. What are the tasks for measuring English oral proficiency in dispensing drugs of Thai pharmacy students?
2. What is the rubric for measuring English oral proficiency in dispensing drugs of Thai pharmacy students?
3. What is the extent of this drug dispensing task's test-usefulness?

Objectives of the study

The study aims:

1. To create and validate the tasks for measuring English oral proficiency in dispensing drugs of Thai pharmacy students
2. To create and validate the rubric for measuring English oral proficiency in dispensing drugs of Thai pharmacy students
3. To establish the extent of this drug dispensing task's test-usefulness

The scope of the study

The scope of this study was on assessing English oral proficiency in dispensing drugs at a drug store situation in Thailand only as the study responded to the needs in developing the assessment required by the faculty. As a result, the instruments were developed based on the local situation in Thailand. Other communication situations of pharmacists, such as dispensing at a hospital, were not included due to the difference in social discourse.

Definition of the operational terms

Assessment instruments refer to test and rubric designed specifically in this study for measuring pharmacy students' ability in communicating with the patients effectively and professionally as to obtain needed information for dispensing the appropriate drugs for them.

P*erformance-based assessment* - In this study, it is the procedure in which the rater who acts as a pharmacist communicates with the patient effectively and professionally as a simulation at the drug store to obtain needed information in order to dispense the drugs to the patients according to the pharmaceutical guideline in a designed assessment tasks developed by the researcher.

Test-Usefulness – It refers to six elements to prove the quality of the test as to what and how useful the test is. The six elements are validity, reliability, authenticity, interactiveness, impact, and practicality.

Thai pharmacy students - Thai EFL students who have completed required drug courses and internship courses and are attending their fifth year at the faculty of Pharmaceutical Science at a university in Bangkok.

Participants of the study

This study employed pharmacist experts, pharmaceutical students, who completed courses needed for dispensing drugs, and language instructors. The participants were divided into three main groups according to the instruments. The participants for semi-structure interview were 3 students and 4 instructors. The questionnaire recruited 132 fifth-year pharmaceutical students of a university in Bangkok and 21 pharmacist experts. The rubric employed the score of 147 students and 6 raters.

Research instrument

Three main instruments for this study were questions for semi-structure interview with pharmacy students, a questionnaire for pharmacy students and pharmacy experts, and questions for interview with raters. This research generated information from rubrics reported in secondary research, rubrics of Thai dispensing assessment and semi-structured interview questions in developing a questionnaire. The questionnaire was aimed to inquire information and opinion from pharmacy students and informant specialists regarding oral skills suitable for pharmacist-patient communication while dispensing drugs. The result from the questionnaire was used to establish appropriate tasks and rubric for the purpose of this research. In addition, two sets of semi-structure interview were applied to students and language instructors, who were the raters of the task performance. The interview with the students was conducted to gain qualitative detail on the task development, while the one with raters was to obtain feedback and insights of the experience and any recommendation they have.

Data collection

In addition to data collection in the interview and the questionnaire part, a simulation technique or role-play was used in the test administration. Students' performance as a pharmacist interacting with an instructor, who acted as a patient, was observed and recorded for rubric assessment use. Each video record of students' performance was roughly five minutes. After the task performance was assessed, the student score from six raters were collected.

Data analysis

Two steps of data analysis were conducted in order to develop tasks and a rubric. The first step was analyzing the data from the questionnaire by computing information to find agreements and differences among pharmacists and students using ANOVA and post hoc test analysis. The result was then discussed with informant specialists to validate the task and the rubric. The second step was to analyze the students' score through Many-Facet Rasch Measurement (MFRM) in order to validate the rubric and to explore other factors that might impact the score, such as the intra-rater and inter-rater reliability.

The significance of the study

It would be beneficial to develop a rubric and a test as a classroom assessment that would facilitate language instructors in evaluating the dispensing skills in English of Thai pharmacy students since this does not involve only the students' knowledge in pharmacy but it also requires them to cooperate their communication skills and English language skills in performing the task. As Douglas (2001) stated that one components of assessing ESP is specific knowledge and that it also involves its interaction with the language use, this study proved that the collaboration with informant specialist is essential and useful in validating that the specific knowledge and its interaction applied in the test.

The aim of this pharmacy test can go beyond the internal use of one university to a possibility of a national proficiency test for pharmacist as none has been developed specifically for Thais. If that occurs, such application of validated

classroom assessment can encourage the development of classroom test to the higher standards, which leads to more reliable result of the students' performance.

In the past, local tests could not be compared to international test in terms of quality; nonetheless, that statement is not accurate anymore (O'Sullivan, 2012b). English is used around the world that the usage is not exactly a foreign language but rather a 'near universal basic skill' (Bolton, 2008). Hence, relying solely on the international test would not be the case when conducting classroom assessment. Apart from believing that local tests providing some local aspects can be a great assessment with the platform and content it brings in, O'Sullivan (2012b) stated that 'Local test can dominate specific target', which conforms to indigenous assessment proposed by Jacoby and McNamara (1999) and what Douglas (2000) stated about including the specific knowledge in assessing English for specific purpose. This agreement led to the idea that a local test that assess the learners based on their performance can occur specifically for Thai pharmacy students. The outcome of this study could provide a potential model for a performance-based test, which can be utilized as an initiative in other healthcare subject areas.

Chapter 2

Literature Review

Language Assessment Concept

Assessment is widely used for numerous purposes in daily life (L. F. Bachman & A. S. Palmer, 2012; Green, 2014; Oosterhof, 2003). Test, which is a part of assessment, is used to accumulate information in order to determine decisions for each consequence (L. F. Bachman & A. S. Palmer, 2012). Among them exists an assessment that is crucial to our education, learning assessment (Oosterhof, 2003). Even though assessment may not be prominently recognized, its role in language education is undeniably teeming to teachers, learners (Green, 2014) and stakeholders.

According to L. F. Bachman and A. S. Palmer (2012), assessment is crucially related to decisions and consequences. Provided information, assessment is a means for making decisions in order to come up with consequences in teaching and learning. The relationship is illustrated in a figure below.

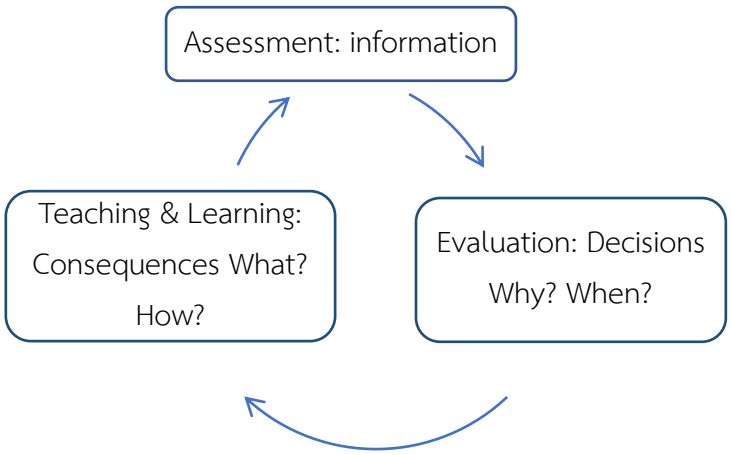


Figure 1: Assessment, evaluation, and teaching and learning

Decisions in assessment terms can be divided into three types ranging from effects on one person to a larger crowd: individuals, program and research. First of all, decisions about individuals or micro-evaluation deal with testing the performance of a person, such as job selection, placement test for a course, achievement test and predicting the performance of individual in the future. Second, decisions about

program, macro-evaluation, involve formative, which is a way to advance the program, and summative, which allows the program to continue or suggests a new program. Last, decisions about research allow us to plan the questions and methodology for future research. It can also enhance our understanding on ‘a particular language phenomenon’(L. F. Bachman & A. S. Palmer, 2012). Essentially, the consequences, aiming for beneficial ones, are the result of those decisions on individuals, programs, institutions and stakeholders.

Classroom Assessment Concept

For language assessment in language instruction and learning, the majority of the decisions tend to be formative, which enable the adaptation to occur during the learning of a course or in classrooms (L. F. Bachman & A. S. Palmer, 2012). Assessment; however, varies from classroom assessments to standardized tests as Green (2014) stated his definition of language assessment that ‘language assessment involves obtaining evidence to inform inferences about a person’s language-related knowledge, skills or abilities’.

Many researchers stated the concept of classroom assessment in their own unique way. Brown (1996) held the use of decisions of classroom assessment as achievement test and diagnostic test. Hill and McNamara (2012) stated the concept of classroom assessment as ‘any reflection by teachers (and/or learners) on the qualities of a learner’s (or group of learners’) work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes.’ Another idea by Frey (2014) is ‘…tests and other formal and informal data-gathering strategies used by teachers to assess their students and themselves. The assessments can be used before, during, or after instruction, at any time during the learning process.’ Relating classroom assessment to the curriculum for a course, Butler and McMunn (2006) believe that ‘A true performance or product assessment, conversely, demonstrates student mastery of a portion of curriculum. ‘Without good assessments, we cannot know whether effective learning has occurred’ (Oosterhof, 2003).

Classroom assessment can be conducted through various activities in which the degree is in between formal and informal depending on the arrangement are semi-formal exercises and quizzes, observations and portfolio assessment (Green, 2014). In addition, an assessment does not need to be proceeded by only instructors, but the learners can also initiate the learning through self-assessment, which allows them to consider their language use, and peer assessment, in which learners can assess each other's performance.

Classroom assessment can include a variety of roles as Oosterhof (2003) mentioned the followings: preliminary evaluations, formative evaluations, summative evaluations, and diagnostic evaluations. These roles of assessment are somehow linked to the purpose of the assessment. In fact, the aim of the curriculum should be associated with the outcome, which is what the learners can (Butler & McMunn, 2006). Thus, specifying the aim of the assessment should be prioritized when developing an assessment. Apart from that, clarifying the targets or skills at the beginning of the course to students can aid the students to fulfill 'the assessed targets' (Butler & McMunn, 2006).

Focusing on the purpose with aspect of intention, L. F. Bachman and A. S. Palmer (2012) asserted that it is essential to recognize the role of assessment as to what relationship it holds in the classroom. The tasks utilized in the classroom can be either for assessment, for teaching or for both. They categorize the role of classroom language assessment into two modes: implicit and explicit.

Although part of this research assessment can be used for a micro-evaluation in case of test takers who wish to improve their communication skills, the main decision of this assessment research is to conduct a macro-evaluation on a part of an English course taught as a second language. It aims to develop the instruments for assessing English communication skills of Thai pharmaceutical science students at a university in Bangkok, which might be further distributed for widely usage outside the classroom. As an illustration, the rubric can be utilized to assess the students' capability during their internship program before the class started as a pre-test.

Table 1: Role of assessment in teaching and learning (Bachman & Palmer, 2012)

Mode	Characteristics	Purpose
Implicit	Continuous Instantaneous Cyclical Implicit: both teacher and students may be unaware that assessment is taking place	Formative decisions, e.g.: Correct or not correct student's response Change form of questioning Call on another student Produce a model utterance Request a group response
Explicit	Clearly distinct from teaching Explicit: both teacher and learners are aware that assessment is taking place	Summative decisions, e.g.: Decide who passes the course Certify level of ability Formative decisions, e.g.: *Teacher*: Move on to next lesson or review current lesson *Teacher*: focus more on a specific area of content *Student*: spend more time on particular area of language ability *Student*: use a different learning strategy

The mode of this assessment research is identified as explicit classroom assessment since assessing pharmaceutical science is a separate task occurred at the end of the unit. While the content in the unit is geared toward aiding the performance in communication skills with some exercises on the tasks between the instructor and peers, the scored task is made apparent to test takers as they are informed of instruction and detail beforehand. The assessment; however, do not assign any grade to pass or fail any test takers. The score attributes to only a part of the grade assigned.

Approach to language assessment

The main essence in the process of developing a test for L. F. Bachman and A. S. Palmer (2012) is Assessment Use Argument (AUA), which is basically a combination of proof used to identify the relationships between a test taker's performance on an assessment, an assessment record, an

interpretation about the ability, the decisions and the consequences. Conforming to AUA as the focal point in developing a language test, L. F. Bachman and A. S. Palmer (2012) grounded their practice on four fundamental principles. The test developers and users need to be equipped with the followings evidence: a rationale for the uses on decisions and consequences of the test provided to stakeholders, AUA that explains the relationship between the performance and interpretation and intended uses, the indications that can confirm the AUA statement, and the joint effort from the stakeholders while developing and trialing the assessment.

Apart from establishing PRICE (planning, reflection, improvement, cooperation and evidence) as a step to effective assessment, Green (2014) also raised the concern of useful assessment, which depends on four qualities: beneficial consequences, validity, reliability and practicality. This is similar to what most researchers concern of, as well as the AUA when relating them to Bachman and Palmer's approach. First of all, the beneficial consequences are a part of the rationale for the uses on decisions and consequences of the test provided to stakeholders. Next, a proof of AUA that explains the relationship between the performance and interpretation and intended uses is in the sense of validity and practicality of the test. Last, the indications that can confirm the AUA statement is related to the reliability of the test. The joint effort from the stakeholders while developing and trialing the assessment can be compared to step to effective assessment in terms of planning, reflection, improvement and cooperation.

Approach to classroom assessment

Oosterhof (2003) stated that in order to obtain the appropriate interpretation, the right parameter and the type of assessment should be carefully selected. Classroom assessment can include a wide range of approaches used differently depending on the aims of the assessment (Oosterhof, 2003). According to Frey (2014), modern classroom assessments are categorized into five main approaches: traditional paper-and-pencil

assessment, performance-based assessment, formative assessment, authentic assessment and universal test design. Similar to Butler and McMunn (2006) who believed that authentic assessment can be 'performance driven', but provide no detail about performance assessment being authentic. Preparing a classroom assessment does not require the developers to choose only one out of these approaches (Frey, 2014). In fact, a combination of various approaches that suits the aim should be the concern (Frey, 2014). 'Modern classroom assessment is multifaceted', Frey (2014) suggests that the aim of the assessment should be a pivotal point that leads to what approaches to be utilized.

Applying ESP performance-based assessment in classroom context

While the definition of classroom assessment encompasses both formative and summative assessment, the focus of this current research is on the latter. Assessment of learning or summative assessment is done after the content is taught. It is the process of collecting information to judge whether learners acquire the knowledge (Frey, 2014). Performance-based assessment can be applied to assessment for learning, particularly when assessing skill or ability (Frey, 2014). It is considered the best approach to assess skills or ability (Frey, 2014).

The origin of performance-based assessment is from the testing for occupations in the scientific field, which is to test English for Specific Purposes (ESP), since traditional assessment on general English content like paper-based assessment did not directly response the target language use of the content specific field. Performance-based assessment began a role in language testing in 1950s with the very first test claimed to be in such category as speaking tests on various foreign languages for US personals. The use of performance assessment became widespread in 1960s when the number of foreign students wishing to study in England and North America expanded. The rationale for performance assessment was better defined in

1970s after the movement for communicative competence led by Canale and Swain (T. McNamara, 1996).

Established as consideration in communicative ability not ‘abstract demonstration of knowledge’ (T. McNamara, 1996), the theory of performance-based assessment is mainly divided to two main approaches: the work sample approach and the cognitive approach. The prior tradition stems from non-language context and does not pay much attention to the linguistic factors, while the focal point of the latter tradition is on the ‘quality of the execution’ of the performance and the ‘underlying state of language knowledge’. In second language performance assessment, the language is used as a medium of the performance as well as the aim of the assessment. (T. McNamara, 1996)

The distinctive characteristics that differentiate performance tests from the traditional fixed response are the display of behavior under ‘stimulus characteristics of tests’, the judging process, the finished products, and especially the ability to cover the detail of the ‘performance process’. Jones (as stated in McNamara, 1996, p. 16) made a remarkable point that when measurement cannot occur with the means of pen-and-pencil tests; performance assessment would provide a ‘better interpretation’ of test takers’ proficiency. While the relationship between the test task and reality is another notable aspect of performance assessment, Fitzpatrick and Morrison (as stated in McNamara, 1996, p.17) theorizes two features contemplating this relationship, which are the degree of reality of the simulation and the relevance of the performance in the simulation to performance in the criterion.

Performance-based tests can be divided into three main types: direct assessment, work sample methods and simulation techniques (T. McNamara, 1996). The degree of authenticity of these performance types varies respectively from more to less. First of all, direct assessment is a process of screening and selecting admission to the workplace or a place for study. To be able to perform the assessment, a provisional admission is required to

access the particular work setting. Second, work sample methods allow the test takers to use the workplace settings. Since this type of performance aims to fulfill the standardization of assessment, the tasks set are directed. Finally, simulation techniques contain the least authenticity among the three. The tasks are based on some amount of 'abstraction from workplace reality'. In conclusion, the difference among the three types of performance tests is the control of the performance task set, which results in their authenticity.

This classroom assessment research employs performance-based assessment in ESP field using the simulation techniques. The first two techniques, direct assessment and work sample, cannot be applied due to the normal condition of classroom assessment, where the instructions and assessment usually occur in the classroom and that professional settings are not easily accessible for language instruction and assessment. In addition, dispensing skills in the condition of direct assessment and work sample would require the real patients, who might not be willing to be in a part of the assessment. Even in the case of a university drug store where most patients know that they might consult with both the pharmacist in charge and the student pharmacists, the pharmacist trainee and trainer need to be aware of the patient's time and privacy. The time spent with patient is limited to no more than five minutes to keep the situation as if the patients are visiting a normal drugstore. Simulation is thus considered the best possible techniques to apply to this classroom assessment on the communication skills of pharmacy students during the dispensing session.

Test Usefulness

One of the essences in developing the test is its usage, which is what and how useful the test is to the users. Based on L. F. Bachman and Palmer (1996)'s model of proving the test usefulness, they suggest six elements that can attest the quality of the test usefulness: validity, reliability, authenticity, interactiveness, impact and practicality. It is essential to note that the most significant elements to be

considered as an indicator in terms of measurement for making decision and interpretation are validity and reliability (L. F. Bachman & Palmer, 1996).

Validity

'Does a test measure what it is supposed to measure? If it does, it is valid.' (Lado, 1961) Validity deals with the extent the scores from the assessment can be interpreted as the performance of the test taker. If the test is valid, the score obtained from the test can offer 'an accurate representation of a candidate's level of language knowledge or skills' (Weir, 2005). The important consideration is '*how* and *how well* we can generalize from the test performance to the criterion behavior' (T. McNamara, 1996). Validity can be observed in many ways as Weir (2005) states that 'validity is multifaceted'. The fact that validity is categorized into many types reflects various means of measuring validity (Alderson, Clapham, & Wall, 1995). In general, validity can be established through the following four types: content validity, concurrent validity, predictive validity, and construct validity.

Content validity

Based on Weir (2005)'s socio-cognitive approach, content validity is another term for context validity, which involves the degree of the sample text being a representative of the general. Content validity is described by Kerlinger (1973: 458) as 'the representativeness or sampling adequacy of the content- the substance, the matter, the topics- of a measuring instrument' (Alderson et al., 1995). This sort of validity concerns with the perception of the experts and that their judgments are considered significantly no matter what directions they point to.

Although the normal process of validation is for the experts to follow the system by comparing the test items with the test specification, what usually occurs is the non-systematic validation of experts with scarcely preparation resulting in the test that cannot represent the general as expected (Alderson et al., 1995). If aiming for content validity, clear test item specification should be used for matching the items and the specifications (J. D. Brown, 1996).

J. D. Brown (1996) suggests making a sound plan to prove content validity by creating a clear test specification, writing more items than what it needs, administer and revise the test using proper strategies, and selecting the true expert. First of all, a clear test specification should contain a general description of the test, a sample item, stimulus and response attributes, and supplemental lists. All of these details can guide the test developer through the process better. Second, when creating items, their number should be at least fifty percent more than the test needs since item selection involves deleting some items during the revision. Next, administering and revising the test using the appropriate strategies such as descriptive statistics, reliability coefficient and SEM. Last, appointing true experts, who have similar viewpoints.

Concurrent validity

J. D. Brown (1996) refers to concurrent validity as criterion-related validity. This validation uses other forms of measurement taken roughly at the same time of the test to indicate the test taker's ability. Various forms of measurement can be the parallel tests, students' self-assessment, rating by teachers or specialists on relevant areas. The value used for concurrent validity is correlation coefficient, which ranges from -1.0 to +1.0. The reliability of the concurrent validity depends on the selection of external measurement. The other forms of measurement should be reliable, valid and test the students on similar language ability.

Predictive validity

Different from concurrent validity in terms of time of gathering external measures, predictive validity uses information after the test was administered to predict the future performance of the test takers. This validation is mostly employed in case of people who take TOEFL and IELTS, which is to predict their ability to use English in English speaking countries. Some other forms of predictive validation are Grade Point

Average (GPA) and language performance of student teacher. Nevertheless, not high correlation can be expected from predictive validity since the result can stem from other factors.

Construct Validity

'Construct validity embraces all forms of validity evidence.' (Messick, 1989).The last validity which most people consider as a contribution other types of validity is construct validity (Alderson et al., 1995). L. F. Bachman and Palmer (1996) referred to term construct validity as 'the meaningfulness and appropriateness of the interpretations that we make on the basis of test scores'. The term construct validation is employed when Messick (1989)mentions all aspects of validity, as to explain that a test should represent the test construct; otherwise it is construct underrepresentation. The purpose of construct validity is to prove the psychological construct, which is the way our brain uses the language and cannot be detected directly (L. F. Bachman & Palmer, 1996). This psychological construct depends on the test task in the assessment. In fact, the construct language ability can determine the level of construct validity (L. F. Bachman & Palmer, 1996). Clearly defining the construct of the test task can lead to the closer interpretation of two domains: the target language use interpretation on that particular task and the general language ability of the test takers.

It is essential for test developers to seek opinions from stakeholders in order to obtain relevant and beneficial consequences of the test (L. F. Bachman & A. S. Palmer, 2012). Many researchers mentioned conducting job analysis in order to acquire an appropriate construct validation. In EAP testing in late 1970s, content validity was a center of attention, while in 1980s, content validity overlaps with construct validity (Weir, 2005). In other words, the test developers, who had insufficient information about language in use, tend to identify the construct validity by utilizing priori. In contrast to Jones who based job

analysis on occupational performance, Davies (as cited in McNamara, 1996, p. 18) referring to EAP context states that '...The best safeguard against an unsatisfactory test is a professional job analysis at the outset.' In addition, Henning (as cited in McNamara, 1996, p. 19) disagrees with the idea of using purely priori and asserts that empirical validation is needed. Obviously, the knowledge of language instructors or test developers alone can appear construct underrepresentation. Obtaining specialist's perspective can include the language in use that test developers may have failed to notice.

Reliability

Reliability was not given equal consideration to validity until 1980s (Weir, 2005). In fact, many researchers recognized reliability as a part of validity. Weir (2005) believed that reliability plays an influential role as a part of validity in which he referred it as scoring validity while Alderson et al. (1995) labeled it as concurrent validation. L. F. Bachman and Palmer (1996); however, refer to reliability as 'consistency of measurement'. It is the ability of the test to be consistent when measuring the test takers' ability. As an illustration, a test with reliability should be able to differentiate the performance of the strong learners from the weak learners. Nevertheless, in a case of mixed-level learners, the test is unreliable if it cannot indicate which learners perform better. In another case, if two forms of the test taken at different situation can justify a learner at the same proficiency level, the tests reflect their reliability.

Types of measurement error

The problem with reliability is to retain the consistency in measurement or to control the measurement error. Types of measurement error or error variance can stem from environment, administration procedures, attributable to examinees, scoring procedures, attributable to the test and test items (J. D. Brown, 1996). The variance enters the list according to the process of employing the assessment.

First, variance due to environment occurred when administering the test, which includes location, space, ventilation, noise, lighting and weather. Following the first variance come the administration procedures, which are directions, equipment, timing, and mechanics of testing. Third, a list of variances attributable to examinees involves health, fatigue, physical characteristics, motivation, emotion, memory, concentration, forgetfulness, impulsiveness, carelessness, testwiseness, comprehension of directions, guessing, task performance speed and chance knowledge of item content.

After the examinees completed the test brings in the next step, scoring. Variance due to scoring procedures is errors in scoring, subjectivity, evaluator biases and evaluator idiosyncracies. The last type of error measurement is variance attributable to the test and test items, which consists of test booklet clarity, answer sheet format, particular sample of items, item types, number of items, item quality and test security. To keep the effect from measurement error at the minimum, Brown (1996) offers a checklist for potential sources of error variance shown in Table 2.

Table 2: Checklist for potential sources of error variance

Checklist for potential sources of error variance
Variance due to environment
location
space
ventilation
noise
lighting
weather
Variance due to administration procedures
directions
equipment
timing

Checklist for potential sources of error variance
mechanics of testing
Variance attributable to examinees
health
fatigue
physical characteristics
motivation
emotion
memory
concentration
forgetfulness
impulsiveness
carelessness
testwiseness
comprehension of directions
guessing
task performance speed
chance knowledge of item content
Variance due to scoring procedures
errors in scoring
subjectivity
evaluator biases
evaluator idiosyncracies
Variance due to the test and test items
test booklet clarity
answer sheet format
particular sample of items
item types
number of items
item quality
test security

Assessing the reliability firstly needs a consideration on the types of test since different strategies are applied. Weir's approach to reliability is classified into 4 ways: test-retest reliability, parallel-form reliability, internal consistency, and marker reliability.

These Weir's 4 ways of assessing reliability are compatible to what Brown (1996) suggests in criterion-referenced tests (CRTs). According to Brown (1996), he categorizes two main types of tests as norm-referenced tests (NRTs) and CRTs. Firstly, norm-referenced test measures the reliability through the degree called a *reliability coefficient*. Three fundamental approaches offered to test developers who wish to tackle error variance are test-retest, equivalent forms and internal-consistency strategies. While NRTs tend to be dependent on variance in score, CRTs do not concern high standard deviation. Four methods applied to achieve the reliability of CRTs are threshold loss agreement, squared-error loss agreement, domain score dependability and confidence intervals. The degree for CRTs reliability measurement is *estimates of test consistency*.

Test-retest reliability

The test-retest reliability is a classic way of assessing reliability by employing one group of students and one test. The value of this reliability ranges from -1 to +1, with 1 as a best possible indicator for strong reliability and consistency. This method; however, is considered 'problematic'(Weir, 2005). As an illustration, the test takers' performance might be better as they remember the test or worse because they are displeased by same test repetition (Alderson et al., 1995).

Parallel-form reliability

Parallel-form reliability is suggested as a solution to the problem of test-retest. To employ this technique, two tests must be developed to be equivalent. The scores from the same group of students taking the two tests are calculated to find the correlation. Two approaches are used

to find an agreement between the two parallel tests: threshold loss agreement approaches and squared-error loss agreement approaches.

Internal consistency

Internal consistency is regarded as the most used technique when comparing to the test-retest reliability and parallel-form reliability (Weir, 2005). Instead of focusing on how to make two tests equivalent, internal consistency or inter-item consistency pays an interest on the internal items' consistency, which can be further investigated in terms of comparison to the population and each individual.

Marker reliability

One of the processes involved in determining the score is the rater or marker. Marker reliability composes of two divisions: intra-rater and inter-rater. When only one marker decides on the score, the sole concern is on intra-rater deals with the consistency within the rater him or herself. Inter-rater would be involved in case of more than one rater to keep the consistency between the raters. Marker reliability can be calculated through correlation where one means a perfect agreement.

No matter what approaches is selected, paying attention to the factors that may affect the result is recommended (J. D. Brown, 1996). The factors to make the test more reliable are keeping the test as long as possible, designing a deliberate well-constructed test, assessing closely identical language material, containing items with high difference indexes in the test, connecting the test to the aims of course the students are in.

Many-Facet Rasch Measurement

Rasch measurement was introduced to language assessment in 1980s, but was not well received by the researchers until 1990s when Many-Facet Rasch measurement was proposed (T. McNamara & Knoch,

2012). The complex version of Rasch known as Many-Facet or Multi-Facet Rasch Measurement (MFRM) was developed by Linacre in 1989. Apart from looking at the test taker's ability, item difficulty and the rating process, MFRM allows the researcher to investigate other factors that can impact the score, such as rater severity (Lynch & McNamara, 1998). In addition, MFRM can also reveal various problems that may occur as in inter-rater problem or intra-rater problem (Lynch & McNamara, 1998). In order to utilize the program, the text file needs to be prepared through either Excel or SPSS program. The MFRM program can be installed only on Windows program.

MFRM can provide insights on facets involved in the test, which are the test takers, raters and criteria (Bond & Fox, 2007; T. McNamara, Knoch, U. & Fan, J., 2019). It also includes information on other statistics such as fixed Chi-square. The information MFRM generates is Wright map, candidate measurement report, rater measurement report, criteria measurement report, and rating scale category functioning. First of all, the Wright map or the variable map presents the overall information of the variables in the analysis by aligning them according to the scale it belongs to.

Second, candidate measurement report offers various views on the candidate through candidate ability, candidate fit, and candidate summary statistics. The ranges of the mean square to consider for the candidate fit can be adapted according to the type of the test. For the clinical observation, the candidate fit is categorized into the mean square range of 0.5 and 1.7 according to the reasonable ranges of Wright & Linacre in Bond and Fox (2007). The mean is set to be 1, the amount less than 1 means less variation while the number more than 1 means more variation. The fit statistics can be observed through the infit and the outfit statistics. The outfit, however, tends to consider the test takers who are in the outliers, which can present a huge amount of difference, while the

infit is the comparison to the group majority. This study is thus focusing on the infit statistics.

Any results of the infit at under 0.5 and over 1.70 will be counted as misfit. Apart from the infit, the mean square is further analyzed as in a normalized distribution, which is set at no more than +2 as more variation or less than -2 as less variation than the expected response.

Similar to the previous item, rater measurement report presents rater severity, rater fit, and rater summary statistics. Next, criteria measurement report displays criterion difficulty, criterion fit, and criterion separation statistics. The ranges of the mean square for the criterion is narrower than the other sections to control the quality according to Eckes (2009) suggested, which ranges from 0.7 to 1.3. Last, rating scale category functioning provides rating scale statistics, which includes the average measures and the Rasch-Andrich threshold. A simple advice for observing average measures is that the average measure of each level should move toward according to the higher level meaning the higher level the higher the average measure not lower while the mean square value is lower than 2. The Rasch-Andrich thresholds is used to look at the step calibration or the step difficulty from one score point to the next. The number should rise more than 1.4 but less than 5.

Authenticity

L. F. Bachman and Palmer (1996) described authenticity as 'the degree of correspondence of the characteristics of a given language test task to the feature of a Target Language Use task'. A test with authenticity means that the situation in the test task can happen in real life (Brown & Abeywickrama, 2010). Authenticity is vital when considering the value of generalizability to the performance in a non-test situation or the general language ability (L. F. Bachman & Palmer, 1996). In fact, the construct language ability can determine the level of authenticity (L. F. Bachman & Palmer, 1996). To establish the authenticity of the test, the important features of the TLU should be specified.

Two aspects to certify a reasonable level of authenticity are the task characteristics and expected perceptions on the part of the test takers and test users (L. F. Bachman & Palmer, 1996), which is similar view to face validity. The degree of authenticity is never viewed as authentic or inauthentic, but rather 'more' or 'less' authentic (L. F. Bachman & Palmer, 1996).

Interactiveness

L. F. Bachman and Palmer (1996)'s definition of interactiveness is 'the extent and type of the test taker's individual characteristics in accomplishing a test task'. The related individual characteristics in responding to the task pertain to language knowledge, strategic competence or metacognitive strategies, topical knowledge and affective schemata. In TLU, the interactiveness is essential as it is an evidence for construct validity (L. F. Bachman & Palmer, 1996). To demonstrate the interactiveness of a test task, the passable level of individual characteristic related to the task use should be specified and conducted during the pre-test through the qualitative method (L. F. Bachman & Palmer, 1996). No definite term of interactiveness and non-interactiveness is used, but rather the term of 'less' or 'more' interactive (L. F. Bachman & Palmer, 1996).

Impact (Consequential validity)

An equivalent term for impact is consequential validity. All the possible effects that can occur in a test are included in consequential validity (H. D. Brown & Abeywickrama, 2010). It is used to examine the effect of the test use and the test itself on individual, society, educational system and other stakeholders (Fulcher & Davidson, 2007). Consequential validity can be observed in two aspects: macro level and micro level (H. D. Brown & Abeywickrama, 2010). The macro level refers to the society, educational system and other stakeholders, while the micro level or washback refers to individual test takers.

Washback

Categorized as a part of the impact, washback is; however, different in a way that it is directly related to the effect stem from the classroom, which can be either positive or negative (H. D. Brown & Abeywickrama, 2010). In a test preparation, a great deal of positive washback can arise from preparing and reviewing the test (H. D. Brown & Abeywickrama, 2010). In addition, feedback given with the scores instead of merely grade can encourage the positive washback as well since the test takers can learn from the teacher's comment on how to improve their performance. In order to facilitate the students' learning, teacher should provide feedback to students to increase positive (H. D. Brown & Abeywickrama, 2010).

No assumption on the impact should be claimed on any tests before a careful investigation is applied (L. F. Bachman & Palmer, 1996). This can include content of teaching, teaching methodology, instruction, test specification, and ways of the assessment. The impact of the test on the individuals can be both the teacher and the test takers. Based on(L. F. Bachman & Palmer, 1996), three facets of the testing procedure that affect the test takers are the experience of taking and preparing for the test, and the decision made about the test takers based on the test scores. L. F. Bachman and Palmer (1996) suggested the following ways to improve the three faces. Firstly, the impact on experience of taking and preparing for test can be directed toward the positive side by allowing some involvement from the test takers in designing and developing the test task. Secondly, providing the verbal feedback personally can be a valuable guidance for the test takers to improve their performance. Lastly, the decision made about the test takers based on the test scores should be equally applied to every individual through the fair test use.

Practicality

Practicality deals with 'the ways in which the test will be implemented, and, to a large degree, whether it will be developed and used at all'(L. F. Bachman & Palmer, 1996) . Since practicality involves the use of all other

element of test usefulness: validity, reliability, authenticity, interactiveness and impact, the association can thus be viewed as in below equation (L. F. Bachman & Palmer, 1996).

$$Practicality = \frac{Available\ resources}{Required\ resources}$$

If practicality $\geq$ 1, the test development and use is practical.

If practicality $\leq$ 1, the test development and use is not practical.

The resources in the equation above can be referred to human resources, material resources, and time. Firstly, human resources include the test developers, raters, test administrators and clerical support. Secondly, material resources are space, equipment and materials. Lastly, time pertains to time used in developing the test until reporting the scores in the first administration and time for specific tasks, such as designing, writing, administering and analyzing.

Relevant Test Usefulness

Considering the classroom performance-based assessment, the qualities of test usefulness relevant to this study are content validity, concurrent validity, construct validity, parallel-form reliability, marker reliability, and wash back (consequential validity). Three different perspectives that can support the internal validity of the test in this study are content validity, concurrent validity, and construct validity. Content validity can attest that the assessment contains a representative of the content intended to measure by using judgment of professionals. Concurrent validity is selected as it is a way to prove external validity of the test while construct validity is important in proving that the test assesses the language feature it claims to do.

Various approaches are suggested to prove reliability of test usefulness; however, not all approach can be applied. Two methods selected in proving reliability of the test are parallel-form reliability and marker reliability. Parallel-

form reliability is applied using squared-error loss agreement because the phi dependability index can be used with the result from only one test administration. Marker reliability will be proven in both intra-rater and inter-rater using Many-Facet Rasch Measurement. Lastly, a part of consequential validity, washback, reflects the feedback from the individuals especially in the classroom assessment. All selected features are essentially fundamental to prove the test usefulness of this classroom performance-based test.

Related assessment

L. F. Bachman and A. S. Palmer (2012) believe that a language assessment should be drafted following the targeted performance considering what way the language is used and in which situation. When developing a test whether for academic or for specific purpose, we may need to consider the availability of the tests in use as well as the purpose of the test. In an attempt of developing a specific purpose test for evaluating pharmacist-patient communication skill of pharmacy Thai students, it is wise to review the existing test aimed for the same and similar purpose to recognize what has been done studying the pros and cons with regards to the current test development. Language for Specific Purposes (LSP) assessment is first reviewed to comprehend the basic idea in developing the test. Apart from the approach for LSP assessment, three health-profession tests involving similar issue are discussed: Objective Structured Clinical Examination (OSCE), Occupational English Test (OET), and Objective Structured Pharmacy Examination (OSPE).

Objective Structured Clinical Examination (OSCE)

OSCE is an exam for health professionals aiming particularly at medical practitioners. It was developed by Harden to replace the traditional paper-pencil exam. This performance-based test on medical skills has been used since 1975. It is used to assess purely the clinical performance and competence according to the medical field of the test taker. The exam is to certify that the test takers attain enough knowledge to work in the medical field in English speaking countries, such as the United Kingdom, United States,

and Canada. Both native and non-native speakers are required to pass this examination to obtain the license to practice.

The Occupational English Test (OET)

OET is an international test of English language for health professionals who wish to work in Australia, New Zealand and Singapore. Composing of four skills, the test was developed in late 1980s by Tim McNamara under the contract of Australian government, which legally allows the assessment to be measured only on language ability and not on professional competence. It was first mainly used in Australia as the preliminary process of screening immigrants to work in the health sector. Non-native speakers of English need to pass OET before they can take the next test, Objective Structured Clinical Examination. In the past, the test use general content in health area to assess the test takers. The current version of OET offers specific content in productive skills, writing and speaking, for twelve professions: dentistry, dietetics, medicine, nursing, occupational therapy, optometry, pharmacy, physiotherapy, podiatry, radiography, speech pathology, and veterinary science, while the reading and listening still employs the general health content.

The focus of this thesis is on the role play, which is in the speaking subtest. Before the current version where real people are used as an interlocutor, OET played the audio to the test takers and let them response through tape recording. This change in test format probably stems from critics(Macqueen et al., 2016; Pill, 2016) about their inauthenticity in testing speaking skill. The one-time audio does not reflect the authenticity of real-life speaking.

Objective Structured Pharmacy Examination (OSPE)

Held by the Pharmacy Council of Thailand, OSPE is an exam that tests clinical performance and competence in pharmaceutical science skills. Similar to OSCE, Thai pharmacist needs to pass OSPE to get the license to practice at the drug store in Thailand. The applicants can apply for the test when they

have earned a degree in Pharmaceutical Science from an institution certified by Thai Pharmacy Council. OSPE is held two times a year, in April and August. It is a criterion-based test. To qualify as a pharmacist and be able to earn the license, the applicants need to earn a passing score of 80 out of 100. Although the students can take this exam while they are in fifth year, this is not an exit exam of pharmaceutical science program.

ESP assessment

Three qualities were raised as characteristics of ESP test: language use according to the context, second, precise specific purpose content, and an interaction between specific purpose language and specific purpose background knowledge (Douglas, 2001). Although separating the language use from the specific content has long been argued (Brunfaut, 2014; Douglas, 2013), we need to be reminded that no clear boundary between LSP and GSP can be specified and that it is the matter of which section along the continuum the focus of the language is in(Brunfaut, 2014). In fact, a performance test can vary from 'very general to very specific'(Douglas, 2001).

Referring to O'Sullivan (2012a), English for Specific Purpose (ESP) or Language for Specific Purpose (LSP) assessment can be categorized in two ways: the field of its use (such as business and law) and the purpose (work, immigration, and study) of the test. This study views the test and rubric development in both views as the field of its use is in pharmacy area and the purpose is for assessing the students' performance after the content is taught in class. One approach that the language use according to the context can be clarified is through Hymes' SPEAKING grid (Douglas, 2001).

In order to subsume that a person can communicate effectively based on communicative competence, they need to recognize both correct and wrong form of language, what, whom, when, where, and how to communicate in a proper way with consideration to speech act and speech event (Paltridge, 2006), both of which belongs to communicative events.

Communicative events consists of three components which is ‘hierarchically’ arranged: speech situation, speech event, and speech act (Cameron, 2012). Speech situation deals with the social context when the speech occurs, which includes physical activities apart from the use of language. Speech events are based solely on the verbal act and many speech events can occur in one speech situation. Speech act, the smallest unit, can be situated in more than one in a speech event. Examples of speech situation, speech event and speech act is described in Table 3.

Table 3: Examples of speech situation, speech event and speech act

Speech situation	Speech event	Speech act
a classroom presentation includes gestures (pointing), writing, nodding, and handing out paper.	presentation, comments, Q&A session	greeting, introducing, asking and answering questions

For the dispensing skills in this study, the speech situation is referred to the drug store. However, the assessment situation does not allow much authenticity of using the real drug store. The classroom is simulated as if the event occurs in the drug store. The dispensing skill is considered a speech event, which contains greeting, asking patients for information, providing information and suggesting drugs appropriate for patient’s symptom. The focal point to consider among the three components is the speech event, which Cameron (2012) adapted from Hymes (1964) and others by selecting the most useful components as a framework called ‘SPEAKING grid’, which stands for setting, participants, ends, act sequence, key, instrumentalities, norms of interaction and genres. Details of each component in the SPEAKING grid (Cameron, 2012) are described in Table 4.

Based on language tests and subjectivity concept of Foucault, T. F. McNamara and Roever (2006) stated that ‘…the test taker is understood as a social being whose subjectivity is a function of subject positions realized in the

test itself.' To be more precise, T. F. McNamara and Roever (2006) believed that 'Certain kinds of 'acceptable' identities are defined by the test and the test is a procedure for conferring those identities-for recognizing the individual in terms of the qualities identified by the test.' Likewise, this study is intended to develop a test of identity. The speaking grid specifies the description in detail to provide the subjectivity of the test. The subjectivity of the task will be described according to the guideline of the speaking grid in order to prove this test of identity for Thai pharmacist students.

Table 4: SPEAKING grid

Component	**Details**
S	setting: where the speech events is located in time and space
P	participants: who takes parts in the speech event, and in what role (e.g. speaker, addressee, audience, eavesdropper)
E	ends: what the purpose of the speech is, and what its outcomes is meant to be
A	act sequence: what speech acts make up the speech event, and what order they are performed in
K	key: the tone or manner of performance (serious or joking, sincere or ironic, etc.)
I	instrumentalities: what channel or medium of communication is used (e.g. speaking, signing, writing, drumming, whistling) and what language/ variety is selected from the participants' repertoire
N	norms of interaction: what the rules are for producing and interpreting speech acts
G	Genres: what 'type' does a speech event belong to, and what other pre-existing conventional forms of speech are drawn on or 'cited' in producing appropriate contributions to talk (e.g. do people quote from mythology or poetry or scripture?)

Research from the fields

The significance that pharmacist has in the society is not only to prescribe drugs but also how he/she can talk to the patient and get the information as to make a decision on which drugs are needed. According to the concept seven-star pharmacist by World Health Organization (2006), communicator is one of the characteristics recommended. It is out of question that communication skill plays a core role in this occupation (Kimberlin, 2006, Graham & Beardsley, 1986). At the meantime, considering the target audience as a part of general communicative needs is a part of learning English for occupational purposes (Kelliny, 1988), which in this case is the patient. Several interests (Sibbald, 1998; Kolsek et al, 2003; Schell & Lind 2003; Ried et al, 2007; Gortney & Lundquist, 2013) have been paid in the education of the pharmacy focusing on the pharmacist's communication skills with the patient in English. These studies can be categorized into three main groups according to the researchers; (1) the specialist in the field of pharmaceutical science, (2) a cooperation research between an English language instructor and a pharmacy with interest on second language learners, (3) language instructors. Based on different focuses of these researchers, various methods have been suggested on how to assess the pharmacist-patient communication, some for the native speakers of English and some for the second language learners.

To understand these performance tests and be able to develop an authentic test with proper construct, we need to recognize the current standard compositions. Similar to Bachman and Palmer's assessment concept (2012), Douglas (2000) believes that basic components of a test essentially for LSP would be 'an analysis of a target use situation' and 'interaction between the language knowledge and specific purpose content knowledge'. Although the concept of assessment construct (Bachman & Palmer, 2012) is stated earlier as the core method of this study, learning from previous practice and test constructs may enhance the understanding of ways in past classroom assessment.

Research from informant specialists

Two studies (Kimberlin, 2006 & Schwartzman et al., 2013) have described the overall view of what content matters to teaching communication to the pharmacists in the US. The former investigated the practice in assessing patient communication skills of colleges and schools of pharmacy in the US and Puerto Rico while the latter described the pharmacy curriculum in the US and Canada. Although the topic of the studies is similar, the objectives of the two studies are slightly different. Kimberlin (2006) is specific about what domains are assessed for pharmacist-patient communication when Schwartzman et al. listed down the content for the communication curriculum, which includes some other issues of pharmacist communication, such as communicating with healthcare providers. In other word, Schwartzman (2013) illustrates the pictures of what is to be taught in the communication curriculum while Kimberlin (2006) steps further into the detail as what skills in the domain, pharmacist-patient communication, is being assessed. The information given; however, are just a percentage of the skills being assessed by the pharmacy programs. It does not include the ratio of the score assigned in the assessment for each category.

In addition, the fact that the assessment's aim is for measuring the native speaker of English could be a reason why no linguistic features are listed. The only domain Kimberlin (2006) includes some linguistic features in is ‘Using effective nonverbal communication’, which is likely to be a flaw, since ‘uses correct language and pronunciation’ is the detail about verbal assessment that does not seem to fit the nonverbal category. As different domains and details of assessment from US schools and colleges are described, it is clearly shown that the way each school performs their assessment is not on the same alignment. Thus, no matter how useful the result from the study could be, it needs to be proceeded with consideration according to different context and needs. The detail these two studies gathered can be used to compare to the dispensing

assessment in Thai as a guideline to what criteria should be taught and assessed. In addition, the difference in context might suggest which criterion is not practical in Thai context.

Table 5: Comparison between content of communication curriculum and skills

Schwartzman, 2013 **Content of Communication Curriculum**	**Kimberlin, 2006** **Skills Identified on Assessment Forms for Pharmacy Student Communication with Patients**
Elements of communication process	Initiating communication Organizing the encounter Concluding the encounter
Psychology of communication	Establishing a trusting relationship
Nonverbal communication	Using effective nonverbal communication
General patient counseling techniques	Eliciting information from the patient Initiating educational interventions Encouraging patient involvement in communication and problem solving Verifying Understanding
Adherence	Promoting adherences to appropriate therapy
Cultural competency	Demonstrating sensitivity to and adjustment of communication based on contextual or cultural factors
Importance of communication in pharmacy practice	
The privileged nature of patient and provider communication/ patient rights	

Focusing on the pharmaceutical content, the next research presents a finding on utilizing performance-based assessment in pharmacy education (Collett et al., 1994). The researchers believe that the written examination cannot reflect the students' real ability in counseling patients. A performance-based assessment for measuring the

students' ability to interview and counsel patients was developed and trialed on 42 sample third-year pharmacy students in England using a role-play method. Before the assessment, a need of preparation was fulfilled for both pharmacy students and the assessors or simulated patients. Apart from attending an eight-week work-based-learning training program, the students also participated in four one-hour tutorials, both of which geared them toward four subject areas used in the assessment. The assessors who acted as simulated patients were given an information pack that contain relevant information about the case, which are medical history, presenting symptoms, social and family history, and current medication of the patient. It also includes a set of initial questions for the assessors to ask the students at the end of the assessment and semi-structured interviews on some students as shown in Table 6 and Table 7.

Table 6: Example of Examiner's answer sheet (topic 2: cough)

Example of Examiner's answer sheet (topic 2: cough)	
1. Who has the cough?	☐
2. How old is he?	☐
3. What sort of cough?	☐
4. Is he coughing up much sputum?	☐
5. What color is it?	☐
6. Has he any other symptoms of a cold?	☐
7. Has he tried anything so far?	☐
8. Is he on any medication?	☐
9. Does he smoke?	☐
Total score	
Examiner's Comments:	

A specific time was assigned to the interview and the students had to complete written questions in the table below based on the interview. The students were required to do this procedure for all four

cases. The answer sheets of both the students and assessors were scored with predetermined marking schedule.

Table 7: Example of a student's answer sheet

Example of a student's answer sheet (topic 4: headache)
1. What relation was the sufferer to the enquirer?
2. What type of headache is the patient suffering from?
3. How long has the patient suffered from a headache?
4. What medication is the patient taking and what condition would these be prescribed for?
5. What other symptoms did the patient have besides a headache?
6. What did you recommend/ advise for this patient?
7. What has the patient tried to relieve the headache?

The data from the study expressed the focus of the researcher on the pharmaceutical content. The questions set for both the students and the examiners were centered on the pharmacy knowledge, which demonstrates the important aspects of the pharmacist-counseling procedure. Neither aspect regarding communication skills nor the language background of the students was mentioned in this study. It can be inferred that all pharmacy students are native speakers of English and were capable of communication skills.

Another study done on verbal communication assessment by a pharmacist presents a fascinating point. Parkhurst (1994) conducted a study on pharmacy students' verbal communication at The Massachusetts College of Pharmacy over two semesters. The participants were a combination of native and non-native speakers of English who were studying their fourth and fifth-year pharmacy courses. Apart from twenty-three hours of classroom observation, the researchers spent sixteen and a half hours observing the participants at four external settings in and around Boston. These were independent retail pharmacies located nearby the doctors' office, national chain pharmacy in a

shopping mall, and one pharmacy in a general hospital and specialized hospital. In addition, interviews were conducted with twenty-five participants to gather in-depth information about reported communication roles, tasks and problems. Parkhurst (1994) analyzed the causes of communication breakdown as the following: inappropriate register (speech style) use, unclear speech production and/or lack of audience awareness, poor listening comprehension, avoidance of communication or repair, inaccurate or missing information, inappropriate/ ambiguous speech act and lack of empathy. She examined the problems as deficiencies leading to communication breakdown, which are deficient knowledge of appropriate register (speech style) use, deficient speech production skills and deficient ability to initiate or repair communication. The study suggested ways to correct these deficiencies. In the study, all pharmacy students regardless of their native language encountered problem in verbal communication. This was due to the fact that communication skills do not deal with only the linguistic features, but also sociolinguistic awareness, such as empathy, and strategies to repair communication. It can be inferred from this study that aside from being a native speaker of English, English verbal communication skills should be taught to all students.

Table 8: Rubric detail of verbal expression

1	2	3	4	5
Communicates in manner that interferes with and/or prevents understanding by patient	Exhibits sufficient control of expression to be understood by an active Listener		Exhibits command of expression (fluency, grammar, vocabulary, tone, volume and modulation of voice, rate of speech, pronunciation)	

When most of the studies that concerns with pharmacist communication emphasize on what skills and knowledge pharmaceutical science students should obtain, some discusses deeper into the evaluation of oral communication skills for pharmacy students. Among

them, Sibbald (1998) offers a rubric with linguistic features as shown in Table 8.

In her research, a global rating of 1-5 is used to determine the verbal expression, nonverbal communication, response to patient's feelings and needs, and degree of coherence. The verbal expression is the section containing details about linguistic assessment, which becomes descriptive when the score is leaning toward the high performance. When such detail about fluency of grammar, vocabulary, tone, volume and modulation of voice, rate of speech, and pronunciation are not mentioned in the low to mid score section, the rubric instead provides the overall description for the rater to interpret the students' performance to. If more than one rater uses the rubric, the problem of inter-rater reliability can occur, since the word 'sufficient' is not clear and more description should be added. Examples of students' performance in each category should be offered to adjust the norm of the raters to what the rubric aims for.

Table 9: Characteristics test for Nonprescription Medication Course

Characteristics	Nonprescription Medication Course
Objectives	Implicit: - To prepare students to assume the role of a pharmacist who will accept accountability for patient care outcomes by identifying preventing and resolving drug related problems - To assess a pharmacist's cumulative content knowledge used to analyze, synthesize, and evaluate information
Procedures for responding	Oral interaction to assist simulated patient with special needs
Structure **Number of tasks**	Two patient interviews
Time allotment	45-minute preparation session, 10-minute oral examination, 5-minute interval for documentation

Collaboration research

After learning about the overview of how the pharmacist-communication assessment is done by strictly pharmacy specialist, it is the turn of collaboration assessment between a professional and a language instructor. A study on developing a communications course for pharmaceutical students who were using English as their second language was a collaboration work between an English instructor and a pharmacist whose specialty is in communication for pharmacist (Graham and Beardsley, 1986). The researchers designed the course with a focus on content to improve the participants' oral communication since the participants' problem lies on the inability to speak professionally. The study relies on the pretest and posttest score of the participants as independent variables to indicate the change in participants' proficiency after being taught. Apart from considering previous studies, the researchers conducted a needs analysis on the learners, the teaching establishment and user institution to determine what content to be included. Various activities were listed including the dictation, listening comprehension, the writing assignment and the role-play, but no sample of the rubric to indicate how the role play was measured. The closest information relevant to the researchers' judgment is a table showing mean and standard deviation scores of pretest and posttest obtained from SPEAK test in the category of pronunciation, grammar, fluency and overall comprehensibility. With the information from the study, it can be concluded that this research collaborated the content of the course on the pharmacy knowledge and some linguistic features, while the main criteria for the assessment holds the linguistic areas as a core.

Due to the lack of validated pharmacy assessment criteria, a collaboration research (Hyvärinen, et al., 2012) between speech communication instructors and pharmacy teachers are conducted in believing that specific profession criteria for assessing communication skills of pharmacy students used to counsel with patient needs to be

developed. The study employed 223 subjects, 4 teachers and 119 students, from a course in communication skills and patient counseling in a pharmacy at a university in Finland. During the three-month course, students were given practical training on the use of the criteria in groups as peer evaluation and self-evaluation. In order to test the reliability of the analytic criteria, the study employed the quantitative data of the role-play final-test score rated by speech communication and pharmacy teachers and students' peer and self-evaluation, which also include open-ended questions to generate student's level of pharmacy knowledge for rating such score. The criteria developed for this course is excessively based on assessing communication skills. As shown in table below, the analytic criteria is divided into five topics; 1) controlling the situation, 2) customer-oriented behavior, 3) giving instruction based on individual needs, 4) motivation and attitudes, 5) reflective action. The analytic rubric does not list the linguistic features or pharmacy knowledge, which may due to the specialty of the respondent researcher, whose topic of interest is in the speech communication. Although well developed and validated, this criterion does not match the author's purpose. The rubric cannot be further used by the present author, who is interested in developing a rubric that would cover the rating scale for linguistic features and pharmacy content.

Research from language instructors

Based on the issue whether the specific knowledge can be excluded from communicative competence when constructing the ESP tests, Pill and McNamara (2015) conducted a study on establishing the health professionals' point of view on minimum requirement of English communication skills. One aim of this research is to define the experts' judgment in assessing clinical communicative competence to strengthen the standard setting process of the specific-purpose test, which in this research is Occupational English Test (OET). The researchers compared the information and scores collected from the speaking sub-test of OET,

which is normally assessed by test raters with language training, with the score generated from the qualified panelists with health professional training. Pill and McNamara (2015) deployed the analytical judgment method in setting the standard between the two groups, language teachers and health professionals. While the score of the prior group can be obtained swiftly from the database, the latter group Akin to the scores from the OET database of the test takers, the participants are thirty-nine educators and practitioners from three health professionals; medicine, nursing and physiotherapy. The participants were asked to listen to the audio tape of test taker's communication with the simulated patient and rated them based on the new four categories; strong, competent, not yet competent, and unsatisfactory, without knowing which four existing categories; A, B, C and D was previously assigned to them. Grade A and B, referred as strong and competent, are considered as pass while C and D, referred as not yet competent and unsatisfactory, are fail.

The results of each profession's rating do not only differ from the scores of regular OET raters, but also among the three areas of profession. It is found that the cut scores among the three professions are different to a slight extent. In general, the health specialists identified more test takers as grade A more than OET raters moving roughly twenty percent of test takers from B to A. In medicine, seventeen percent more test takers received grade A while fewer passed the test at ten percent. The nursing overall pass rate is not as high as medicine. From the grade given, the finding; however, indicates that the proposed cut scores for nursing should be lower than the current one. This leads to the suggestion by Pill and McNamara (2015) that OET should adopt different cut scores if they want to fit the requirement of each profession.

The cut scores are important to language teachers as it can explicate the minimum requirement of test taker's abilities for the intended job from the specialists' point of view. It is another means to validate the exam as authentic and whether the exam is a good

mechanic for interpreting students' abilities with the value from their professionals attached. Although OET has been qualified to use as a high-stake test to select people to work in health professional fields, the standard for passing does not exactly match what the study from Pill and McNamara (2015) found. Each profession has their own standard in passing the exam. Deploying general health context to test the students can ease the testing procedure in terms of developing and rating. When the result showed otherwise the standard each profession requires, a language instructor as a rater should consider the differences and adjust the grade classification accordingly as to reflect the authenticity in terms of professional judgment.

Through qualitative research, Woodward-Kron and Elder (2015) investigate the authenticity of the Occupational English Test (OET) by comparing the role-plays of the test to the similar part in Objective Structured Clinical Examination (OSCE), which is a clinical medical exam used to qualify the doctors to work in English speaking countries. Considering the difficulties in obtaining information from actual setting and the range of cases, OSCE is regarded as a viable representation of real-life interaction. The participants are twelve international medical graduates who were preparing for a test and work in Australia. Randomly selected twelve successful performances, the researchers compared the similarity of the two tests by exploring the discourse structure and the management of communication tasks based on analytical approach. The finding reveals that in order to reach their aims in communicating, the OSCE and OET candidates deploy similar genre structure and genre features. Six phases of the genre structure and genre features are categorized as opening, statement of concern, exploration of condition, discussion of diagnosis, treatment and investigation, and closing. Apart from defining the turn takers and the initiator, the study also distinguishes the nature of language use by relating it to the purpose of each phase.

As the similarities between the two tests have shown, it is implied that if OSCE is claimed to be authentic, so does the OET. However, one unique difference between the two tests that cannot be ignored is the interactional patterns and lexicogrammatical choices. The pattern of language use of OET is found to be formal, while OSCE is rather informal. This clash stems directly from prompt materials provided to simulated patients since OET test use abstract language while OSCE use colloquial language. The language pattern is a non-negligible issue in constructing a prompt since it can reflect the authenticity of the test toward the real-life language use. When comparing the two tests, OSCE might actually depict the authenticity of the referred situation better than OET as not many patients know medical and formal terminology and would tend to use simple language in communicating. Considering authenticity, the pattern of linguistic choices can be one of the occupationally relevant communication skills or another approach to prove the reliability of the ESP test.

Viewing English for Specific Purposes test as a boundary object, Macqueen, Pill and Knoch (2015) explored the OET test in view of stakeholders' perception, language representation in the test and the test's effects. They interviewed three stakeholder groups, which are senior representatives of two professional bodies, supervisors of English as an Additional Language health professionals and successful OET non-native-speaker of English candidates who are currently employed, about the test relevance and the language demand in their workplace. Based on the semi-structured interview, this qualitative research generated four themes: communication skills versus language proficiency, professional journeys, note-taking and role-play.

The researcher answered the first question which deals with the stakeholders' perception to the test. Some stakeholders view aspects of communication skills and language proficiency as separate constructs in the test, while the test-takers believe they 'overlap'. As an illustration,

several successful test-takers use communication strategies to deal with language proficiency test. The researchers state that the fact that some test-takers merge the strategies could be a result of ‘a degree of incommensurability between the applied linguistics and health professional notions of communication’. The second question involving the adequate representation of language in the workplace is answered by comparing OET to International English Language Testing System (IELTS), which is another test acceptable for healthcare professional application in Australia. It is found that the construct of IELTS and OET tests are coherent in terms of screening healthcare workforce. Nevertheless, some stakeholders are under the impression that OET is more relevant to the test takers than IELTS.

A few concerns were raised considering the authenticity of the test comparing to the real-world task. Note-taking in the listening part of the test is another issue for authenticity. The non-genuine features found in the test process are the lack of being selective when note taking, no visuals provided, the surreal subdued test room, and the fact that this listening only is a one-way communication that does not allow the test taker to verify the information received. Other aspects that participants believe should have exited in the OET speaking test are intra-/ inter-professional communication and the variety of accents in the real-world experience. Nonetheless, the task was viewed by the test takers as an adequately portrayal of the practice for the real-world situation. In addition, some past candidates perceived the OET test preparation as a positive wash back since it offered some degree of experience in Australian culture, thus preparing them for the workplace context.

Although Macqueen et al. (2015) found that OET test is adequately genuine representation of the real-world communication, increasing some possible features would improve the test’s authenticity greatly by using the real person to role-play the case. This can give the test-takers the experience as if they are talking to the real patient as they

will be provided with visuals and ability to communicate and verify the information received with the real person, not the audio tape. Instead of testing the listening alone, integrated task should be adopted in the assessment for health-related professionals. Another important feature of pharmacy assessment reported in Macqueen et al.'s study (2015) is the transition of the register use, inter-/intra-professional interaction. The fact that many health professionals may not know adequate layman terms to effectively communicate with patients promoted the awareness of assessing the test takers on professional-patient communication. In Thai context, the dispensing process at the hospital does not require pharmacists to do much communication with patients other than confirming the identity and explaining the drug use. This current research would focus solely on inter-professional interaction at the drug store, excluding the hospital, situation since our scope is the dispensing process.

Based on a previous study (Douglas & Myers, 2000) involving assessing communication-skill performance of veterinary students, it is found that the linguists tend to rely on the language and communication aspects while the experts use their knowledge and professionalism as their baseline. For language teachers in the English for Specific field, finding what matters in the perspectives of language educators and experts on judging one performance is the ideal solution for indigenous criteria, which needs amount of research to support this area. One of such study is conducted by Pill (2015) to investigate indigenous criteria aiming for more authentic health-professional assessment. The questions of his research explore the long-unresolved-issue aspects of performance that matter to health professionals and applicable to language teachers in terms of assessing. The study employs the data from past OET test takers who are non-native speakers of English intended to work in Australia. The assessment involves two parts; video recording and written report. The focus of the current study is on the oral performance. The

video records are assessed by health professionals and language teachers, whose language backgrounds are presumed as native speakers of English. The participants are directed by the general criteria in assessing oral performance as “stronger/weaker aspects of performance” with no other scopes suggested in order to draw various comments and not guide the participants to considering language or communication aspects. Using thematic analysis to examine the comments, the researcher offers three headings: the performance, the goals and the foundation of the consultation as shown in Figure 2.

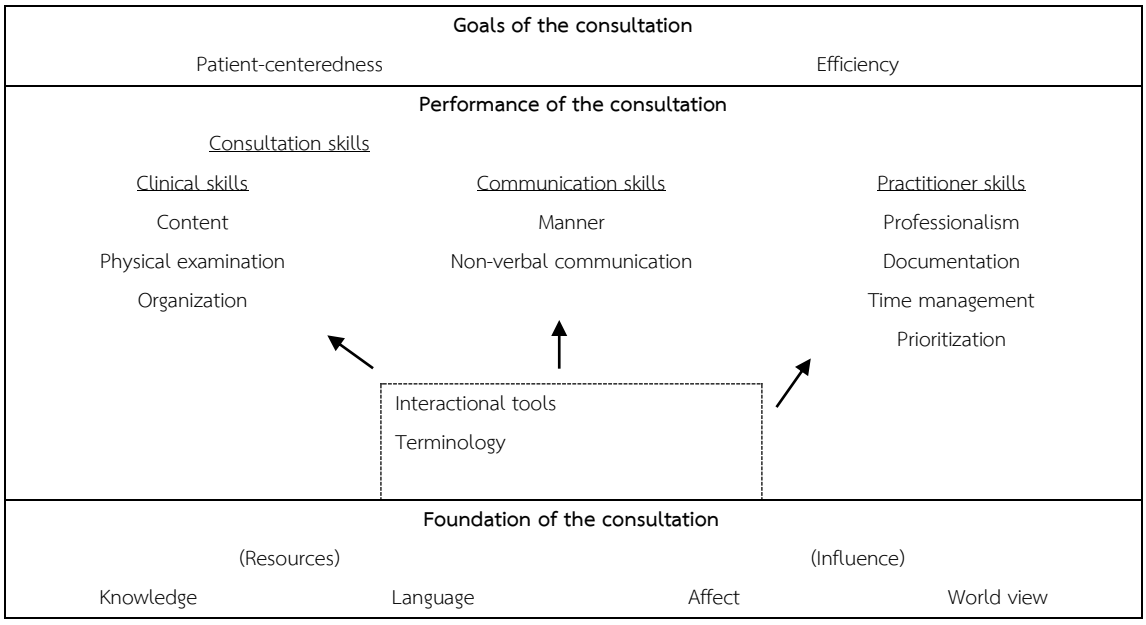


Figure 2: Three Headings of Using Thematic Analysis to Examine the Comments

In an attempt to drawing on indigenous criteria without bias, this study provides a great insight into the health professionals’ judgment by avoiding guidance to language or communication skills. Nonetheless, not much information from the participants is documented. It would be great to consider the participants’ language background as a variable to consider whether it affects their decision in anyway. This study also points out the importance of non-verbal communication. The OET changed the format of speaking sub-test from audio only to real

interlocutor, which somehow reflects the importance of non-verbal communication. The consideration to include non-verbal communication in the assessment would depend on the test amenities; however, if we want to adjust the criteria following the professionals' judgment, administrative staff might as well provide well-equipped facilities to the test takers as the OET adapted its format to suit the authenticity of the communicative performance assessment.

Revealing different result from the previous research (Douglas & Myers, 2000) about the focus of the experts on knowledge and professionalism, the professionals in this research note paraphrasing and signposting techniques as one of the criteria they consider in assessing performance in communicating with patients. Including this linguistic feature, which stems directly from the professionals' point of view, in the current criteria can fill in the gap between linguistic and professional criteria. Applying this criterion in the rubric; however, needs to be prudent. Paraphrasing technique is an advanced skill that can differentiate the students' proficiency level. The detail on the quality of paraphrasing should be stated and ascertained the accuracy by comparing to the well-established rubric system, such as TOEFL.

Another topic to be concerned for assessment is cultural issue. The researcher categorizes culture under a topic called world view. From this study, failing to acknowledge the cultural issue can result in incompetence in both clinical and communication aspects. Without further information from the test taker about their cultural background, the raters have to be sensitive on assessing such performance.

Table 10: A Summary of assessment content and students' language and professional background in the studies

Researcher(s), year	Assessment content			Test takers/ participants/ students' language background			
	Pharm/ Med	Ling	Com	NS	NNS	H-Pro	P-Pro
Collett, J.H., Rees, J.A., Mylrea, S. & Crowther, I., 1994	√			√			√
Graham, J.C. & Beardsley, R. S., 1986		√			√		√
Hyvärinen, M.-L., Tanskanen, P., Katajavuori, N., & Isotalus, P., 2012			√		√		√
Kimberlin, C.L., 2006	√			√	√		√
Macqueen, S., Pill, J. & Knoch U., 2015		√	√	√	√	√	
Schwartzman et al., 2013	√			√			√
Sibbald, D., 1998	√	√		√			√
Parkhurst, C., 1994	√			√	√		√
Pill, J. & McNamara, T., 2015		√			√	√	
Pill, J., 2015	√	√		√	√	√	
Woodward-Kron, R. & Elder, C., 2015	√	√			√	√	

Note: Pharm/Med means Pharmacy/ Medical Ling means Linguistics
Com means Communication skills
NS means Native Speakers of English
NNS means Non-Native Speakers of English
H-Pro means Health professionals related
P-Pro means Pharmacy professionals related

Construct of test task and the rubric

Defining a construct can offer a crucial interpretation. In fact, test developer should clarify the construct starting at the beginning to obtain a meaningful result and validation(Weir, 2005). 'The construct definition for a particular assessment situation becomes the basis for the kind of interpretations we can make from the assessment practice'(L. F. Bachman & A. S. Palmer, 2012) To properly assess the

student's performance, test developers should state a comprehensible detail on the language use and language ability as that can be a proof of validity of the test.

Apart from making a critical decision on 'the nature of the knowledge, skills or abilities that should be assessed', it is necessary to choose the most suitable item and task formats in order to obtain the evidence of the test taker's language performance (Green, 2014). Defining the types of language use in the task would draw on such result. According to L. Bachman and A. Palmer (2012), language use is categorized into two types considering the way of interaction: internally interactive and externally interactive. Both of these interactions of the language use involve the language ability, personal attributes, topical knowledge, affective schemata and cognitive strategies. L. F. Bachman and A. S. Palmer (2012) defined the term internally interactive as 'interactions among attributes within individual language users' and externally interactive as 'interacting with characteristics in the language use situation'. The difference between internally and externally interaction is the involvement of the outside factors in each situation of the language use, such as written or spoken language.

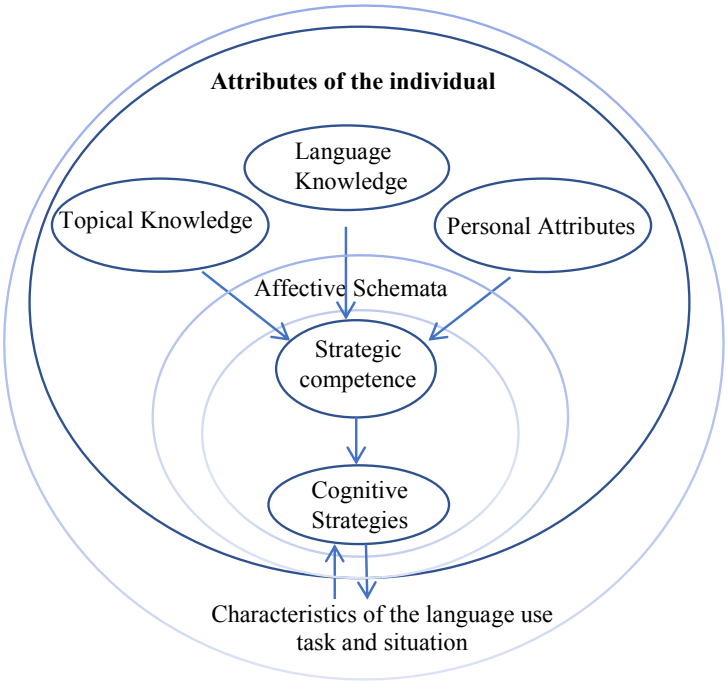


Figure 3: Non-reciprocal language use

Another two important terms that L. Bachman and A. Palmer (2012) and Weir (2005) mentioned are reciprocal and non-reciprocal language use. When the interaction of language use occurs 'directly' with more than one language user, it is called reciprocal language use, while a single language user with no other language users is called non-reciprocal language use. L. F. Bachman and A. S. Palmer (2012) illustrated the interaction among these attributes in the non-reciprocal language figure 2.

Definition of attributes of individuals

Topical knowledge

Topical knowledge is known as content knowledge, real world knowledge or 'knowledge structure in long-term memory'. This attribute is necessary to communicator since the data is a base of the language use. In some assessment, the topical knowledge can have an influence on the test taker's performance and the score of that performance.

Personal attributes

Personal attributes refer to 'a wide range of personal test taker attributes...relevant to the decisions we make about assessment design and development'(L. F. Bachman & A. S. Palmer, 2012). It includes age, sex, nationality, resident status, length of residence, native language, foreign language aptitude, level and type of general education, socio-psychological factors, personality, cognitive style, language use strategies, ethnolinguistic factors, multilingual ability and type and amount of preparation or prior experience with a given assessment (L. F. Bachman & A. S. Palmer, 2012).

Affective schemata

Affective schemata reflect how a person feels and view a particular task and topic based on the person's experience. The affective response on some topics or tasks can generate emotional response,

which affects the person's capability to employ language knowledge and strategic competence. When designing assessment tasks and topics, the type of task should be considered in terms of affective schemata as it can lead to the intended positive and undesired negative affective responses.

Table 11: Areas of language knowledge

Areas of language knowledge
1. Organizational Knowledge
a. Grammatical Knowledge
i. Knowledge of vocabulary
ii. Knowledge of syntax
iii. Knowledge of phonology/ graphology
b. Textual Knowledge
i. Knowledge of cohesion
ii. Knowledge of rhetorical or conversational organization
2. Pragmatic Knowledge
a. Functional Knowledge
i. Knowledge of ideational functions
ii. Knowledge of manipulative functions
iii. Knowledge of heuristic functions
iv. Knowledge of imaginative functions
b. Sociolinguistic Knowledge
i. Knowledge of genres
ii. Knowledge of dialects/ varieties
iii. Knowledge of registers
iv. Knowledge of natural or idiomatic expressions
v. Knowledge of cultural references and figures of speech

Cognitive strategies

Cognitive strategies allow the language users to implement plans while communicating with another people, understanding and responding according to the discourse of that situation.

Language knowledge

Language knowledge is as a matter of fact a part of language ability, which includes strategic competence. Viewing this attribute would need to consider concept of language ability. The framework for describing language use is originally based on L. F. Bachman (1990)'s concept on 'the ability to use language communicatively'. Language knowledge is divided into two main areas: organizational knowledge and pragmatic knowledge. Details of these two main areas are illustrated in Table 11.

Organizational knowledge deals with the basic foundation when forming a language. It contains two parts: grammatical knowledge and textual knowledge. Grammatical knowledge allows the language user to use vocabulary, syntax and phonology or graphology knowledge to form a sentence or an utterance. Once those utterances are grouped to create an organized text, the textual knowledge proceeds. The textual knowledge includes knowledge of cohesion, which connects sentences with connecting words, and knowledge of rhetorical or conversational organization, which enable the language user to organize the text, such as comparison-contrast texts and turn taking in conversation.

When the language is produced, the real usage needs the user to contemplate the pragmatic knowledge, which involves functional knowledge and sociolinguistic knowledge. Functional knowledge copes with 'the relationship between utterances or sentences and texts and the intentions of language users'. Knowledge of ideational functions accommodates the language user in informing, expressing, and exchanging 'information about ideas, knowledge, or feelings'. If language users desire to make some effect on the surrounding, the knowledge of manipulative functions can be applied. It is consisted of instrumental functions, which are utilized when language users want people to do something, regulatory functions, which deal with directing people in rules

and laws, and interpersonal functions, which are useful for initiating, keeping, and modifying interpersonal relationships.

Human is able to clarify language pertinent to specific language use situation with sociolinguistic knowledge, which comprises knowledge of genres, knowledge of dialects or varieties, knowledge of register, knowledge of natural or idiomatic expressions and, knowledge of cultural references and figures of speech.

Strategic competence

Lastly, strategic competence, initially a part of language knowledge, is recognized as 'a set of metacognitive strategies' that deals with individual's problem-solving management through the use of various attributes and characteristics of the language use situation. Strategic competence is composed of three steps: goal setting, appraising and planning. The detail on these areas on metacognitive strategy use is described in Table 12 (L. F. Bachman & A. S. Palmer, 2012).

Table 12: Areas of metacognitive strategy use

Areas of metacognitive strategy use	
Goal setting **(deciding what one is going to do)**	Identifying the language use and assessment tasks to be attempted Choosing one or more tasks from a set of possible tasks (sometimes by default, if only one task is understandable) Deciding whether or not to attempt to complete the task(s) selected.
Appraising **(taking stock of what is needed, what one has to work with, and how well one has done)**	Appraising the characteristics of the language use or assessment task to determine the desirability and feasibility of successfully completing it and what resources are needed to complete it.

Areas of metacognitive strategy use	
	Appraising our own knowledge (topical, language) components to see if relevant areas of knowledge are available for successfully completing the language use or assessment task Appraising the degree to which the language use or assessment task has been successfully completed.
Planning **(deciding how to use what one has)**	Selecting elements from the areas of topical knowledge and language knowledge for successfully completing the assessment task Formulating one or more plans for implementing these elements in a response to the assessment task Selecting one plan for initial implementation as a response to the assessment task.

Internally interactive reciprocal language use

The aim of this assessment research is developing tools for assessing pharmacy students' communication skills while dispensing drugs to patients at the drugstore. Such communication is regarded as reciprocal language use considering the number of interlocutors, which in this case is two or more as in a pharmacy student and a patient.

The conversation of the pharmacist interacts with the patient is internally interactive considering pharmacy students as a focal point. As an illustration, the conversation happens at the drugstore where the patient comes in and consults the pharmacist about the condition he or she experiences. The pharmacist asks the patient questions that can help identify

the disease. The pharmacist then suggests the medication for the possible disease. When the patient accepted the recommended medication, the pharmacist can dispense the drug explain the usage and provide the caution of the drug use. Both interlocutors exchange information between each other throughout the whole conversation. This dispensing conversation is internal interactive for both interlocutors. The patient may need to order the event of the symptom before explaining it to the pharmacist, who after listening to the information uses topical knowledge on pharmaceutical science to gather more information in questioning, identifying the possible disease of the patient and recommending the medication. Lastly, after the patient makes a decision, the pharmacist uses his or her personal attributes in deciding the suggestion for the patients, including explaining the drug usage and drug caution. Thus, based on the goal of this study, the general language-use framework of this assessment is internally interactive reciprocal language use. To attain a construct to serve such assessment in context, a social interaction needs to be investigated (Chalhoub-Deville, 2003), the detail which will be listed in a SPEAKING grid.

Relevant Construct

Following L. F. Bachman and A. S. Palmer (2012) conceptual framework of language use, all categories of individual attributes need to be contemplated regarding the language use task and situation in order to achieve the closest interpretation of the test takers' underlying language competence. Apart from the Assessment Use Argument (AUA), the construct relevant to this study involves topical knowledge, personal attributes of the test taker, grammatical knowledge, knowledge of genres, and strategic competence.

First of all, topical knowledge is necessary to the assessment since it can result in the difference in the test taker's performance(L. F. Bachman & A. S. Palmer, 2012). In this assessment, the topical knowledge of pharmacy students concerns mainly the pharmaceutical science knowledge. As suggested by T. McNamara (1996) and Douglas (2000), implementing the topical knowledge that involves knowledge in the specific field needs cooperation

from the specialist informant. This means that consulting with specialist informants would occur throughout the process of material development to ensure the test construct validity (More discussion on the process of consulting with the specialist informant can be found in Chapter three.).

The second factor is personal attributes of the test taker. This classroom assessment assumes that the age, level of general education, native language and prior relevant experience of the test takers are the same. Another observation of this research assessment is on the prior experience on given assessment. Lim (2013) found that participants on the rubric-referenced oral production do not use the rubric during the task performance since they have to focus on constantly producing the language. The participants; however, can use the rubric to review their performance and improve themselves based on the criteria listed (Lim, 2013). Apart from explaining the rubric to students in detail, this study plans to give equal number of exercises to pharmacy students. This is to ascertain that the students are given the same treatment in terms of prior experience on the assessment.

In this study, the relevant construct as a part of language use according to the context is the grammatical knowledge in organizational knowledge and sociolinguistic knowledge in pragmatic knowledge paying specific attention at knowledge of genres. Grammatical knowledge can inform the ability of the pharmacist to produce sentences or utterances. This can be a clear indicator to differentiate the level of the test takers. Another necessary composition to be included is knowledge of genres. The task is developed with the concern over the particular pattern of the language used in Thai pharmacists' community, which in this case are drug store situations. Finally, strategic competence is vital to be included in the construct since it deals with individual's problem-solving management through the use of various attributes and characteristics of the language use situation. In dispensing case, goal setting, appraising and planning are all relevant in communicating with patients. Thus, these details will be listed in SPEAKING GRID for drug dispensing as a relevant construct of the test task.

Assessment Use Argument (AUA)

In order to justify the test usefulness and the validate the test task and rubric, a framework of argument to validate the test is recommended and widely used by the researchers to support the development of their assessment or to prove the existing assessment. Frameworks to validate the test has been suggested by various researchers with the most widely applied from L. Bachman and A. Palmer (2012) as AUA and Kane (2006) as argument-based validation. This study employed AUA framework of L. Bachman and Damböck (2018), which focuses on classroom assessment.

AUA is a conceptual framework to justify the test task and the rubric and support test usefulness. Without AUA, test usefulness can appear vague and unreliable. AUA consists of four claims: the intended consequences of using the assessment, the intended decisions to be made, the intended interpretations, and the intended assessment records. Each claim composes of the intended outcome and one or more of its qualities. In addition to that, backing can be provided to strongly support the AUA. Backing can be divided into two types: backing from procedures that are followed during the development and use of the assessment, and backing that is collected during assessment development and use specifically for the purpose of supporting claim in the AUA.

Claim 1 or the intended consequences of using the assessment considers the beneficial effect on whoever involves and the quality of the assessment. Three main types of claim for classroom-based assessment can be stated as followed: improving instruction and learning, assuring that students are ready for instruction at the appropriate level, and assuring that students who are certified at a given level of ability have actually achieved that level of ability. Apart from stating the claim, washback or impact on instruction and learning should be provided as warrants to the first claim.

Claim 2 is the intended decisions to be made. For this claim, the decisions to be made are stated as to what it is, who it is for, by whom, when

it is made, and what qualities are. The backing for the second claim aims mainly at the stakeholders by assuring that the decisions are values-sensitive and equitable, both of which can be considered in high-stake summative decisions and low-stakes formative decisions. Examples of values-sensitivity are notes from meetings with different stakeholders and documentation on relevant rules and regulations, while the examples of equitability are documentation for procedures like setting standards and scores, monitoring how assessments are implemented, and informing students and other stakeholders.

Claim 3, which is the intended interpretations states what should be assessed and how it should be assessed. What should be assess concerns how to identify information about student's ability in terms of relevance and sufficiency, how to define the area of language ability, and the source of definition in a meaningful way. How it should be assessed deals with the type of performance and the kinds of tasks in terms of generalizability and impartiality. The backing for the third claim involves the following qualities: relevance and sufficiency, meaningfulness, generalizability, and impartiality.

Claim 4 is the intended assessment records, which deals with the kind of assessment records and its intended quality. The kinds of assessment record involve appropriate assessment performance, assessment record, and scores. The quality of the assessment record directs the intention to the consistency, which can be proved by identifying the specific possible sources of inconsistency and ways to minimize the effects. Providing backing for the fourth claim can cover consistency in various aspects as in administration, assessment tasks, criteria and procedures, and raters.

Chapter 3

Methodology

Research methodology

Presenting in this chapter are participants, research instruments, data collection and data analysis. The detail of each process is described in Table 13 as follows.

Table 13: Overview of the research design

Research Objectives	Type of Data & Instruments	Participants	Instruments
1. To create and validate tasks for measuring English oral communication competency of Thai pharmaceutical science students	Qualitative (Semi-structured interviews)	Pharmaceutical science students and informant specialists	Content analysis
	Qualitative (Consultation)	Informant specialists	Content analysis
2. To create and validate the rubric for measuring English oral communication competency of Thai pharmaceutical science students	Quantitative (Questionnaire)	Pharmaceutical science students and pharmacists	Mean, standard deviation, ANOVA and Games Howell post hoc test
	Qualitative (Consultation)	Informant specialists	Content analysis
3. To establish the extent of this dispensing task's test-usefulness	Quantitative (Scores)	Raters (language instructors) and pharmaceutical science students	Many-Facet Rasch Measurement
	Quantitative (Questionnaire)	Pharmaceutical science students	Content analysis
	Qualitative (Semi-structured interviews)	Raters (language instructors)	Content analysis

In order to fulfill the first research question, the research design started with the semi-structured interviews with informant specialists and pharmaceutical science students to obtain general information about their content knowledge, the Thai dispensing exam both used in the classroom and the license exam, the students experience in dispensing drugs in real situations. These data were accumulated to develop a dispensing SPEAKING grid, which is used as a guideline in task development and test administration. The tasks were developed mainly based on the information from the informant specialists.

Next, the secondary research and the Thai dispensing rubric with the information from the interview were utilized to develop a questionnaire to determine the criteria to be included in the rubric. The questionnaire was given to pharmaceutical science students and pharmacist experts, who were responsible for training the students during their internships at the drug stores. The result from the questionnaire was analyzed to find the mean, standard deviation, ANOVA and Games Howell post hoc test in order to eliminate the criteria. The consultation hours were conducted with informant specialists to finalize the criteria and the scores assigned.

The last research procedure was to establish the extent of this dispensing task's test-usefulness which was first through Many-Facet Rasch Measurement (MFRM). The scores on students' dispensing performance from the raters were analyzed to find the average fair score of three facets, which were students, raters and criteria. Second, the students were asked to fill in the questionnaire, which allowed them to comment on the test. Lastly, the interviews with the raters were conducted to investigate the ease of rubric usage and the rating process. To support the test usefulness, the Assessment Use Arguments were developed based on the result from the MFRM, the students' comments from questionnaires on the test and the interview with the raters.

Participants

The participants of this classroom-based assessment consisted of four groups: informant specialists, fifth-year pharmacy undergraduate students, language instructors as raters, and pharmacists.

Informant specialists

Two informant specialists are involved in the interview and consultation hours on developing the task. Both specialists are qualified to provide information and recommendation on the test task. One of the specialists teaches the dispensing course to fourth-year pharmaceutical students while the other one supervises students when they practice their internship at the drugstores. Both of them have been teaching pharmacy students for over five years and are very familiar with the official dispensing test that the students need to pass in order to obtain the pharmaceutical license in Thailand.

Fifth-year pharmacy undergraduate students

The first group of participants is fifth-year pharmacy undergraduate students at a university in Bangkok, Thailand. Each year, around a hundred and forty students attend the targeted pharmacy school. These students took three courses of English, which are two fundamental English courses and one course of English for pharmaceutical profession. Their English proficiency varied from B1 to C1. Before entering the university, students chose their major during the entrance exam. They have two fields of pharmacy as their option; pharmaceutical science, which is for those who want to work in the drug industry, and pharmaceutical care, which is for those who wish to work in the drug store. The entrance scores to these two departments are not much different.

Language instructors as raters

Originally, the raters of the study followed the number of the language instructors for five classes. Among six raters employed in this study, only one rater did not teach the class. An extra rater was added in to help when the difference in score is more than seven percent. The extra rater has been trained for the rating process for dispensing as well

as language instructors of the class. All the raters were given the detail of the test task to study before gathering in for two one-hour training sessions. Four raters have experience teaching this class before.

Pharmacists

In validating content section, this classroom performance-based assessment employs nonprobability sampling or purposive sampling in selecting pharmacists as experts. According to Henry (2008), one of the aims in using non-probability sampling is 'obtain evidence about individuals whose experiences are particularly relevant to the study's research questions'. Purposive sampling is utilized when 'particular settings, persons, or events are deliberately selected for important information' (Maxwell, 2008). In the meantime, purposive sampling is defined by Layder (2012) as problem sampling. As the issue of representativeness of the whole population may exist, Layder (2012) argued that the pattern of the sampling is directed by the problem of the research and what matters is not the size but the quality of the sample.

The concern of this study is the quality of the participants whose data are appraised as great contribution to the task development. 'As long as the participants are representative of the group and its culture..., then there is no fixed criterion for the number of participants.' (Flowerdew, 2002). Apart from the fact that this study selected a specific group of informants, the number of participants this study employed adapted from standard setting (Cizek & Bunch, 2007) and Angoff approach is believed to yield substantial data for rubric development. Below is the explanation supported with a research aiming at closely similar objectives.

Selecting participants, who are knowledgeable in the task assessment and can provide sensible perception toward the assessment standard, is extremely necessary in qualifying an assessment (Cizek & Bunch, 2007). The qualification of being a pharmacist working in the field

is the first aspect in selecting the participants. Such qualification means that the participants passed the Thai version of pharmaceutical licensure exam and are legally practicing as a pharmacist. In addition, the type of assessment can be a judgment to select the participants (Cizek & Bunch, 2007). This performance-based assessment is intended to test pharmacy students for their dispensing-skill performance. Although the test follows the construct of the Thai pharmaceutical professional licensure and affects part of the grade, it is not considered a high-stake test. Apart from the participants' requirement in holding a pharmaceutical license to practice in Thailand, the participants representing this group of populations are required to work mostly at the drug stores not at the hospital. Since the detail of pharmacist's work between the hospital and the drug store is different. Dispensing drugs at the hospital does not require the pharmacist to ask much detail of the patient's health record or to make a decision on what medication should be dispensed. Thus, two important qualifications for the participants in this study are holding a pharmaceutical licensure to practice in Thailand and working at the drug store.

Although the number of more than 30 representatives is recommended for behavioral research, a small sample of 10-20 is acceptable in case of controlled experimental research (Gray, 2013). An example of research aiming at similar objectives can be a guidance to the number of the representatives. A qualitative study (Elder et al., 2012) on the criteria health professions employed a purposive sampling when assessing their clinical communication skills. The number of the participant was generated on three fields of health professions: medicine, nursing, and physiotherapy. Different numbers of participant in each field were conveniently used in collecting the information as 13, 8 and 12 respectively. The total number of 33 participants in this research fits the recommended number for behavioral research; however, the number may be slightly inadequate when considering each field individually. In

fact, the number fits the latter category on controlled experimental research.

Another approach to the number of participants was partly influenced by Cizek and Bunch (2007) who suggested procedures in conducting standard setting to certify a testing program or a credential exam. Different from the normal number of participants at 20-25 in high stake assessment, the standard setting for credential programs require only two groups of panel, each consisted of 8-10 members (Cizek & Bunch, 2007). Although this performance-based assessment is aimed to be used in classroom, its model was generated from Thai credential exam. Adopted Cizek and Bunch (2007) guideline seems applicable.

Finally, the modified Angoff method is an approach for setting the standards for polytomous test item, the items that allow more than two choices of score as in correct or incorrect (Simon, Ercikan, & Rousseau, 2012). This research is a performance-based assessment which allows a continuum of the score given on each criterion. The judgment of Angoff approach that requires number of participants to set the standard for the test at 10-20 (Simon et al., 2012) appears suitable for this study. In conclusion, integrating these recommendations, the amount of participant at twenty is a justifiable number since this study specifies the field of the participant and considers the quality of the participants who can provide valuable sources of information.

The faculty at the targeted university established a cooperation with the drug stores around Thailand to accept pharmacy students for training, especially the drug stores in the cities which are tourist attraction. A part of the agreement the faculty made with the drug stores is the evaluation of the students' performance, which includes the students' ability in communicating with patients. At the drug stores, pharmacy students had to serve all kinds of patients while being under supervision of a pharmacist in charge. These pharmacists responsible for training the students were thus the aimed participants since they are

familiar with assessing pharmacy students when practicing their dispensing skills at the drug stores. The consent form and questionnaire were sent to a list of fifty qualified drug stores obtained from a Pharmacy school of a university from all over Thailand, which are Bangkok, Chiangmai, and Suratthani. Personal information obtained from the pharmacists was only their sex and the number of time in working at the drug stores.

Research instrument

Apart from consultation with informant specialists, this study utilizes a questionnaire to validate the content of the tasks and the rubric through various statistic methods. In order to develop the questionnaire, content is gathered from various processes. Before describing the questionnaire as the main instrument, details on instruments establishing the questionnaire are discussed in content development for a questionnaire, which is composed of an investigation on literature review, Thai dispensing assessment and semi-structured-interview.

Literature review

Previous studies vital to this research are those that provide a sample rubric and dispensing task. Two relevant studies (Kimberlin, 2006; Schwartzman et al., 2013) presented the overall view of what content matters to teaching communication to the pharmacists in the US. The former investigated the practice in assessing patient communication skills of colleges and schools of pharmacy in the US and Puerto Rico while the latter described the pharmacy curriculum in the US and Canada. The content of the two studies were compared to find the similarities as a part of selected content was shown in table below.

The skills stated by Kimberlin (2006) are mainly adapted into the questionnaire since it specified the detail better than Schwartzman et al. (2013), which mentioned the general topic of communication. The content from the secondary research are selected with the concern on the dispensing skills only. The topics that are irrelevant to dispensing skills with patient are not

included in the questionnaire. Apart from that, information from the interview with a pharmacy instructor was considered as well.

Table 14: Comparison between content of communication curriculum and skills

Content of Communication Curriculum	**Skills Identified on Assessment Forms for Pharmacy Student Communication With Patients**
Elements of communication process	Initiating communication Organizing the encounter Concluding the encounter
Psychology of communication	Establishing a trusting relationship
Nonverbal communication	Using effective nonverbal communication
General patient counseling techniques	Eliciting information from the patient Initiating educational interventions Encouraging patient involvement in communication and problem solving Verifying Understanding
Adherence	Promoting adherences to appropriate therapy

Thai dispensing assessment

The secondary research suggests the main topic for dispensing assessment while further information stems from Thai dispensing assessment. The guided dispensing assessment was obtained from the pharmacy students' practice book, which is given to all students at the beginning of the practice course. As every pharmacy student needs to pass the exam to obtain the license, pharmacy faculty offers courses that would facilitate the students in passing the requirement. Although the pharmacy council of Thailand does not force every pharmacy school and student to use the same rubric or supply a standard rubric, they provide a guidance describing the standard requirement, which the university adopted to develop the rubric. It is thus essential to refer

to the Thai rubric used as a standard for the students to follow when they practice dispensing drugs since the rubric developed conforms to the rules of pharmacy council of Thailand.

Table 15: The Thai dispensing rubric (translated)

Assessing Service Skills

Topic	Score weight	Needs improvement	Fair	Good	Excellent
1. Greeting (Starting the conversation).	1				
2. Demonstrating appropriate posture and body language.	2				
3. Establishing patient's identity or identifying the medication user.	3				
4. Using appropriate open-ended and closed questions.	3				
5. Appropriate and easy-to-understand language.	2				
6. Interviewing patients about drug allergies, underlying disease, pregnancy and breast-feeding in order to find possible problem from drug usage.	2				
7. Asking patient about current medical or herbal remedies and/or ask about past medication.	2				
8. Summarizing the issue and asking the patient to verify the information.	2				
9. Summarizing the diagnosis and suggesting a therapeutic regimen.	2				
10. Preparing the drug before dispensing and selecting the appropriate container.	2				
11. Preparing a correct, complete, and easy-to-read label together with appropriate caution.	3				
12. Rechecking to reduce the medication error.	3				
13. Explaining the instruction while handing medication to patient and assessing his/her understanding of the medication.	3				
14. Suggesting the appropriate practice for patient (*non-pharmacotherapy*)	2				
15. Ending the conversation. End of service.	1				

Assessing Dispensing Knowledge

Topic Disease or symptom	Score weight	Needs improvement	Fair	Good	Excellent
1. Knowledge about the disease or symptom.	2				
2. Reasoning for dispensing particular drug(s) (IESAC).	3				
3. Knowledge about the drug(s) dispensed.	2				
4. Knowledge about the appropriate practice for patient.	2				
5. Overall impression of test taker's knowledge and understanding	2				

The rubric is originally written in Thai and translated to English to compare the content with secondary research. It is divided into three main parts: assessing service skills, assessing dispensing knowledge and comments. The score weight and the performance levels are given for all topics in the first two sections. The performance levels are divided into four levels: needs improvement, fair, good and excellent. In this rubric, starting and ending conversation are two simple topics that are assessed at only two levels of needs improvement and excellent.

Not all the topic in Thai dispensing assessment was included in the semi-structured interview and the questionnaire. For one reason, the rubric pays more attention on assessing pharmaceutical knowledge while the study focuses on assessing dispensing skills in English. In fact, topics included in the questionnaire were contributed from Thai rubric and the literature review following by the consultation with informant specialists.

Semi-structured interview with students

In addition to the fact that semi-structured interview was developed from previous research on pharmacy dispensing practice and the rubric that is used in Thai dispensing assessment, some of the questions stem from the challenges faced in classroom instruction and assessment. All of the questions are designed to guide the format of the dispensing assessment in English according to the participants' perspectives. Nine questions cover various

considerations in dispensing assessment such as time, rater and rubric. The guided questions are listed as the followings:

1. How much time do you think it is needed for the whole dispensing process in Thai?
2. How much time is allowed for dispensing exam in Thai?
3. How much time do you think is needed for dispensing exam in English?
4. How many raters are there in Thai dispensing assessment?
5. Do you prefer role-play with your friend to the rater?
6. Do you think that the rubric for dispensing exam in English should be similar to Thai's?
7. Did you (participant as a test taker) study the rubric for Thai dispensing assessment before the exam?
8. Do you understand the Thai dispensing rubric?

It is found that the level of pharmacy students, which was categorized according to their English grades received when they were in their first year, ranges mainly from grade A to C. To get the sample of all level, the semi-structured interview was conducted on three volunteers who were a representative of different English proficiency level as in A, B, and C. These three representatives were selected through random sampling method and volunteer participation. This study included the whole year of pharmacy students as a group of participants. The volunteer participation on a specific process was inquired per case as applied in semi-structured interview. First of all, students were listed in three groups according to their grades; one student from each group was randomly selected and asked for their volunteer participation. If the chosen student denies volunteering, another student was randomly selected until volunteer participation is answered.

Questionnaire

The questionnaire consists of three parts: personal information, communication skills and pharmaceutical science content. Developed to fit both the students and the pharmacists, the questionnaire was divided into two forms of questionnaire, which is different only in the personal information part. The first part of questionnaire asks the students to fill in their major and their English grade, while the pharmacist's version request information about their work field and time of work experience.

The communication skills and pharmaceutical science content section was developed based on the Thai dispensing rubrics and survey on skills identified on assessment forms for pharmacy student communication with patients (Kimberlin, 2006) . Some topics derived from Thai dispensing rubric, some from Kimberlin (2006), and some from both Thai dispensing rubric and Kimberlin (2006) . It contains different speech acts according to the speech events. The questionnaire was validated before distributing to participants using the IOC.

Table 16: Questionnaire content

	Items	Thai Rubric	Literature reviews
1	Greet patients.	✓	✓
2	Identify yourself as a pharmacist to the patients.		✓
3	Introduce your name to patients.		✓
4	Confirm patient's identity (whose medication is for?).	✓	✓
5	Ask for patient's name.		✓
6	Call patients by name.		✓
7	Offer warm greeting		✓
8	Use pace and silence appropriately.		✓
9	Speak loudly enough to be easily heard.		✓
10	Use appropriate tone of voice.		✓
11	Use correct English language.		✓
12	Use correct English pronunciation.		✓

13	Use words the patient will likely understand and avoid medical jargon.	✓	✓
14	Modify communication to meet special needs of patients (e.g., elderly, low health literacy, cultural differences).		✓
15	Use open-ended questions and closed-ended questions appropriately to avoid bias (e.g. This drug is very strong. Did you throw up when you took it? – Inappropriate What's your symptom? - appropriate).	✓	✓
16	Use written information to emphasize and help oral communication (Write the name of medication and the instruction on the drug label).	✓	✓
17	Summarize information (the medication and the instruction).	✓	✓
18	Ask if there is anything else patients would like to discuss.	✓	✓
19	Invite patients to contact if questions or concerns arise.		✓
20	Thank patients.		✓
21	End the conversation politely.	✓	✓
22	Demonstrate appropriate eye contact.		✓
23	Demonstrate appropriate posture and body language.	✓	✓
24	Wear appropriate attire.		✓
25	Display appropriate health-professional manner.		✓
26	Ask patients about their concerns or reasons for visit.	✓	✓
27	Give patients opportunity and time to talk.		✓
28	Ask for a complete record of patients' current health conditions and therapies.	✓	✓
29	Ask questions to assess patients' understanding of key information about medications.	✓	✓

30	Ask questions to assess patients' experience with medications currently being taken.	✓	✓
31	Emphasize key information.	✓	✓
32	Provide reasons for advice and options for treatment.	✓	✓
33	Provide appropriate recommendations based on IEASC[1].	✓	
34	Discuss one drug or therapeutic regimen at a time.	✓	
35	Provide complete and clear instructions on medication.	✓	✓
36	Verify patient understanding of new information provided.	✓	✓
37	Work with patients to schedule the doses.		✓

Apart from personal information, the questionnaire asked the participants to respond to two sections: Overall communication skills and Pharmaceutical Knowledge. The participants were asked to rate each item on a four-Likert scale, which are numbered 4-1 as in very important, important, somewhat important, and not at all important respectively. In each section, a space was provided if the respondent would like to add any suggestions.

All of the items were calculated to find mean and standard deviation among the experts and the students, which are divided into two departments: pharmaceutical care and pharmaceutical science students. In each part, the result of the expert is presented first following by the students as a whole and the students according to their department. The items are considered based on the mean value it receives. They are divided into three groups as in ‘highly important’ when the value is higher than three, ‘important’ when the value is more than two and ‘not important’ when the value is less than two.

Data collection

Two groups of participants, fifth-year pharmacy students and pharmacists, are needed to answer the questionnaire. Different data collection is applied to these two groups.

Fifth-year pharmacy students

Apart from the semi-structured interview, two further steps for collecting data are questionnaire and the students' score. The dispensing assessment is a part of English for Pharmacy Profession II. The content and the assessment are provided to students at the first order, which is Chapter One in their textbook. In order to understand the students' perspectives before the content of the class guide their response, the questionnaire was distributed to students in the first class of their English for Pharmacy Profession II. Permission for fifteen minutes was granted by instructors to explain the detail of the questionnaire and allow students to inquire information before filling in the questionnaire, which was attached with the consent form. The questionnaire was distributed around the classroom. Students who do not want to participate in the research can refuse to do the questionnaire.

The second part of data collection, students' score, is a part of classroom assessment. The performance of the students was video-recorded so that the second and or the third rater can assess their performance. The third rater assessment was needed when the difference in score between the first and the second rater is greater than fifteen percent. The student data have been collected through video recording in a semester of academic year 2016 accounting for 142 students and 8 cases.

Pharmacists

While the student's questionnaire version could be distributed in the classroom, the pharmacist's version was conducted in another way. A name list of drug stores around Thailand was obtained from the Pharmacy faculty of a university in Bangkok. The pharmacist experts of this study exclude pharmacists who work in the hospital since the skills relevant to the assessment is the dispensing skills at the drug store. A list of fifty qualified drug stores from all over Thailand was obtained from the faculty of pharmacy at a university in Bangkok. This research employed only the drug store listed by the university

since the pharmacists at these drug stores are responsible for training the pharmaceutical students during their fourth-year internship program.

A consent form and a total of sixty-five questionnaires in Thai were mailed to fifty drug stores except for one drug store in Bangkok which requests for an in-person delivery. The participants have three options to return the questionnaire. The first method is to mail it back using the enclosed envelope with the postage attached. The second and the third method are taking pictures of the questionnaire and returning it by e-mail or by Line application respectively.

Raters (Language instructors)

A total of six raters were used in this study, all of which are language instructors for English for pharmaceutical course. Five instructors from five sections are used in the study, while one rater who did not teach the course was used as an extra. The two-hour training session was offered to the raters with additional information to study. Five raters who teach for five sections were assigned to role-play with students from another section. The raters did the role-play and scored the students for the first round. The raters then watched the video of the students in their section and rated them for the second round. Attaining internal consistency of reliability, the index for the performance scale was recommended at .93 or higher (Ryan & Lopez, 2001). This means that if the difference in score is greater than 7% of the raw score, which is 3.5 out of 50, the raters are asked to review their score to retain the internal reliability. If their decision is unchanged, the third rater steps in and the scores were calculated based on the scores from three raters.

Data analysis

Two procedures of data analysis are conducted in order to develop and validate tasks. The first procedure is analyzing two sets of data; the data from the interview with informant specialists and pharmaceutical students, and the data from the questionnaire by computing information to find agreement among pharmacist

experts and students through ANOVA and Games Howell post hoc test. The result is then discussed with informant specialists to finalize the task validation and to determine the rubric criteria. The second procedure is to validate the rubric composes of quantitative and qualitative studies. For the quantitative part, the scores from the raters were gathered in the excel sheet for MFRM analysis. The specification file included three facets. The first facet is the scores that reflect the candidates' dispensing performance. The second facet is the number of the raters, which is used to calculate the rater difficulty. The third is the criteria, which can determine the difficulty level of each criterion. The score of the students analyzed with MFRM results in Wright map, rater measurement report, and criteria measurement report.

For the qualitative section, four raters were available for the interview on the phone for approximately 15 minutes each. A set of questions (See Appendix) was sent to the raters at least a week before the appointment. The interview was transcribed and analyzed according to the themes, which are confidence and comfort, easiest and hardest section, Pharmaceutical science skills criteria, language use criteria, strategic competence criteria, adequacy and ability to assess, time constraint, and recommendation. This part of data was matched to the quantitative result to explain the findings in more detail.

Task development

The development of the task started with the interviews with informant specialists to gain the overall picture of the Thai dispensing situation, the curriculum, the license exam, and the needs of the stakeholders. After that, the interviews with the students were conducted voluntarily on three students from different English proficiency level. The interview with informant specialists and students was analyzed in terms of content relating to Hymes' SPEAKING grid (Cameron, 2012), which is the base the task outline was developed on. Apart from setting guideline for the task, this set out the detail for the administration of the test. The information from this process outlined the content of the task and test specification, which was developed with

consideration on data analysis from the interview and questionnaire results. The main content of the task was provided by informant specialists and adapted concerning the detail listed in the dispensing SPEAKING grid. The tasks were discussed and trialed with informant specialists. Additional matters arise during the consultations were added to the task detail.

Rubric development

Considering the previous rubric used in classroom assessment and Bachman's reciprocal language use and ESP' characteristics, the description of the rubric in analytical scale was developed based on the data comparison to the result from the questionnaire. The questionnaires result from the pharmacists and the students are analyzed to find mean and standard deviation of each speech act. Any item with the mean value less than two was eliminated. The data is further analyzed in one-way ANOVA to find the correlation among three groups: pharmacy experts, pharmaceutical care students, and pharmaceutical science students following by post hoc analysis. The final process to approve the rubric and assigning the scores was to consult the informant specialists and language experts. The scores of three parts were divided among three sections with slightly more percentage to language use and strategic competence. The scores assigned to each criterion were decided first on the pharmacy skills by informant specialists and on the other two sections by the language experts.

Task and rubric validation

Four different aspects to attest validity of this test are content validity, concurrent validity, construct validity and consequential validity. Firstly, content validity was conducted through questionnaire and semi-structured interview with pharmacists. The result from the questionnaire was analyzed to find mean, standard deviation of each item in the task. The items were eliminated and selected based on the qualitative data and several consultations with informant specialists.

Secondly, construct validity obtained through Item Objective Congruence by language instructors and pharmacists. The item selection was relied on IOC index. Consulting the content of the test task with language instructors, whose focus is on the language assessment, and informant specialists who aimed to test dispensing skills, can affirm the context validity of the test task.

Table 17: Overview of validating and proving test usefulness

Instrument	Test usefulness	Participants	Data analysis
Questionnaire	Content Validity	Pharmacists and Pharmacy students	Mean, standard deviation, ANOVA, Post hoc analysis
Interview	Content Validity	Pharmacists and Pharmacy students	Content analysis
Item Objective Congruence	Construct Validity	Language instructors and pharmacists	IOC index
Students' feedback	Consequential Validity (Washback)	Test-takers	Content analysis
Score	Marker Reliability	Test-takers and raters	Many-Facet Rasch Model
Interview	AUA	Language instructors	Content analysis

Thirdly, consequential validity is performed on test takers after the exam through a feedback questionnaire, which demanded information about the preparation process. These comments on the dispensing test from the students were analyzed to support AUA, which is to attest the task and rubric validation. The comments described the perspectives toward the usefulness of the test, which reflected the washback or consequential validity the test had on students' current and future performance in dispensing drugs. In addition, the interviews with language instructors as one of the stakeholders were added to support the AUA as well as the washback.

While parallel-form was used on the content validation, the marker reliability was inspected through MFRM, which produced the fair scores for the students and reported on the rater and criteria measurement. The score of the test takers and raters will be compared and analyzed to measure the correlation in Many-Facet Rach Model measurement (MFRM), which reports on Wright map, candidate measurement report, rater measurement report, criteria measurement report, and rating scale category functioning. The information on the rater and criteria measurement can also be used as backings for AUA as well.

Assessment Use Argument (AUA)

The result from the data analysis was provided and explained to conform to AUA, which consists of four principles, which are the intended consequences of using the assessment, the intended decisions to be made, the intended interpretations, and the intended assessment records. First of all, the intended consequences of the developed test were explained and provided to the stakeholders, such as pharmacy faculty, pharmacy instructors and language instructors. Second, a rationale for the intended decisions, which is linked to the consequences, was explained in accordance with backing for claim 2. Third, the intended interpretation about three aspects of students' dispensing ability was discussed with backing for claim 3. Finally, the description on the assessment record was offered with possible backing for possible sources of inconsistencies.

Chapter 4

Findings

This chapter presents the findings and the discussion, which is composed of three main sections according to the objectives.

Objective number one

To create tasks for measuring English oral communication competency of Thai pharmaceutical science students.

Two methods are employed for this purpose: semi-structured interview and consultations with specialist informants. The interview was performed on both informant specialist and the students. The data obtained from the specialist was conducted first as it is a guidance to the semi-interview and the questionnaire for the students. After the interview, the task was designed based on the factors suggested from the interview following by the consultation with informant specialists. The cases were discussed with informant specialists to fill in the information that may be needed.

Interview with informant specialists

The information on classroom assessment on drug dispensing stems from an interview with informant specialists who are teaching and managing internship courses for pharmacy students at the university, where the assessment is developed, validated and evaluated.

Basic information

Pharmacy curriculum is a six-year program that offers two majors: Pharmaceutical Care and Pharmaceutical Sciences. Pharmaceutical Care trains the students to work with patients in situation of hospital and drug stores, while Pharmaceutical Sciences aims to develop pharmacists for pharmaceutical companies. Both majors provide the same fundamental pharmaceutical knowledge needed for dispensing drugs in the first five years and require students to do a practice in pharmacy field for three courses: after their third

years, during their fourth year and their fifth year. The students study their major subjects starting in their fifth year.

Preparation for dispensing assessment

According to Thai law, pharmacy students can dispense drugs to patients under a pharmacist's supervision to obtain the 400-hour training requirement before they can apply for Objective Structured Pharmaceutical Examination (OSPE). The main purpose of these courses is to prepare students to learn about the system and to ask patients questions according to different settings: primary care units, hospital and drug stores. After students complete the curriculum, they can take OSPE to obtain the license and be able to dispense the drugs on their own.

After 3rd year	4th year		5th year	
Summer	1st semester	2nd semester	1st semester	2nd semester
Professional Pharmacy Practice I (4 weeks each)	Professional Pharmacy Practice II (5 weeks)		Professional Pharmacy Practice III (5 weeks)	
Primary care units	Hospitals		Drug stores	

Figure 4: Internship schedule

The curriculum equips students with dispensing skills by offering communication skills course in Thai and two internship courses called Professional Pharmacy Practice I, II and III as shown in Figure 4. The internship practice focuses not only on the dispensing skills, but also the managing skills as the primary care units, hospitals, and drug stores. The students can take the Pharmacy Practice I, which is an internship at primary care units or primary hospitals, after their third year during their summer semester. Pharmacy Practice II distributes the internship hours at hospitals to students throughout

their fourth year, while Pharmacy Practice III is the internship at the drug stores occurs in the fifth year of the study.

Professional Pharmacy Practice I

This course is in the curriculum delivered in Thai (การฝึกปฏิบัติงานวิชาชีพเภสัชกรรม 1). It focuses on the interview, physical check-up, and basic knowledge on chronic disease. Places for pharmacy students to practice are primary care units and primary hospitals in Thailand. At the community health center, some opportunities are provided to students on dispensing drugs to patients. Most of the task in this course is however the practice on dispensing drugs according to the prescription. Interestingly, most of the pharmacy students' role is dispensing and taking care of elderly patients with chronic diseases. The letter grade based on criterion is assigned after the students completed the course.

Professional Pharmacy Practice II

Pharmacy practice II (การฝึกปฏิบัติงานวิชาชีพเภสัชกรรม 2) is operated at the hospitals around Thailand. Each student needs to fulfill two hundred hours of training. The purpose for this course is for the students to learn the pharmacy system at the hospital. The dispensing role of the students does not involve diagnosing the possible disease since they dispense drugs according to the prescription of the doctor at the premise. The letter grade based on criterion is assigned after the students completed the course.

Professional Pharmacy Practice III

This course (การฝึกปฏิบัติงานวิชาชีพเภสัชกรรม 3) assigns students to drug stores around Thailand. Two hundred hours are requirement for the students to fulfill. The focuses of this course are on basic interview with acute disease or ambulatory patients (those who can visit the pharmacist in person) and evaluate the patients and dispensing drugs. This course provides greater opportunity to students to practice dispensing skills comparing to Pharmacy

Practice I. The letter grade based on criterion is assigned after the students completed the course.

Dispensing assessment in Thai

"Dispensing is a skill that needs a continuous practice."

Dispensing assessment in Thai is a part of professional Pharmacy Practice III. At this university, students are required to work at a university drug store for about twenty-two hours in nine weeks. The assessment is conducted when a student has a face-to-face conversation with a real patient under instructor's supervision. The instructor supervises the dispensing by listening to the conversation, guiding the students in some cases the questions to the right issue as soon as possible and make certain that the drug dispensed is appropriate for the case.

Duration

The dispensing assessment with the instructor allows students six minutes to interview the instructor who acted as a patient seeking for medication. The second type occurs under instructor supervision at the university drug store where real patients come in. Although most of the patients who come into the university drug store gladly incorporate the pharmacy practice, the time allowed for this real situation is no more than five minutes in order to prevent the patient from feeling uncomfortable

Schedule

From the total of nine shifts, which lasts two hours and a half each, the assessment is scheduled for two times during the practice at the university drug store: the third or the fourth shift and the eighth or the ninth shift.

Classroom assessment research might be able to adopt these well-established international and national tests into the curriculum; however,

none of them is a suitable option when considering attributes of individual and the characteristics of the language use task and situation. First of all, OSCE and OSPE are tests that assess the professional knowledge. The latter is done in Thai which does not respond to the need of English language assessment. In addition, the characteristics of the test takers in this research do not differ as much as OET, which is meant for any foreigners who wish to work in English speaking countries. Some personal attributes of this assessment research can be controlled in terms of age, nationality, native language, level and type of general education, type and amount of preparation or prior experience with a given assessment. The target group is Thai fifth-year pharmaceutical science students at a university in Bangkok. They are Thai resident in the age range of twenty-one to twenty-two who speak Thai as a native language and English as a foreign language. They specialize in Pharmacy and have participated in a three-month internship at the drugstores in Thailand. These similarities among the participants' personal factor, thus, encourages the present researcher to develop the assessment that can better correspond the group's variability and offer the closest interpretation of the participant's performance.

Needs for dispensing drugs in English

As mentioned above, all pharmacy students are required to obtain the training hours at drug stores for two semesters. The faculty has established cooperation with the drug stores around Thailand to accept these pharmacy students, especially the drug stores in the city which is a tourist attraction. A part of the agreement the faculty made with the drug stores is the evaluation of the students' performance, which includes the students' ability in communicating with patients. At the drug stores, pharmacy students had to serve all kinds of patients while being under supervision of a pharmacist in charge. Some complaints were received from the supervisor regarding the students' performance in communicating in English. Many drug stores reported that students are not able to communicate with foreigners. In fact, one of the

suggestions from these drug stores is for the students to improve their English communication skills while dispensing drugs to patients. The faculty thus requested for an English course that can deal with the issue.

Semi-structured interview with students

The semi- structured interviews on students were conducted in Thai based on nine guided questions. Questions regarding their internship at the drugstores were asked to confirm that they have an experience in dispensing at the drug stores and the dispensing assessment as a course of their curriculum requirement. About fifteen to twenty minutes were spent on each interview. In this study, a representative of three groups of students are named as student A, B, and C, for advanced, intermediate and pre-intermediate level respectively.

Needed Time on dispensing process in Thai

A: *"I think five minutes is more than enough. If dispensing process takes longer than five minutes, the patient might not like the fact that they have to visit the drug store longer than the others."*

B: *"Fifteen minutes if the patient allows."*

C: *"I think five minutes would be enough."*

This question is intended to ask students about the time they think is needed for dispensing. From the result, the level of English proficiency does not really determine the time needed for dispensing. Student A and student C think of the situation as similar to the real Thai dispensing exam, while Student B is concerned about the dispensing skill. All the participants know that patients do not like to spend longer than five minutes talking to the pharmacist.

This fact matches the information from informant specialist who teach the students the dispensing skills. Two reasons can be claimed for the five-minute reason: the teachers' saying about the time and the students' experience in the drug store. While the previous reason may need some confirmation from the students, the latter reason is obviously stem from

students' experience. From the answer to this question, the time students needed for dispensing skills in Thai seems to rely on the students' experience and their communication skills in dealing with the patient.

Time allowed for dispensing exam in Thai

A: *"Three to five minutes."*

B: *"Ten minutes including preparation time."*

C: *"It depends on the cases. Most of the time is about three to five minutes. Easy cases can be done within three minutes. If the patient allows the time then it can be five minutes. The time allows for the Auspy test to obtain the pharmaceutical license is three minutes."*

This question aims at verifying the fact about the time for dispensing. Although the answer of one participant did not match the others, the time for dispensing exam can be identified at three to five minutes. The possible reason for student B to state different timeline is because the exam was viewed to the student as a whole process which includes preparation time as well.

The informant specialist stated the dispensing time during the internship and the exam at five minutes per patient. In fact, the license exam allows only three minutes for dispensing, while the university version allows up to five minutes. This instruction can prepare the students to get accustomed to the limit timeline for the exam.

Needed time for dispensing exam in English

A: *"I think English would require some more time at seven minutes since I might need more time to think about what words to use."*

B: "*Fifteen minutes."*

C: *"I have problem communicating in English which is why I would need a lot more time than the Thai version. Ten minutes would be ideal for me but the patient might have left the store by then."*

This question intends to gather some idea about the time needed for students to dispense drugs in English. The answers vary from seven to fifteen

minutes. The answer of Student C is similar to the time needed for Thai dispensing exam at fifteen minutes. Student A and B requires a similar time set at seven to ten minutes as they share the same concern which is on the vocabulary recognition while speaking.

In general, it is noted that the students believe they need more time for dispensing in English than in Thai. This is due to their concern on their English proficiency. This interview was; however, done at the beginning of the class, which is before they were taught dispensing skills in English. Their concept on taking a dispensing exam in English might not be clear to them as to what language to be used and how much time they would really need. In other words, they might be too concerned about their performance and the time, which leads to the idea that more time is required for dispensing exam in English.

Number of raters in Thai dispensing assessment

A: *"One rater. I think one rater is enough."*

B: *"One rater, but I think it would be better to have more than two persons."*

C: *"One rater is fine with me considering the stress in the exam situation. However, if the concern is on the score, two raters are much better than one."*

Same response was obtained from three students as one rater at a time. This question is to confirm the number of the rater in their regular dispensing exam. Two students believe that if possible, they would prefer having more than one rater to counterbalance the score.

Apart from this information, the students described the situation of the dispensing exam as a bit stressful and authentic since their rater waited for the real foreign patient to visit the store to conduct the test. Considering the situation, it is reasonable that only one rater is allowed. The patient might not feel comfortable consulting with more than two pharmacists. Although the store they visit is known as a university drug store, the patient might feel awkward being in such situation and might affect the student pharmacist's performance.

Preference on role-play interlocutor

A: *"I prefer having my friend doing the role-play. I do not need to choose the partner. The teacher can pair the partner for me. Also, I believe it's less stressful than doing the role-play with the teacher."*

B: *"I would rather meet the real patient as in real situation and not the prepared conversation with a friend."*

C: *"I prefer the instructor considering the experience. I don't think that I can communicate in English."*

This question aims at establishing the role-play interview. The answers vary from preference on friends to rater or the real patient. Student A, who seems to have no problem in using English prefer to use friend in the role-play to avoid the stress with the teacher, while Students C who is concerned about his English performance and dispensing skills prefer the rater.

Both student A and student C answered this question based on the same concern, stress and English proficiency. Although their English competence is totally different, both of the student answered the question based on their proficiency-level ground. Student A who is capable of communicating in English concern about the stress and not his friend whom Student A unconsciously set the capability to be the same as his. On the other hand, student C, who is worried about the English competence and that he/she might not be able to act as well due to the proficiency level, prefer an instructor to a friend to avoid the problem. Student B, on the other hand, preferred the instructor option as the performance should be tested in the real situation not the prepared one.

Similarity of dispensing exam in English and Thai's

A: *"The process should be similar."*

B: *"Having an instructor as a patient if the instructor has ability to perform in English is my preferred option. I think the process in dispensing should be the same as in Thai."*

C: *"It would be better to have the dispensing process similar to the Thai version so that it is easier to perform."*

This question intends to generate students' idea on the dispensing process in English. All of the three students agree on similar exam process, which in this case means the process of gathering information from the patient to the drug explanation. In other words, students prefer to perform the same dispensing process as in Thai with the language change.

It was no surprise that the students would prefer the same dispensing process as in Thai version. They have been taught and prepared for their license exam for two semesters. Requirements on what to ask and specific detail are learnt as well. Thus, it would be best for the students to focus on the language while applying the same pharmaceutical knowledge in the dispensing exam.

Study the rubric before the dispensing assessment

A: *"Yes. The rubric is actually in the internship manual."*

B: *"Yes, some preparation on that like medical history, allergy."*

C: *"Yes. The rubric is provided in the internship manual."*

The question aims at investigating the rubric use of students. All the students use the rubric as it is provided in the internship manual. The students use them to follow the procedure guided as a recommended procedure.

Using the rubric can be another indicator to explain why the students prefer to have the same procedure of dispensing as in Thai version. If they are used to using the Thai rubric, they would prefer using the similar version in English. The number one factor that may reflect the different performance between the Thai and English dispensing exam is their English proficiency level.

Thai dispensing rubric comprehension

A: *"Yes. In fact, I can ask my rater after the test about what to improve as well."*

B: *"Yes."*

C: *"Some confusion on different types of drugs that can be dispensed to patients."*

This question aims at verifying the students' understanding of the Thai rubric after couple uses in the last two semesters. Two students agree that they understand the rubric while one student stated that he was not sure how some content of rubric is scored. A furthered question was asked concerning their understanding after their performance is reviewed. The students do not have problem with the detail in the rubric but question the way the score is given.

This is meant to identify any obstacles the students might have with the rubric in the English version. The level of students' understanding is necessary in determining whether they can achieve what the rubric required them to. The performance of the students relies on their thorough understanding of the rubric.

Other concerns

Apart from the nine questions above, some further information about the dispensing was provided which matches the information provided by informant specialists. All of the three students were asked a furthered question whether they want to do a prepared case or an impromptu case.

Student A

Student A would like more time to practice dispensing with the real patients during the internship. Student A provided information about the exam that the students have to take an exam two times: one before and one after the practice.

"If considering the score, preparing the case in advance would allow the chance to receive a higher score. But if we want to assess the real performance, the impromptu case would be better."

Student B

Student B explained the time for the dispensing skills as five minutes allowed to study the rubric after the assessment and another five minutes to ask the rater about how to improve the performance.

Student B believed that his or her performance can be worse if dispensing is conducted in English comparing to Thai.

"I would like some time to prepare for the role-play to get the best score. But if I can choose whether to do a prepared case or an impromptu one, I would choose the impromptu case as that can reflect my real ability in dispensing drugs in English."

Student C

"I do not have any confidence in communicating with the foreigners. During the internship, I would step out when the foreigner came into the drug store."

One interesting point the student made is that in case of recording the student's performance on the video, the student doesn't feel the difference in stress.

"It's not easy to follow the requirement in the rubric when talking to real patients. I sometimes wonder about the rating scale as in some cases, the patient provided information without having to ask, which seems to contradict the rubric scale that require me to ask for such information."

All three students have concerned over their performance which affects the scores. The impromptu case appeared to be their preference to reflect their real ability on the condition that they are allowed plenty of practice to make them feel comfortable with the exam and the rubric. This meant that more practices should be provided to these students before the real dispensing exam occurs, which affects the scores they concern on. In addition, rubric usage for case samples should be supplied to students as well. Such supplement can increase the transparency of the test.

Dispensing tasks based on SPEAKING Grid

The SPEAKING grid was developed in order to generate the task that fits the dispensing situation. A total of eight cases were selected from a university drug store's note, which is a compilation of every day patient cases that the students record during the internship program. Apart from the consideration on the SPEAKING grid, the selection of these cases is based on the construct and the test fairness. Since the test can't be administered to every student at the same time, the time allotment and the administration need to be well organized. A total of five raters were employed to test 145 students. The test was administered in two time slots: morning from 9:00 – 12:00 and afternoon 13:00 – 16:00. The cases were divided according to the time slots as four cases in the morning and four cases in the afternoon. The task is developed based on the model of SPEAKING Grid for dispensing task analyzed from the content on the interview and the consultation with informant specialists.

Apart from the use for generating dispensing tasks, the information in the SPEAKING grid was used to create manual for dispensing assessment. The manual is composed of three sections, which are guidelines for administrator, raters, and students. The manual can be distributed to all parties involved for their understanding about the whole dispensing process and the rule. The speaking grid contains detail that it is appropriate for raters and administrator. The manual, however, directs at how the assessment process occurs, which provides general idea and suits those who are new to the process.

S for Setting and scenes

Three scenes are involved during the whole administration: the preparation room, in front of the exam room, and the exam room. Every test takers are gathered in a preparation room before entering an exam room. The preparation room should be spacious enough to serve at least 80 students at a time. The test takers are required to wait in the preparation room until they are called to wait in front of the exam room. In the preparation room, the test takers can study and talk with their peers but they are not allowed to use their

cellphone until they finish the test. The test taker who finished the test can collect their belongings from the preparation room and leave the room without discussing with the peers who have not taken the test.

Two chairs are set in front of the room for the test takers to sit while waiting for their turn. While one person is taking the exam in the room, two persons are sequenced to be waited in front of the exam room. During the wait, the test takers can talk to each other but are instructed to use low volume to avoid interrupting the test in the exam room. Once a test taker leaves the room, another test taker can enter the room when permitted by the rater.

The exam room is situated in a classroom where the normal classes are taught in. The setting is the same as a normal classroom except that one table is moved to be next to the instructor table, so that the speech can occur in a normal classroom to simulate a drug store situation, where the distance of the participants in the drug store is often close to each other. Unlike the drug store situation where both the pharmacist and the patient might stand and talk, this speech is set up with two chairs and two tables. The participants sit on separate table which is facing each other. In fact, the test taker sits across from the rater who acts as a patient. Each test taker acting as a pharmacist is allowed five minutes to perform the dispensing skills.

P for Participants

The participants in the test room are a rater and a test taker. The rater acts as a patient, while the test taker acts as a pharmacist. The rater performs as a patient needs to at the same time rate the test taker during the speech. Both of the speakers can happen to take notes of essential information in order to fulfill the task. Apart from the rater and the test taker, at least two persons are needed to administer the test. This administration process involves sequencing students according to the time slot they are assigned and separating the test taker who finished the test from the test takers who have not taken the test.

E for Ends

The purpose of the speech is for the test taker to achieve the dispensing skills in English, which in this case is dispensing drugs according to the symptoms obtained from the patient. In order to achieve such outcome, the test taker needs to ask questions needed to decide the appropriate drug(s) for the patient. A list of needed information was provided in the rubric given to the test takers for practicing. Apart from naming the drug and the dose, the reasons for dispensing the drug, the instruction and the recommendation are required as well, all of which are listed in the rubric in the part of pharmaceutical knowledge.

A for Act sequence

The sequence of the act in dispensing generally composes of six speech acts: greeting, asking for help, questioning, answering, suggesting, and ending the conversation. The speech starts with the greeting from both speakers. During or after the greeting, the offer for help might be proposed by the pharmacist, or the patient can be the one asking for help. After that the pharmacist needs to obtain patient's information and continues the act by questioning with the patient answering the questions. After the information was obtained, the pharmacist makes a decision and performs the suggesting speech act to the patient and questioning whether the patient agrees with the drug recommendation. The patient might perform the questioning act on the drug, the usage and the suggestion. Lastly, if no further questions are proposed, the ending the conversation is performed. However, the act sequence might differ in some situation that the patient may start the questioning instead of the pharmacist.

K for Key

The tone of the speech act is considered as various tones. Frist, a serious tone as this simulates a situation at the drug store, where a patient

comes in to seek for help from a pharmacist. Apart from the serious tone, the pharmacist would need to perform sympathy, professional and formative tone. As an illustration, the pharmacist should manifest the service as friendly and welcoming the patient to address the problem and willing to give needed information to patients, while maintaining the manner as a professional wise. Although it is video-recorded, the attire is not considered in this situation. The test taker does not need to wear a pharmacist gown. A normal university uniform is allowed.

I for Instrumentalities

The main channel of communication is listening and speaking. Reading might be involved at the beginning of the speech act as the test taker needs to read the prompt of the act as in age, sex, and initial symptom of the patient. The writing is also needed when the test taker needs to take notes of the patient's description and to write down the label for the patient. As a matter of fact, not all the test takers take notes during the dispensing test. Taking notes during the exam is not a requirement since no scores are given on the notes or the label writing.

The language is formal since the role of pharmacist needs to be professional but not too hard medical terminologies. The test taker might utilize some medical words, which can obstruct the conversation if the patient does not have such understanding. Thus, when medical words are involved, the test taker needs to use layman terms to make certain that the patient understands. In other words, medical terminology is allowed as long as the test taker can guide the patient through the conversation without meaning obstruction.

N for Norms

In real situation, the authority in this kind of conversation might appear to be the pharmacist due to one of the status relationship between the pharmacist and the patient, as a medical provider and a help seeker, which is

superior and inferior relatively. The norm of interaction is that the pharmacist possesses the right to question the patient regarding their medical information. In the meantime, the patient should also cooperate by providing information on such topic as much as they can. In other words, the power of pharmacist and patient are unbalanced, but complement each other since they need to both provide and receive information from each other.

In the test situation, the norms might not be so similar to the real situation in terms of being superior as the test takers know that the patient is acting to require the information for assessing their performance not for seeking for help. However, the rater needs to provide information and acquire the information needed for assessing the test taker. The power for pharmacist and patient is complement each other as well.

The test taker as a pharmacist needs to listen to the patient carefully, while the speed of their speech should not be too fast that it obscures the meaning. However, the speed of the whole conversation should not exceed five minutes as most patient do not expect to spend time at the drug store for too long. The test taker should avoid a long silence since this can result in the decrease in the professional trust. In terms of volume, the norm of interaction is that pharmacist should use a loud enough volume for the patient. Similar to the real situation, the test takers should be aware of using appropriate volume since the patient's information can be heard by the others. This is important since some questions and information can be sensitive to the patient. Last but not least, the voice of the test taker as a pharmacist should be in a welcoming and friendly tone. These specifications on the volume and tone of voice are in fact included in the rubric.

G for Genres

The genre of this speech event is considered a practitioner-patient interaction, or to be more specific medical consultation. This is due to the information provided and obtained that deals with patient's medical record. In addition, the aim of the speech event is to dispense the appropriate drugs to

the possible disease the patient describes. All the speech acts are performed to fulfill the medical service as to dispense the drugs to patients based on the information provided as a consultation to the pharmacist.

Objective number two

To create the rubric for measuring English oral communication competency of Thai pharmaceutical science students

Participants' information

The questionnaire was distributed to two groups of participants: pharmacy experts and pharmaceutical students. The main content of the questionnaire is the same for both groups. The difference is the personal information, where pharmacy experts were asked for the year of experience working in the drug store and students were asked which of the two departments they are in. Apart from that, the sex of the participants was one of the information obtained from the personal information section.

Pharmacy experts

The consent form and questionnaire were sent to a list of fifty qualified drug stores obtained from a Pharmacy school of a university. Twenty-two questionnaires were returned out of sixty-five questionnaires. One of the questionnaires missed some information leaving 21 questionnaires available for data analysis. The majority of the returning method is mailing, while two questionnaires were returned by e-mail, and one questionnaire was delivered and picked up in person. Most of the participants was not from Bangkok. 10 participants are male and 11 are females at 47.6% and 52.4% respectively. The dispensing experience at the drug store of the participants varies from 8 months at .60 to 34 years at 34.

Students

The students consist of two fields of study: Pharmaceutical Sciences and Pharmaceutical Care. At the beginning of their class time, students were asked to fill in the questionnaire and allowed them the chance of not

participating. A total of 132 students respond to the questionnaire, but three of them were found to be missing some part of the data leaving 129 students available for data analysis. A total of 39 participants were male at 30.2 percent and 90 were female students at 69.8%. 70 students are in the field of Pharmaceutical Sciences at 54.3% and 59 students are in the field of Pharmaceutical Care at 45.7%.

Questionnaire

Apart from calculating the mean, the items in the questionnaire were analyzed with ANOVA by comparing among three groups: pharmaceutical care students, pharmaceutical science students, and pharmacy experts. This is to find the difference among groups to analyze whether the three group agrees. In addition to the difference among groups, Games Howell post hoc test analysis was applied when the p-value in ANOVA is significant at .05 to find the difference between three pairs: 1) pharmaceutical care students and pharmaceutical science students 2) pharmaceutical care students and pharmacy experts 3) pharmaceutical science students and pharmacy experts. Post hoc test analysis can reveal where the difference exists if it is between each pair.

A total of six sections are composed of two main knowledge areas to be assessed, which are overall communication skill and pharmaceutical skill. The first section, overall communication skill composes of four divisions: initiating communication, verbal communication, concluding the encounter, and non-verbal communication. The second section, Pharmaceutical skills content is divided into two sections: Eliciting information from patients and Initiating education interventions.

Initiating communication

Initiating communication includes all the speech acts defining as important in pharmacy practice, which are greeting patients, identifying yourself as a pharmacist to the patients, introducing your name to patients, confirming

patient's identity, asking for patient's name, calling patients by name and offering warm greeting.

From the experts' response, three items are regarded as highly important as the means are higher than three, which are greeting patients at 3.57, confirming patients' identity at 3.81, and offering warm greeting at 3.52. Two items are regarded as important are identifying yourself as a pharmacist to the patients at 2.52 and asking for patient's name at 2.10. Two items are identified as not important are introducing your name to patients at 1.52 and calling patients by name at 1.67.

From the students' response, four items regarded as highly important are greeting patients at 3.47, identifying yourself as a pharmacist at 3.15, confirming patients' identity at 3.97, and offering warm greeting at 3.29. Two items are regarded as important are asking for patient's name at 2.29 and calling patients by name at 2.16. Only one item identified as not important is introducing your name to patients at 1.98.

Table 18: Response on initiating communication

Variables		Experts	Students	Pharm care	Pharm Science
1. Greet patients	Mean	3.57	3.47	3.59	3.32
	SD	.676	.674	.648	.681
2. Identify yourself as a	Mean	2.52	3.15	3.26	3.02
pharmacist	SD	.928	.811	.811	.799
3. Introduce your name to	Mean	1.52	1.98	2.13	1.80
patients	SD	.680	.834	.931	.664
4. Confirm patient's	Mean	3.81	3.97	3.97	3.97
identity	SD	.402	.174	.168	.183
5. Ask for patient's name	Mean	2.10	2.29	2.46	2.08
	SD	.831	1.025	1.045	.970
6. Call patients by name	Mean	1.67	2.16	2.33	1.97
	SD	.856	.917	.880	.928
7. Offer warm greeting	Mean	3.52	3.29	3.47	3.08
	SD	.512	.712	.631	.749

The result of the students agrees accordingly with the pharmacists on the items perceived to be irrelevant. While two items received a mean score under two which means not important, the item asking for patients' name receiving a very low score at 2.10 from pharmacists and 2.29 from students is considered to be relevant to the low-score item, since it involves name asking and calling. However, one item is found to be different between two groups is identifying yourself as a pharmacist.

When comparing three groups, which are pharmacy experts, pharmaceutical care students, and pharmaceutical science students, using ANOVA, the result shows that the p value of five items are lower than .05, which are identifying yourself as a pharmacist, introducing your name to patients, confirming patients' identity, calling patients by name, and offering warm greeting. Post hoc test analysis is then the next step in analyzing the data of the three items.

Table 19: Result of One-way ANOVA test on initiating communication

Variables	Mean Square	F	P-value
1. Greet patients	1.215	2.747	.067
2. Identify yourself as a pharmacist	4.434	6.544	.002
3. Introduce your name to patients	3.617	5.617	.004
4. Confirm patient's identity	.230	4.755	.010
5. Ask for patient's name	2.551	2.609	.077
6. Call patients by name	4.326	5.387	.006
7. Offer warm greeting	2.868	6.461	.002

Firstly, the result of post hoc test between pharmaceutical care students and pharmaceutical science students shows that two items are viewed differently by the two groups. The Games-Howell value of introducing your name to patients, and offering warm greeting are .052 and .006.

The second is to compare pharmaceutical care students and pharmacy experts. The result reveals that these two groups viewed three items differently, which are identifying yourself as a pharmacist, introducing your

name to patients, and calling patients by name at .008, .006, and .011 respectively. Lastly, comparing pharmaceutical science students and pharmacy experts shows that one item is in disagreement, which is offering warm greeting at .005.

Table 20: Result of post hoc test on initiating communication

Variables	Variance	Games-Howell
Identify yourself as a pharmacist	Pharm care and Pharm science students	.213
	Pharm care and experts	.008
	Pharm science students and experts	.093
Introduce your name to patients	Pharm care and Pharm science students	.052
	Pharm care and experts	.006
	Pharm science students and experts	.264
Confirm patient's identity	Pharm care and Pharm science students	.984
	Pharm care and experts	.194
	Pharm science students and experts	.219
Call patients by name	Pharm care and Pharm science students	.065
	Pharm care and experts	.011
	Pharm science students and experts	.379
Offer warm greeting	Pharm care and Pharm science students	.006
	Pharm care and experts	.920
	Pharm science students and experts	.013

From the result, pharmacy experts do not view identifying yourself as a pharmacist as much important as the pharmacy students in both fields. This is probably due to the status as a pharmacist that these experts have hold and been working as a pharmacist for a while. In addition, every drug stores need to display the sign of pharmacist(s) in charge. They might believe that there's no need to identify themselves as a pharmacist. For the students' view, it is probably due to the learning content following the law that emphasizes on identifying yourself as a pharmacist, since Thailand in the past has problem of drug stores with unlicensed pharmacists.

For introducing your name to patients, the result shows that Pharmacy experts and Pharmaceutical science students view this item in the same way while Pharmaceutical care students rate this item as important. Similar to

introducing your name to patient, pharmaceutical students viewed calling patients by name differently from the other two groups. This might due to the department they are in. Pharmaceutical students are trained to work in the drug store. One of the training purposes is for them to pay attention to customer lending them service mind. Although, the course taught to these students are the same, the result proves that the department these students are in can be a factor in deciding to introduce their name to patients and call patients by name. The pharmacy experts with the experience of working at the drug store might lead to the conclusion that such intimacy as in name introduction and calling each other's name is not needed for the dispensing situation.

Although the pharmacy experts do not view confirming patient's identity as much important as the pharmaceutical care students, the mean score is considered highly important. The pharmaceutical care students may rate this item differently higher than the pharmacy experts because they are training for the license exam which emphasizes greatly on confirming patient's identity. For offering warm greeting, pharmaceutical science students rate the item significantly different than the other two groups. As a part of the service mind, this group of the students may not be coached to attain the idea as strong as pharmaceutical care students and pharmacy experts who are working in the field. In addition, their goal after graduation is to work in the drug industry where greeting patients is not present. It is understandable that the students in pharmaceutical science field have such view since the occupation they are after do not have such requirement.

Verbal communication

Verbal communication consists of nine items, which are using pace and silence, speaking loud enough, tone of voice, English language, English pronunciation, avoiding medical jargon, modifying communication to meet special needs, using open-ended and close-ended questions, using written information.

Table 21: Response on verbal communication

Variables		Experts	Students	Pharm care	Pharm Science
1. Use pace and silence	Mean	3.33	3.28	3.36	3.19
appropriately	SD	.658	.612	.591	.629
2. Speak loudly enough to be	Mean	3.29	3.49	3.57	3.39
heard	SD	.644	.574	.554	.588
3. Use appropriate tone of	Mean	3.24	3.57	3.64	3.47
voice	SD	.625	.543	.539	.537
4. Use correct English language	Mean	2.57	3.36	3.41	3.29
	SD	.811	.610	.602	.617
5. Use correct English	Mean	2.86	3.17	3.31	3.00
pronunciation	SD	.655	.639	.603	.643
6. Avoid medical jargon	Mean	3.57	3.81	3.89	3.73
	SD	.598	.391	.320	.448
7. Modify communication to	Mean	3.52	3.71	3.83	3.56
meet special needs of patient	SD	.680	.474	.380	.534
8. Use open-ended and close-	Mean	3.48	3.63	3.74	3.49
ended questions appropriately	SD	.680	.485	.440	.504
9. Use written information to	Mean	3.29	2.98	3.09	2.86
emphasize and help oral	SD	.784	.729	.717	.730
communication					

From the experts' response, most of the items are rated highly important except for two items which are using correct English language at 2.57 and using correct English pronunciation at 2.86. From the students' response, only one item is rated under three, which is using written information to emphasize and help oral communication. Since none of the items are rated lower than two, all of the items are included in the rubric.

Table 22: Result of One-way ANOVA test on verbal communication

Variables	Mean Square	F	P-value
1. Use pace and silence appropriately	.493	1.302	.275
2. Speak loudly enough to be heard	.899	2.671	.073
3. Use appropriate tone of voice	1.424	4.693	.011
4. Use correct English language	5.822	6.544	.000
5. Use correct English pronunciation	2.468	6.293	.002
6. Avoid medical jargon	.925	5.253	.006

7. Modify communication to meet special needs of patient	1.458	6.001	.003
8. Use open-ended and close-ended questions appropriately	1.219	4.797	.010
9. Use written information to emphasize and help oral communication	1.603	2.995	.053

From the result, six items are viewed significantly different among the groups, while one item, using written information to emphasize and help oral communication, is considered to be partially significant at. 053. The six items are using appropriate tone of voice, using correct English language, using correct English pronunciation, avoiding medical jargon, modifying communication to meet special needs of patient, and using open-ended and close-ended questions. Six items are further analyzed with post hoc test, while the one with partial significant is eliminated.

Table 23: Result of Games-Howell post hoc test on verbal communication

Variables	Variance	P-value
Use appropriate tone of voice	Pharm care and Pharm science students	.184
	Pharm care and experts	.031
	Pharm science students and experts	.285
Use correct English language	Pharm care and Pharm science students	.474
	Pharm care and experts	.000
	Pharm science students and experts	.003
Use correct English pronunciation	Pharm care and Pharm science students	.014
	Pharm care and experts	.020
	Pharm science students and experts	.667
Avoid medical jargon	Pharm care and Pharm science students	.068
	Pharm care and experts	.074
	Pharm science students and experts	.521
Modify communication to meet	Pharm care and Pharm science	.005

special needs of patient	students	
	Pharm care and experts	.143
	Pharm science students and experts	.974
Use open-ended and close-ended questions appropriately	Pharm care and Pharm science students	.010
	Pharm care and experts	.227
	Pharm science students and experts	.995

Firstly, the result of post hoc test between pharmaceutical care students and pharmaceutical science students shows that three items are viewed significantly different by the two groups at .014, .005, and .010 for using correct English pronunciation, modifying communication to meet special needs of patient, and using open-ended and close-ended questions appropriately respectively. Secondly, comparing pharmaceutical care students and pharmacy experts show that three items are in disagreement, which are using appropriate tone of voice, using correct English language, and using correct English pronunciation. Lastly, comparing pharmaceutical science students and pharmacy experts show that using correct English language is viewed differently by the two groups at .003.

Comparing two groups of students to the pharmacy experts, the pharmaceutical care students' view is significantly different from the pharmacy experts on three items, while the pharmaceutical science students' view differs on only one item, which is using correct English language. It is clear that the pharmacy experts do not pay as much attention as the pharmacy students on using correct English language, while the students in both fields strongly believe in this item. This may stem from their experience in dispensing drugs in English assuming that their English usage is not correct all the time when dispensing drugs to patients, but they can still offer the service to patients. This is item is a similar subject but more specific on the pronunciation, the result shows that Pharmaceutical care students value this item differently higher than the other two groups. This might be the result from the department they are in

to pay attention to the understanding of the patients as it is a part of service mind.

Concluding the encounter

Table 24: Response of concluding the encounter

		Experts	Students	Pharm care	Pharm Science
1. Summarize information	Mean	3.71	3.79	3.77	3.81
	SD	.463	.427	.456	.393
2. Ask if there is anything else	Mean	3.05	3.19	3.31	3.03
patients would like to discuss	SD	.590	.682	.671	.669
3. Invite patients to contact if	Mean	2.90	2.98	3.09	2.86
questions or concerns arise	SD	.768	.739	.775	.681
4. Thanks patients	Mean	3.57	2.97	3.14	2.76
	SD	.507	.809	.822	.751
5. End the conversation politely	Mean	3.52	3.38	3.49	3.25
	SD	.512	.640	.583	.685

Concluding the encounter composes of five items summarizing information, asking if there's anything else patients would like to discuss, inviting patients to contact if questions or concerns arise, thank patients, and ending conversation politely. The experts' response on concluding the encounter shows that one item among five is rated slightly under three, which is inviting patients to contact if questions or concerns arise. Apart from similar rating on the same item, the other item students perceived their significance to be lower than three is Thank patients. Receiving a mean score of more than two from both pharmacists and students, none of the item in this section is excluded from the rubric.

Table 25: Result of One-way ANOVA test on concluding the encounter

Variables	Mean Square	F	P-value
1. Summarize information	.081	.432	.650
2. Ask if there is anything else patients would like to discuss	1.432	3.29	.040

3. Invite patients to contact if questions or concerns arise	.841	1.542	.217
4. Thank patients	5.591	9.738	.000
5. End the conversation politely	1.045	2.747	.067

The result of ANOVA shows that two items are viewed differently by the one of the groups, which are asking if there is anything else patients would like to discuss at .040 and thank patients at .000. These two items are further studied in post hoc test analysis.

Table 26: Result of Games-Howell post hoc test on concluding the encounter

Variables	Variance	P-value
Ask if there is anything else patients would like to discuss	Pharm care and Pharm science students	.050
	Pharm care and experts	.198
	Pharm science students and experts	.996
Thank patients	Pharm care and Pharm science students	.019
	Pharm care and experts	.015
	Pharm science students and experts	.000

For the first post hoc item, the difference is shown between pharmaceutical care students and pharmaceutical science students at .050. The result is similar to previous item that pharmaceutical care students rate the items higher score than the pharmaceutical science students. For thank patients, the result reveals that the three groups view the item differently from the others as the importance is ranked into three levels. The highest mean this item received is from the pharmacy experts at 3.57 following by the pharmaceutical care students at 3.14 and pharmaceutical science students at 2.76. The high mean score of pharmacy experts may stem from the work experience, while the pharmaceutical care students score is higher than three but still significantly different from the pharmacy experts. This might due to the

fact that the students are not groom toward the service mind set, especially the pharmaceutical science students. In addition, the status of being a pharmacist dispensing drugs to patients might relay them a higher status as offering help and that saying thank you is not that necessary for them.

Non-verbal communication

Non-verbal communication consists of four items which are demonstrate appropriate eye contact, demonstrate appropriate posture and body language, wear appropriate attire, display appropriate health-professional manner.

Table 27: Response on non-verbal communication

		Experts	Students	Pharm care	Pharm Science
1. Demonstrate appropriate eye contact	Mean	3.24	3.20	3.31	3.07
	SD	.539	.654	.603	.691
2. Demonstrate appropriate posture and body language	Mean	3.38	3.41	3.47	3.34
	SD	.590	.607	.557	.659
3. Wear appropriate attire	Mean	3.43	3.48	3.56	3.39
	SD	.507	.546	.500	.588
4. Display appropriate health-professional manner	Mean	3.62	3.61	3.67	3.54
	SD	.498	.520	.473	.567

For both the pharmacists and the students, the mean of every item is rated higher than three, which means that they are all considered highly important while the highest value in this section is displaying appropriate health-professional manner. This is probably the word ‘manner’ can be referred to the overall behavior, which includes eye contact, posture, body language and attire. The result of ANOVA shows that the groups view the item in the same way, so none of the item in this section is further analyzed in post hoc test analysis.

Table 28: Result of One-way ANOVA test on concluding the encounter

Variables	Mean Square	F	P-value
1. Demonstrate appropriate eye contact	.985	2.469	.088
2. Demonstrate appropriate posture and body language	.289	.792	.455
3. Wear appropriate attire	.473	1.637	.198
4. Display appropriate health-professional manner	.267	1.005	.368

Eliciting information from patients

Eliciting information from patients contains five items, which are asking patients about their concerns or reasons for visit, giving patients opportunity and time to talk, asking for a complete record of patient's current health conditions and therapies, asking questions to assess patients' understanding of key information about medications, asking questions to assess patient's experience with medications currently being taken.

Table 29: Response on eliciting information from patients

		Experts	Students	Pharm care	Pharm Science
1. Ask patients about their concerns or reasons for visit.	Mean	3.57	3.76	3.83	3.68
	SD	.676	.480	.450	.507
2. Give patients opportunity and time to talk.	Mean	3.76	3.78	3.89	3.64
	SD	.539	.419	.320	.483
3. Ask for a complete record of patient's current health conditions and therapies.	Mean	3.90	3.87	3.94	3.78
	SD	.301	.340	.234	.418
4. Ask questions to assess patients' understanding of key information about medications.	Mean	3.62	3.63	3.79	3.44
	SD	.498	.501	.413	.534
5. Ask questions to assess patient's experience with medications currently being taken.	Mean	3.43	3.55	3.70	3.37
	SD	.598	.558	.492	.584

From the pharmacists' result, the means of these items are 3.57, 3.76, 3.90, 3.62 and 3.63 respectively. For the pharmacy students, the means are 3.76, 3.78, 3.87, 3.63 and 3.55. All of the items are rated as highly important by the two main groups. The result of ANOVA shows that the groups agree on one item, which is asking patients about their concerns or reasons for visit at .074 and differ on the rest of the items.

Table 30: Result of One-way ANOVA test on eliciting information from patients

Variables	Mean Square	F	P-value
1. Ask patients about their concerns or reasons for visit.	.683	2.645	.074
2. Give patients opportunity and time to talk.	.936	5.210	.007
3. Ask for a complete record of patient's current health conditions and therapies.	.438	4.100	.018
4. Ask questions to assess patients' understanding of key information about medications.	1.906	8.421	.000
5. Ask questions to assess patient's experience with medications currently being taken.	1.847	6.221	.003

The result of post hoc test between pharmaceutical care students and pharmaceutical science students shows that all four items are viewed differently by the two groups. However, no other disagreement was found on any items between two pairs. From the post hoc test result, Pharmaceutical care students have the highest mean score on all of the items except the last item following by pharmaceutical science students as the second-high rank and the pharmacy experts as the last. The rank is changed in the last item as the pharmacy experts' score are higher than the pharmaceutical science students, while the highest is still pharmaceutical care students. The different result points out that pharmaceutical care students tended to rate all items in this section higher than pharmaceutical science students and pharmacy experts.

Table 31: Result of Games-Howell Post hoc test on eliciting information from patients

Variables	Variance	P-value
Give patients opportunity and time to talk.	Pharm care and Pharm science students	.004
	Pharm care and experts	.583
	Pharm science students and experts	.654
Ask for a complete record of patient's current health conditions and therapies.	Pharm care and Pharm science students	.024
	Pharm care and experts	.855
	Pharm science students and experts	.315
Ask questions to assess patients' understanding of key information about medications.	Pharm care and Pharm science students	.000
	Pharm care and experts	.356
	Pharm science students and experts	.360
Ask questions to assess patient's experience with medications currently being taken.	Pharm care and Pharm science students	.003
	Pharm care and experts	.928
	Pharm science students and experts	.158

Initiating educational interventions

Initiating educational interventions consists of seven items: emphasizing key information, providing reasons for advice, providing appropriate recommendation based on Indication, Efficacy, Adherence, Safety, and Cost-effectiveness, discussing one drug or therapeutic regimen at a time, providing complete and clear instruction on medication, verifying patient understanding of new information provided, working with patients to schedule the doses. For pharmacy experts, the means of the items are 3.86, 3.43, 3.10, 3.67, 3.71, 3.24, and 2.67 respectively. For the students' response, the means are 3.71, 3.36, 3.50, 3.70, 3.88, 3.59, and 2.85 respectively.

Based on ANOVA, the result shows that two items are in agreement among groups, which are discussing one drug or therapeutic regimen at a time, and providing complete and clear instruction on medication. The rest of the items is further studied with post hoc test analysis.

Table 32: Response on initiating educational interventions

		Experts	Students	Pharm care	Pharm Science
1. Emphasize key information.	Mean	3.86	3.71	3.83	3.56
	SD	.359	.491	.380	.565
2. Provide reasons for advice.	Mean	3.43	3.36	3.53	3.15
	SD	.598	.705	.607	.761
3. Provide appropriate recommendation based on Indication, Efficacy, Adherence, Safety, and Cost-effectiveness.	Mean	3.10	3.50	3.63	3.36
	SD	.768	.627	.543	.689
4. Discuss one drug or therapeutic regimen at a time.	Mean	3.67	3.70	3.76	3.63
	SD	.577	.494	.464	.522
5. Provide complete and clear instruction on medication.	Mean	3.71	3.88	3.89	3.86
	SD	.463	.331	.320	.345
6. Verify patient understanding of new information provided.	Mean	3.24	3.59	3.69	3.47
	SD	.768	.568	.468	.653
7. Work with patients to schedule the doses.	Mean	2.67	2.85	3.09	2.58
	SD	.658	.708	.654	.675

First of all, the result of Games-Howell test between pharmaceutical care students and pharmaceutical science students shows that four items are viewed differently by the two groups, which are emphasizing key information at .007, providing reasons for advice at .008, providing appropriate recommendation based on Indication, Efficacy, Adherence, Safety, and Cost-effectiveness at .040, and working with patients to schedule doses at .000. Second, comparing pharmaceutical care students and pharmacy experts show that three items are in disagreement, which are providing appropriate recommendation based on Indication, Efficacy, Adherence, Safety, and Cost-effectiveness at .017, verifying patient understanding of new information provided at .046, and working with patients to schedule the doses at .039. Last, comparing pharmaceutical science students and pharmacy experts shows that

two items are in disagreement, which are emphasizing key information at .020, and providing reasons for advice at .008.

Table 33: Result of One-way ANOVA test on educational interventions

Variables	Mean Square	F	P-value
1. Emphasize key information.	1.368	6.477	.002
2. Provide reasons for advice.	2.310	5.129	.007
3. Provide appropriate recommendation based on Indication, Efficacy, Adherence, Safety, and Cost-effectiveness.	2.698	6.645	.002
4. Discuss one drug or therapeutic regimen at a time.	.279	1.100	.336
5. Provide complete and clear instruction on medication.	.243	1.956	.145
6. Verify patient understanding of new information provided.	1.826	5.203	.007
7. Work with patients to schedule the doses.	4.467	10.171	.000

For emphasizing key information and providing reasons for advice, pharmaceutical science students view the items differently from the others as in significantly lower in the mean score for both items, while pharmaceutical care students agree with pharmacy experts. The next two items, which are providing appropriate recommendation based on Indication, Efficacy, Adherence, Safety and Cost-effectiveness, and working with patients to schedule the doses reveal that pharmaceutical care students think differently from pharmaceutical science students and pharmacy experts even though no difference can be pointed out when comparing the whole student population with pharmacy experts.

Finally, verifying patient understanding of new information provided indicates a scaling result. Although the whole student population seems to disagree with pharmacy experts on this item, the result shows that pharmaceutical science students differ from the experts but not different from pharmaceutical care students. This means that the level of agreement is different as in a scaling system. In other words, pharmaceutical care students have the highest mean score at. 3.69 following by pharmaceutical science

students as the second-high ranking at 3.47 and the pharmacy experts as the last at 3.24.

Table 34: Result of Games-Howell Post hoc test on verbal communication

Variables	Variance	P-value
Emphasize key information.	Pharm care and Pharm science students	.007
	Pharm care and experts	.947
	Pharm science students and experts	.020
Provide reasons for advice.	Pharm care and Pharm science students	.008
	Pharm care and experts	.782
	Pharm science students and experts	.008
Provide appropriate recommendation based on Indication, Efficacy, Adherence, Safety, and Cost-effectiveness.	Pharm care and Pharm science students	.040
	Pharm care and experts	.017
	Pharm science students and experts	.368
Verify patient understanding of new information provided.	Pharm care and Pharm science students	.100
	Pharm care and experts	.046
	Pharm science students and experts	.429
Work with patients to schedule the doses.	Pharm care and Pharm science students	.000
	Pharm care and experts	.039
	Pharm science students and experts	.854

Objective number two

To create the rubric for measuring English oral communication competency of Thai pharmaceutical science students.

After the interviews with informant specialists and the students to obtain SPEAKING grid and the questionnaire with the target participants, the rubric was developed based on the case report of training pharmacy students with the detail following the criteria to be assessed. The rubric development was conducted through three steps, selecting the criteria, dividing into three categories, and assigning

the scores. Three sections were determined according to the three components of ESP (Douglas, 2000). The criteria were selected from the questionnaire comparing to Thai dispensing rubric and divided into three categories, which are pharmaceutical science skills, language use, and strategic competence

A total of 17 criteria were selected into the rubric. 9 criteria were selected for pharmaceutical science skills, composes of *patient awareness, allergy, underlying disease, impression, dispensing, reason(s) for dispensing, instruction, caution*, and *verifying understanding*. Based on the questionnaire, language use is composed of 4 criteria, which are *grammar, pronunciation, question type*, and *word choice*. Strategic competence is composed of 4 criteria: voice, initiating communication, concluding the encounter, and non-verbal communication. The scores were designed to be heavily relying on the specialty of the language instructors at 40% on language use following by strategic competence at 30% and strategic competence at 20%.

The number of assessment criteria represents the diagnostic focus of the assessment, which was used in a formative manner as part of the English language course the students attended. As can be seen in Table 34 below, the criteria were grouped into the three sections while the scores assigned into dichotomous system and partial credit scores. According to the informant specialists, the students need to perform these criteria by eliciting information and initiating educational interventions using their pharmaceutical skills they have learned. The differences in scores of pharmaceutical science skills criteria were decided by the informant specialist in referral to Thai dispensing skills that criteria with scores of 2 require more critical thinking from the pharmacy students, which are *patient awareness, dispensing, reason for dispensing, instruction,* and *caution.*

The scores on language use are partial credit scores that are equal to all of them at 1 to 4. *Grammar* and *pronunciation* are judged on the level of intelligibility, while *question type* counts on the questions the students used whether the open-ended or close-ended questions are appropriate to each situation, since misleading questions may influence the patients to the wrong impression or symptom. *Word choice* is assessed by the use of lay-man terms with patients. This criterion attempts

to test whether the students can communicate effectively using easy-to-understand language for patients.

Table 35: Criteria and score range

Criteria	Score range
I: Pharmaceutical science skills	15 (30%)
- Patient awareness	2-0
- Allergy	1-0
- Underlying disease	1-0
- Impression	1-0
- Dispensing	2-0
- Reason for dispensing	2-0
- Instruction	2-0
- Caution	2-0
- Verifying understanding	2-0
II: Language use	20 (40%)
- Grammar	2-3-4-5
- Pronunciation	2-3-4-5
- Question type	2-3-4-5
- Word choice	2-3-4-5
III: Strategic competence	15 (30%)
- Voice	.5-1-1.5-2
- Initiate communication	1-2-3-4
- Conclude the encounter	2-3-4-5
- Non-verbal communication	1-2-3-4

Different partial credit scores were given to strategic competence due to the different components to be judged in each criterion. The tone, volume, pace, and silence were categorized into *voice*. In the questionnaire, greeting and identifying patient's identity were separated, but were grouped into one criterion as *initiating communication*. *Concluding the encounter* was titled for summarizing information,

asking if any questions arise, and thank patients. Lastly, the eye contact, gestures, posture, and professional manner were grouped under *non-verbal communication*.

Objective number one

To create and validate tasks for measuring English oral communication competency of Thai pharmaceutical science students.

After the rubric was developed, the cases were developed with the information following the criteria in the rubric. The informant specialists provided a set of clinical reports that the internship students recorded to fulfill their training requirement, which contained about 40 cases. The researcher read through and selected 8 cases with the similar symptoms or purpose of visiting, which are muscle pain, birth control pills, stomachache, and headache. The researcher checked that the information in the report is enough for the raters to use role-play with the students according to the criteria listed. For example, slight information on the allergy was added to increase variety to the response.

Since dispensing tasks are not the researcher's specialty, the selected cases with detail were consulted with informant specialists several times to include the detail needed and the detail that might be asked by the students. In addition, the cases were trialed with intern pharmacy students who are not the target user as they are from different university. The process of seeking opinions not only from informant specialists but also the students is to validate that the detail in the tasks are sufficient for the raters to role-play when assessing pharmacy students' dispensing skills.

Objective number two

To validate the rubric for measuring English oral communication competency of Thai pharmaceutical science students.

Participants' information

Two groups of participants are students and raters. The rubric was used by six raters, all of which except one is a language instructor to students who took the test. As a part of a classroom assessment, 147 students from five sections

took the test with the rater who was not their classroom instructor. Their performance was recorded for the second rater. However, due to the technical problem, five records of students' performance could not be found. The students were given an option of retaking the exam for the second rater or use the score from one rater. One student opted to take another exam while the other four students chose to use the score from one rater. This means that from a total of 147 students 143 students are scored by two raters, 4 students by one rater, and 1 student performing two times rated by different raters.

Rubric information

The rubric with the full point of fifty consists of three main sections: pharmaceutical science knowledge, language knowledge, and strategic competence at fifteen, twenty, and fifteen points respectively. Pharmaceutical science knowledge comprises of nine criteria, which is dichotomous at two systems: one or zero and two or zero. The language knowledge contains four criteria using four scales from two to five. The strategic competence consists of four criteria four criteria with three types of scale: one dichotomous criterion at either one or two, two criteria with four scales from one to four, and one criterion four scales from two to five.

Many-Facet Rasch Measurement (MFRM)

The score of the students are analyzed using MFRM, which results in Wright map, candidate measurement report, rater measurement report, criteria measurement report, and rating scale category functioning. Three facets are analyzed in this model: students as candidates, raters, and criteria from the rubric.

Wright map

The Wright map in figure 4.1 presents the data in this study relatively to the degree of the rating. The first column, 'Measr', shows the logit scale of all candidates, raters, and criteria. The second column labeled 'Candidate' shows the number of the students on the scale using stars and dots. A star represents two candidates while a dot equals one candidate. The candidates

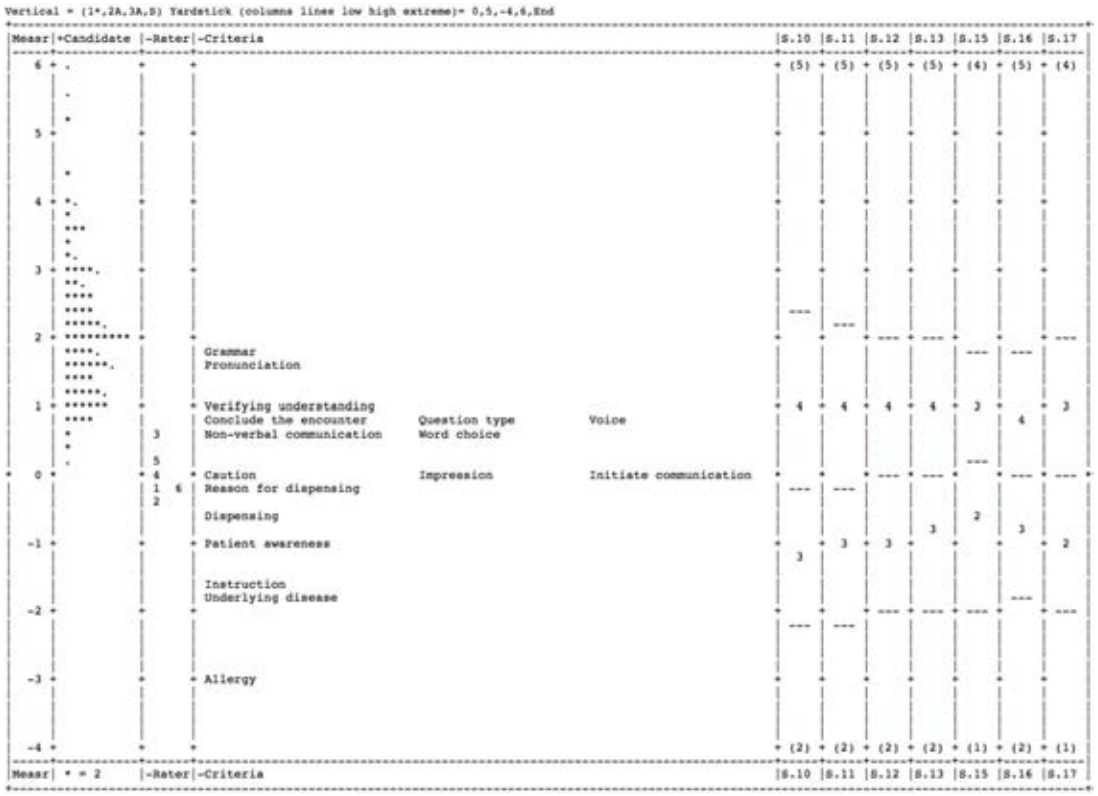
are placed along the scale according to their score. That is the higher position they are the higher the score they receive. This Wright map shows that the candidates are positively oriented. None of the candidates are below the zero logits, which is the risk of receiving less than 50% score from the average severe raters on the criterion of average difficulty (T. McNamara, Knoch, U. & Fan, J., 2019).

The third column, 'Rater', indicates the position of their judgment as the higher the scale, the harsher the rater is. In the figure, it can be seen that rater number 3 is the highest which means the harshest while rater number two is the lowest, which means the most lenient. The raters are homogenous considering that the raters are within the two logit scales apart. Similarly, the fourth column labeled 'Criteria' shows the position of how hard each criterion is rated by the raters with the highest as the hardest and the lowest as the easiest. In the figure, Grammar is the highest, which means that it is the harshest criterion for the candidate to achieve, while allergy is at the lowest meaning that it is the most lenient criterion for the students to receive the score. It can be seen that the group of harsh criteria are partial credit scale while the lenient scale criteria are dichotomous, which position at zero or below. This result complies with what Eckes (2009) found that the harsher the criterion is, the more variable it measures.

The wright map includes partial credit scale but not the dichotomous scale. The last seven columns indicate the level of the score given to candidates in seven partial credit scales as the lowest are either one or two and the highest as either four or five. From the figure, the highest interval score is wider than the other intervals at three to four logit scales, while the second levels are about 2 digits. S.10-13 refers to the four partial credit models for Language use, which are Grammar, Pronunciation, Question type, and Word choice respectively. S.15-17 are applied to partial credit scale in the strategic competence section which are Initiate communication, Conclude the encounter, and Non-verbal communication respectively. The average interval of each score point is about two except the highest level where the digits are

higher than three. Each criterion has its own scale of level that the candidates need to perform to fit each level. For S.15 criterion or Initiate communication, the Wright map shows that its digits of level 3 is the smallest scale comparing to other criteria. This means that for this criterion, it is not so hard for the candidates to achieve level 4 when the scale of level 3 is slightly narrow.

Figure 5: Wright map - Speaking scale



Candidate measurement report

Candidate ability

The candidate ability is shown in figure 4.2, which is a part of the full table that can be found in Appendix. The first column is the total score each candidate receives from the rater. This information needs to be studied together with the second column, total count. As an illustration, the first candidate receives a total score of 99 from two raters, which has total count of 34 as each rater judges their performance on 17 criteria. The last candidate in figure 4.2 receives a total score of 48 with a total count of 17, which means that the candidate was rated by one rater.

Table 36: Candidate measurement report (extract)

Total score	Total count	Obsvd Average	Fair (M) Average	Measure	S.E.
99	34	2.91	2.93	6.00	1.05
99	34	2.91	2.93	5.69	1.05
99	34	2.91	2.93	5.27	1.05
99	34	2.91	2.93	5.27	1.05
98	34	2.88	2.91	4.48	.77
97	34	2.85	2.91	4.39	.66
96	34	2.85	2.90	3.98	.66
97	34	2.85	2.90	3.98	.66
97	34	2.85	2.90	3.98	.66
48	17	2.82	2.89	4.54	.83

The third column, Obsvd Average, is the average rating score across 17 criteria. The number is obtained from the total score divide by total count. The fourth column labeled Fair (M) Average is the fair average score MFRM calculated to indicate the score the candidate would receive if being assessed by average raters. The last two columns state the measurement, which is the position of the candidate on the logit scale, and the standard error, which is the precision of the measurement. The result of the measurement and the standard error is high and the claim can be hard to state.

Candidate fit

The first columns show the infit score of the students. Both infit and outfit score can be referred to the variation of the expected response. The mean score level from .5-1.7 follows the suggestion used on clinical observation since the test is to assess the pharmacist's performance in dispensing drugs. However, the variation under .50 is not considered since that means less variation to the group. The result shows that the infit score of 13 students out of 147 are more than 1.70, which means more variation. The next step is to look at the normalized distribution of the infit, which is shown as Zstd in the column after the infit. The Zstd of the misfit group is furthered considered and found that 6 students out of 13 students are significantly less

compatible to the expected model response at over 2. This result means that the variation of this small group is too much, unpredictable and underfit. The figure below is the sample data. A complete report of the data can be found in the Appendix section.

Table 37: Fit statistics from candidate measurement report (extract)

Infit MnSq	ZStd	Estim. Discrm	Correlation		Num
			PtMea	PtExp	
1.11	.4	.99	.16	.17	526
1.21	.5	.77	.05	.17	419
1.15	.4	.86	.11	.17	213
.92	.2	1.09	.22	.17	214
.96	.1	1.07	.26	.23	231
1.44	.9	.65	.17	.28	427
.83	-.1	1.10	.34	.28	211
.72	-.4	1.20	.37	.28	216
1.13	.4	.92	.28	.28	226
.38	-.9	.59	.18	.28	418

Candidate summary statistics

The summary provides statistical information on the separation index and the fixed Chi-square, which shows that at least two candidates are significantly different. This means that the test can differentiate the students' capability. What important in this part is candidate separation index or the label strata in the column. The candidate separation index at 3.46 means the number of distinguishable group found in this group.

Table 38: Summary statistics from candidate measurement report

		Population	Sample	Fixed
RMSE		.45	.45	
	SD	.99	.99	
	Separation	2.34	2.35	
	Strata	3.45	3.46	
Chi-square				920.7
	d.f.			146
	Significance			.00

Rater measurement report

Rater fit and ability

The first two columns after the raters' column, which are total score and total count, refer to the score and the counting of the score each rater gave to the candidates. It can be seen that rater 1 and 2 has more total score and total count than the other raters because they rated a larger number of candidate. The third column is the observed average score that each rater gave to students on each criterion. The fourth column is the fair average score the rater will give if they encounter the average candidate. The fair average can be used to compare the score among the raters.

Table 39: Rater measurement report (extract)

Raters	Total score	Total count	Obsvd Ave	Fair (M) Ave	Measure	S.E.
3	1862	850	2.19	2.59	.56	.06
5	2104	867	2.43	2.66	.27	.06
4	2204	884	2.49	2.72	-.06	.07
6	797	323	2.47	2.74	-.16	.11
1	2721	1037	2.62	2.76	-.30	.07
2	2820	1071	2.63	2.76	-.31	.07

The last two columns, measure and standard error, assign the position of the rater on the logit scale, and the standard error, which is the precision of the measurement. It can be seen that the logit scale difference between the most severe rater at .56 and the most lenient rater at -.31 is .87, which is considered low meaning that the ability in assessing the test takers using this rubric of the raters in this group tends to conform with the group standard. The standard error is rather low with the highest on rater 6 at .11.

The Rater fit

The result indicates that the infit score of rater 4 is higher than the whole group at 1.38 and the normalized distribution, which is shown as Zstd is less compatible to the expected model response at over 2. The two columns before last shows the correlation result among the rater, which is

similar to what the Wright map shows that the raters tend to have similar opinion to the whole group.

Table 40: Fit statistics from rater measurement report (extract)

Raters	Infit MnSq	ZStd	Estim. Discrm	Correlation		Upper CI	Lower CI
				PtMea	PtExp		
3	.76	-4.8	1.26	.58	.57	0.68	0.44
5	.86	-2.3	1.05	.52	.52	0.39	0.15
1	1.01	.1	.96	.45	.48	-0.16	-0.2
2	1.09	1.3	1.02	.47	.47	-0.17	-0.38
6	1.19	1.6	.86	.50	.52	0.06	-0.44
4	1.38	5.1	.77	.54	.52	0.08	-0.45

Another way to prove the certainty of the rater measurement is to calculate confidence interval (CI) based on standard error and severity logits. The CI width, which stands for the degree of uncertainty, in this data set indicated that rater 6 is less reliable at 0.44 than the rest of the group which was at the same width for rater 3 and 5 at 0.24 while rater 1, 2, and 4 CI width differed slightly at 0.28.

Rater summary statistics

The summary shows that at least two raters are significantly different. The rater candidate separation index or the label strata in the column at 6.42 means the number of distinguishable groups of the raters found in this data.

Table 41: Summary statistics from rater

		Population	Sample	Fixed
RMSE		.07	.07	
	SD	.31	.34	
	Separation	4.15	4.57	
	Strata	5.87	6.42	
Chi-square				150.4
	d.f.			5
	Significance			.00

Criteria measurement report

Criterion difficulty

Similar to the result in the Wright map presented above, this section provides information about the position of the criterion relative to its difficulty. From the result, 'Grammar' and 'Pronunciation' are the hardest criteria for the test takers to achieve following by Verify understanding, Question type, Voice, Conclude the encounter, Word choice, Non-verbal communication, Caution, Initiate communication, Impression, Reason for dispensing, Dispensing, Patient awareness, Instruction, Underlying disease, and Allergy.

Table 42: Fit statistics from criterion measurement report (extract)

Infit MnSq	ZStd	Estim. Discrm	Correlation		Num
			PtMea	PtExp	
.91	-1.1	1.12	.64	.60	10 Grammar
.85	-1.9	1.18	.67	.59	11 Pronunciation
1.82	7.5	-.28	.19	.50	9 Verifying understanding
.80	-2.7	1.23	.65	.57	12 Question type
.79	-3.3	1.39	.54	.38	14 Voice
1.00	.0	.99	.56	.56	16 Conclude the encounter
.76	-3.2	1.28	.66	.57	13 Word choice
.80	-2.7	1.22	.64	.56	17 Non-verbal communication
1.34	1.9	.57	.16	.31	8 Caution
.90	-1.1	1.05	.53	.51	15 Initiate communication
1.11	1.0	.87	.20	.30	3 Impression
1.12	.6	.92	.24	.28	6 Reason for dispensing
1.06	.2	.98	.17	.18	5 Dispensing
1.13	.4	.87	.08	.14	4 Patient awareness
1.00	.3	1.03	.09	.07	7 Instruction
1.02	.1	1.00	.15	.14	2 Underlying disease
1.01	.2	.99	.07	.08	1 Allergy

Criterion fit

In order to gain quality for the criterion, a different range of infit in this section is between 0.7-1.3. From the infit result, verifying understanding is misfitting at 1.82 and caution at 1.34. When considering the mean square standard of the two items, verifying understanding is found to be significantly misfit.

Criterion separation index

The summary shows that at least two criteria are significantly different. The criterion separation index, or the label strata in the column at 7.33 for sample group means the number of distinguishable levels of criterion found in this data.

Table 43: Summary statistics from criterion measurement report

		Population	Sample	Fixed
RMSE		.24	.24	
	SD	1.22	1.26	
	Separation	5.09	5.25	
	Strata	7.12	7.33	
Chi-square				628.5
	d.f.			16
	Significance			.00

Rating scale category functioning

The function of this rating scale needs to be divided into two parts according to the method of the rating, which are dichotomous and partial credit scale. In this rubric, two sets of dichotomous scales are used and three sets of partial credit scales are applied. The different result between the two is mainly the report on Rasch-Andrich threshold, which is the information on the step difficulty on partial credit scale.

Table 44: Patient Awareness Rating

Data			Quality control			Response
Score	Total	%	Avg Meas	Exp Measr	Mnsq	Category
0	4	1	2.26	1.56	1.3	no
2	292	99	3.04	3.05	1.1	yes

For the first part of the rubric, the pharmaceutical knowledge is assessed with dichotomous scale at two systems: 1 and 0 or 2 and 0. All criteria have their own table. Examples of the dichotomous criteria are shown below while the rest of the criteria can be found in the Appendix.

Table 45: Allergy rating

Data			Quality control			Response
Score	Total	%	Avg Meas	Exp Measr	Mnsq	Category
0	3	1	4.30	4.15	.9	no
1	293	99	5.13	5.13	1.0	yes

The partial credit scale is applied to language knowledge composing of four criteria rating at four levels from 2 to 5 as in need, fair, good and excellent. The criteria are grammar, pronunciation, question type, and word choice. The table below indicates the average measure and the Rasch-Andrich Thresholds. The average measures of all criteria in this section are advanced properly. The Rach-Andrich Thresholds, which is the information to consider for step calibration of each score level, are found to be increased properly for all of the criterion, except pronunciation.

As stated, the step calibration should be at least 1.4 and not over 5. The result can be found in the Rasch-Andrich Threshholds section. The lowest score the students can receive is 2 and the higher is 5 meaning that there are three steps among these score levels. As an illustration, the score stated at level 3 means the step calibration from level 2 to 3. A problem is found in

pronunciation criterion as the difference between the step calibration of score level 2 to 3 at -1.71 and 3 to 4 at -.39 is only 1.32.

Table 46: Language use rating

Score		2	3	4	5
Grammar	Average Measure	-.73	-.47	.50	1.81
	Mean square	1.1	.80	.8	.9
	Rasch-Andrich Thresholds		-1.81	-.36	2.17
Pronunciation	Average Measure	-.70	-.31	.56	1.98
	Mean square	1.0	.9	.8	.8
	Rasch-Andrich Thresholds		-1.71	-.39	2.11
Question type	Average Measure	-.39	.06	1.25	2.12
	Mean square	.7	.5	1.0	1.0
	Rasch-Andrich Thresholds		-1.67	-.12	1.79
Word choice	Average Measure	-.11	.28	1.27	2.30
	Mean square	.8	.7	.8	.8
	Rasch-Andrich Thresholds		-1.80	.20	1.60

The strategic competence composes of two rating scales. The first criterion, voice, originally uses an partial credit scale of four as in .5, 1, 1.5, and 2. It was adapted to dichotomous scale, which is a score of one and two. To adapt to the MFRM program, the score was round up, which means that any person who received .5 would be round up to 1 while 1.5 was round up to 2.

Table 47: Voice rating

Data			Quality control			Response
Score	Total	%	Avg Meas	Exp Measr	Mnsq	Category
1	80	27	.22	.54	.6	Fair
2	216	73	1.67	1.55	.8	Good

The next criterion is initiating communication which has the same scale as the non-verbal communication. They compose of partial credit scale of 1 to 4. A problem was found in the step calibration from score 3 to 4 for initiating

communication as the step calibration from score level 2 to 3 at .76 and score level 3 to 4 at 1.05 is only 0.29.

Table 48: Initiating communication and non-verbal communication rating

Score		1	2	3	4
Initiating	Average Measure	.25	.92	1.66	2.55
communication	Mean square	.7	1.1	1.6	1.0
	Rasch-Andrich Thresholds		-1.80	.76	1.05
Non-verbal	Average Measure	-.11	.36	1.29	2.36
communication	Mean square	.9	.6	1.3	.8
	Rasch-Andrich Thresholds		-1.79	.10	1.69

The partial credit scale of 2 to 5 is applied to concluding the encounter. The step calibration of each score level, is found to be increased properly for this criterion, which is more than 1.4 logits apart between each score level. In addition, the unidimensional model of this study was proved by the percentage of standardized residual outside 2, which is below 5 % at 3.46.

Table 49: Concluding encounter rating

Score		2	3	4	5
Concluding encounter	Average Measure	.26	.34	1.22	2.10
	Mean square	.9	.6	1.3	.8
	Rasch-Andrich Thresholds		-1.50	.04	1.46

Objective number two

To validate the rubric for measuring English oral communication competency of Thai pharmaceutical science students.

Participants' information

The raters who used the rubric are the subject of this part of the study. As stated, the rubric was used by six raters, all of which except one is a language instructor to students who took the test. Four raters out of six could be reached for the interview. The raters were asked whether they agree to be interviewed and recorded. The interview questions were sent to the raters for them to study

before the interview session was conducted individually on the phone. One of the four raters taught the course one time and the assessing the dispensing task was her first time, but the rater has extensive experience in teaching and rating other medical performance tests, such as English for medical professionals. The rest of the raters has been teaching the course and rated students on dispensing tasks for more than two semesters. The interview with the language instructors was described by numbering them according to the order of the interview as R1 to R4. The numbering does not relate to the number of the rater in the MFRM section.

Structured interview questions

The rubric contains a total of three parts. The main questions were established for all three parts including the overall rubric usage. The process of asking started with the overall rubric usage to obtain the general idea of the raters on the rubric. It was followed by the Pharmaceutical science knowledge, Language use, and Strategic competence, which is similar to the format the criteria are listed in the rubric. The question set for the overall rubric usage is slightly different than the other parts as listed below:

- Did you feel comfortable using the rubric?
- Which section was the easiest to rate?
- Which section was the hardest to rate?
- Does the rubric specify adequate information for the rating? If not, what should be added in?
- Can this rubric be used to assess and differentiate the student's performance in dispensing drugs in English?
- Are there any criteria you think should be added in the rubric?
- Are there any criteria you think it's unnecessary?

Most of the questions for the other three parts are similar except for Pharmaceutical science knowledge, which focuses on the confidence rating each criterion specifically and asking the raters for their opinion on the score weight given to each criterion. The rest of the questions are the same as language use and strategic competence as listed below:

- Were you confident rating the students in this section?
- Do you think the scale for each criterion is appropriate?
- Are there any criteria you found it hard to rate the student?
- Any recommendation?

Overall rubric usage

Comfort

Did you feel comfortable using the rubric?

R1: Yes, I felt pretty comfortable once you've done a couple as well. It's quite straight forward making sure that you cover everything on the rubric. I didn't think there's any major problem.

R2: I'm a new rater to this test and my first time using it. I didn't feel that much comfortable. I have to do the role play and at the same time assess the test taker, especially when the test takers are not students from your section. The new audience might not be accustomed to my accent and the way we speak. The test takers might be nervous as well. I was wondering whether I could cover all of the criterion. Also, I was concerned about the time constraint that I had to finish assessing 30 students within a specific amount of time with the students pressing in line as they want to get it done. As a matter of fact, I didn't have enough time to write down some detail that I wanted to. I think it's because I'm new to this test and also very slow.

R3: Yes. I think it's clear since the section is divided in an partial credit way.

R4: Comfortable.

Easiness

Which section was the easiest to rate?

R1: Probably the first section, pharmaceutical science knowledge.

R2: The language use except for the word choice. The second easiest is the strategic competence, which is something I'm comfortable with.

R3: The first section because I don't have knowledge on this. What I did is to find out whether the test taker perform the following criteria or not. There's no shading area I have to decide like the other two sections that I'm used to.

R4: Pharmaceutical knowledge because it requires me to choose only yes or no. Also, the test takers have been through similar test in Thai so it's not so hard for them to follow the criterion in this section of the rubric.

Hardness

Which section was the hardest to rate?

R1: The strategic competence since it's slightly more subjective scale rather than what they said this thing or no they didn't. So that's a bit harder to rate within than five-minute role play. But as I said because we did more than 20, you do get a good indication of fairly quickly

R2: The hardest for me is pharmaceutical science knowledge even though the score weight is not that much at 15 points. That still slowed me down when assessing the whole section. I'm really slow at checking various things within the limited time. Some of the students complained that the rater they did role play with told them all the information, which left them wondered how to perform the task what questions they can ask to attain the score. As we believed that when going to see a doctor or a pharmacist, we usually present them with one obvious symptom and the rest is left for the doctor or the pharmacist to ask, while some of the patient can be panic and tell the doctor everything. This is the point on the raters, who acted as a patient, whether it is okay to state all the information on the card or state the disease they think they have. That means the test taker does not need to fulfill the impression. Stating all the information right away seems like guiding to the test taker. That's why I think this part is hard.

R3: Language use.

R4: Strategic competence. It seems that some topic has been covered in pharmaceutical science knowledge section. So, when the test taker didn't perform a particular part, it doesn't seem fair to mark them down.

Adequacy

Does the rubric specify adequate information for the rating? If not, what should be added in?

R1: Definitely.

R2: I think mainly it's enough. To add anything in, we have to concern the time constraint. Not to mention the nervousness on both sides since the raters have to study the cases as well. The preparation for the rater is needed that the rater knows how to answer the response in a planned way. Having to act and assess at the same time, I was worried whether I can be fair to them. How much information can I provide to the test takers? To me, it's like I have to think over both as a patient and as a rater. Normally, when I do this kind of test, I can be an observer assessing them as the students act out as a patient and a doctor. I don't have to think whether I did something wrong or unfair to the test takers since they are the one who takes that responsibility.

R3: Of course.

R4: I think it is good enough.

Ability to differentiate

Can this rubric be used to assess and differentiate the student's performance in dispensing drugs in English?

R1: I think so. The only issue maybe that there are only 8 situations that they were informed about before. I think if you make it completely unprepared unrehearsed, you'll get a much stronger indication because the students would have no idea what the role play may consist of. They just have to react there and then. That would be my only suggestion to that.

R2: I think it does. All the criteria listed help in the dispensing performance. For example, if they can employ the question type appropriately, they can come up with sufficient information. It's the question whether they can encourage the patient to give more information. We can't judge whether they

can dispense a correct medication but we can for the dispensing skills in English.

R3: I think so, but it would rather focus on the language ability since the test taker only need to perform the pharmaceutical knowledge and not being graded on how well they did it.

R4: Yes.

Added criteria

Are there any criteria you think should be added in the rubric?

R1: From a language point of view, I don't think so. You might need to talk to the pharmacy specialists for the first section. For the language use and the strategic competence, I think the rubric covers the areas I would expect.

R2: I'm not sure whether the cases are complicated at the same level since I'm new to this assessment. Is it possible to add overall impression or general impression? If possible, I want to add this, especially for the raters who did the role play. This criterion is a part of the test for medical students when they perform as a patient and a doctor.

R3: No. This is a lot already.

R4: I think the criteria is enough. It's the question on the score weight. Even though we don't expect them to use an exact correct English, we cannot really differentiate the intermediate students from the advanced level when looking at the scale.

Unnecessary criteria

Are there any criteria you think it's unnecessary?

R1: I don't think so. There's nothing stood out as unnecessary. I don't think there was any problem there.

R2: No.

R3: The language use and strategic competence might be able to merge as one section as language and language use. For example, language can focus on the linguistic ability when the language use can be the voice and the tone.

R4: Verifying understanding doesn't seem natural for the pharmacist would do. When I go to the drug store, the pharmacist never asks me to repeat the instruction.

Pharmaceutical science knowledge

Confidence

Were you confident rating the students in this section?

R1: Yes, for the most part because the criteria were quite clearly labeled. So even if I don't have the subject knowledge, I was still able to assess them based on the criteria. I would say 8 out of 10.

R2: As I said, I'm not confident rating this section because it's absolute. I wasn't sure about asking how much is enough.

R3: Very confident. But I don't know whether the test taker dispensed the appropriate drugs or give the wrong instruction.

R4: Confident.

Appropriateness

Do you think the scale for each criterion is appropriate?

R1: I think it's a nice balance among the three components. In terms of dispensing the drugs, student went into a varying amount of detail, but I recalled that there are two points to give to instruction. I think you can spread that out to 0,1, or 2. I think 2 is fine for caution because normally they would have two things to say about what not to do, when they shouldn't take this or what they should do if something happens, so I think 2 is fair enough.

R2: Not really for instruction and caution. I saw differences in giving detail on caution and instruction. Some student gave a very detailed caution and instruction receiving the same score as the others who only touched the topic briefly.

R3: Yes. I think it's appropriate especially when we are language instructor, we have to focus on the language. This scoring is summation as 1 or 0 and 2 or 0. To give someone a score of 2 concerns the fact that they can perform that task

at the certain kind of quality that you're satisfied with. But again, I cannot judge the quality in terms of pharmaceutical knowledge.

R4: I think it's appropriate except verifying understanding. I think it might be unnecessary to have this criterion.

Hardness

Are there any criteria you found it hard to rate the student?

R1: The only issue that came up is when the students got the wrong impression. From my point of view, it might be that what they said it's actually also correct but I recalled quite detailed guideline about what acceptable responses were in terms of drug and medication. To be fair, that was covered.

R2: I felt that patient awareness is very demanding since you have to check whether the test taker elicit the questions or express concern and that the students will receive either 0 or 2 points. I'm concerned whether this is sufficient enough, especially deciding between 2 and 0, which is absolute. When sometimes it's not a definite yes or no and as a rater I wasn't sure what score to give them to make it fair.

R3: None. But I cannot define the quality of scoring yes.

R4: I think impression is hard. Some students could not obtain all the information, which sometimes lead them to the wrong diagnose and it's hard for the rater to guide them back to the correct impression. Even though I tried to give them hints, those who don't have enough pharmaceutical science knowledge just didn't get it. That's the competent of their professional knowledge.

Recommendation

Any recommendation?

R1: No.

R2: General impression should be added. This is for the patient to rate the pharmacist as a whole considering such as the talk, the medication suggested and the instruction given.

R3: I still like dichotomous since I don't want to waste time thinking about the topic I have no deep knowledge in. Dichotomous is my preferred way.

R4: No.

Language use

Confident

Were you confident rating the students in this section?

R1: 9 out of 10 for the confidence.

R2: Very confident. I like how the score is spread out.

R3: Confident

R4: Confident

Appropriateness

Do you think the scale for each criterion is appropriate?

R1: I think that's fine. It allows for a bit more of a range of the score from the students. If the range is wider, it will make it quite hard to assess when you're doing it. So, I think it's a practical thing as well.

R2: I think the level is spread out evenly. I feel comfortable with the label for each score level.

R3: I would prefer it at three levels: 1,3,5 as in fair, good, excellent.

R4: I think it's appropriate at the four levels.

Hardness

Are there any criteria you found it hard to rate the student?

R1: I think the criteria are the kind of things we would normally assess students on. Maybe not question type, but question type was appropriate for this one. I don't think there's any problem.

R2: Word choice can appear easy, but I sometimes wondered whether some words are included in the layman terms. Students might assume that the instructor understands everything and did not switch the medical terms to layman. In addition, I talked to students after the test. Most of the students

tend to think that their word choice is easy enough to comprehend. An example students asked me whether this is easy to understand is hypertension, which I used regularly. However, the word, high blood pressure is easier to understand. I'm used to rate the students in this section because I'm a language instructor.

R3: None.

R4: I don't think so. These are the criteria we taught them in class. They are pretty clear.

Recommendation

Any recommendation?

R1: The only issue as I said was the fact that it's timed assessment. The assessor was also involved. So, we have to multi-task, but then that's why you have the video recording.

R2: I think this section is fine. The percentage of the score make me comfortable rating as a language instructor.

R3: Only three levels of score might be given to the students.

R4: None

Strategic competence

Confidence

Were you confident rating the students in this section?

R1: 9 out of 10.

R2: Very confident.

R3: Confident. It's easy to rate.

R4: Confident.

Appropriateness

Do you think the scale for each criterion is appropriate?

R1: Arguably, I think you could maybe give fewer points to the concluding the encounter. Maybe if you spoke to the pharmacist, they might give you another

answer. For me, once they have told you once they have given you the treatment and explained it, said thank you and get out, I don't know if 10 percent of the whole points is too much versus voice. Voice would be something very important I think, especially a Thai person speaking to a native speaker who might not be familiar with the Thai accent.

R2: I think it's appropriate.

R3: I think it's appropriate. But I would prefer it to be three levels probably the same format as language use.

R4: I think it's fine. They are appropriate.

Hardness

Are there any criteria you found it hard to rate the student?

R1: No.

R2: Silence can be hard to interpret. If pharmacist allows some silence for a patient, it can be interpreted in many ways as in both negative and positive way. Another topic is the professional manner and posture. The posture can be hard to judge since the administration is set sitting down and some student was concerned with the range of the recording. Professional manner is a question since the pharmacist might need to use different approach in talking to each individual. For example, they might need to be extra friendly when talking to the elderly.

R3: None.

R4: None.

Recommendation

Any recommendation?

R1: Just the voice and concluding the encounter. Maybe, have them slightly more balanced. Maybe not necessary the same but give more to voice and slightly less to concluding the encounter.

R2: You might need to consider the second rater who might be harsher since they do not need to act out.

R3: I felt that some part is similar to the pharmaceutical knowledge. It can be a double penalize or a double score. I think this section can be merged with the language use and call it communicative competence while the pharmaceutical knowledge is the content knowledge that students bring in.

R4: The time constraint is the issue for me. I need sometimes to rate, but I was pushed in with the next student. I feel I need more time in between the students. This issue doesn't happen when I was a second rater. It was easier for me when doing the second rater since I didn't need to worry about the detail I need to cover while acting as a patient.

Discussion

The interview with the raters show different aspects the raters concern is on, which are discussed according to these following topics: confidence and comfort, easiest and hardest section, Pharmaceutical science knowledge criteria, language use criteria, strategic competence criteria, adequacy and ability to assess, time constraint, and recommendation.

Confidence and comfort

One rater was found to not being comfortable using the rubric while the rest did not report such problem. This stems from the fact that this is the first time the rater taught this course and is not used to the rubric and the task. The other raters have been teaching the course for more than three semesters. Considering the number of time in teaching and assessing, it is understandable why the other raters feel comfortable and confident using the rubric, while the new rater was not. Although this new rater has taught medical courses for many years and is familiar with the medical content, the rater was not adapted well with this task. This is due to the confidence of the first-time teaching and the different administration where other courses lay the responsibility of the role-play on the students not on the rater, so that the raters in that course can focus on the rubric rather than multi-tasking between doing role-play and assessing the students. In addition, the task in that medical course assesses

only linguistic features and communication skills but not on the content knowledge.

If the system of rating is taking into account, it is possible that the new rater is used to the partial credit system and doesn't like dichotomous. The rest of the group do not have problem with dichotomous system and are confident and comfortable using the scale. For the rest of the rubric which is partial credit, all of the raters find it comfortable and were confident using the rubric.

Easiest and hardest section

Three of the raters viewed pharmaceutical science knowledge as the easiest, while one rater viewed it as the hardest. The majority believes that pharmaceutical science knowledge is the easiest part because they only need to check whether each criterion in that part is performed or not as in a yes or a no. The rater who found it to be the hardest is due to the fact that some decision is undecided as a definite yes or no, but rather as in percentages as somewhat yes or somewhat no. In fact, the easiest part for this rater is language use, which is an partial credit scale.

The hardest section for two raters is strategic competence, while one rater believes that it's the language use. The reasons for strategic competence are being subjective and some similarity to the pharmaceutical knowledge. One rater viewed language use as the hardest because it's the rater's specialty and many levels can be applied to that instead of just a yes or no.

Pharmaceutical science knowledge criteria

Apart from being rated by the majority as the easiest section to rate, five criteria at a score of 2 or 0 are criticized by the raters, which are patient awareness, instruction, caution, verifying understanding, and impression. First of all, patient awareness was mentioned by a rater who does not feel comfortable using the rubric. The problem the rater has with this criterion is that the detail the test takers need to do in order to receive the score from.

Second and third are the instruction and the caution, both of which is similar that various amount of detail can be provided. A rater stated that it should be possible for the language instructor to rate the quality of the content knowledge. The raters felt that the different amount of detail should be reflected in the score as in partial credit scale as in 0,1, and 2, instead of dichotomous at 2 or 0.

Furthermore, verifying understanding is being attacked as unnatural in a normal situation. To solve the problem, the rater suggested that the students should be able to get the score on this part without having to specifically ask the patient to repeat the information before ending the conversation. If the patient feels that their understanding is being verified, then the score should be given without an explicit question. Finally, impression is a hard criterion for both the raters and the test takers. The question for the rater is when the test takers state different impression other than the suggested answer. Even though the rater tried to give additional detail to guide them to the right impression, some of the test takers couldn't get the clue. The question to the language instructor is that if the test takers state the wrong impression, should score be given to them as they have performed the criterion? The raters strongly believe in the answers suggested by the informant specialist and felt undecided to give the score to the test takers who could not show enough pharmaceutical science knowledge with a correct impression.

Language use criteria

All of the raters agree that they are confident on rating this section. The majority believe that the score level is appropriate while one rater suggested the score level to be decreased from 4 levels to 3 levels with the same full score. The consideration for the change needs to be discussed with the whole group of experts and raters. In addition, one rater reported that the word choice criteria can be hard to assess since there is no clear specification of which words are considered not too hard and understandable to layperson.

This problem could be solved with a list of medical and layman terms provided.

Strategic competence criteria

Every rater inserted their confidence on this section and some suggested a change to the criteria. First, the score of concluding an encounter might be lessen since the performance for this criterion is similar to the criterion; provide reason(s) for dispensing. In the pharmaceutical knowledge section, to receive the score from this criterion is to state all the symptoms the test takers ask from patients, which is similar to summarizing information. If this criterion is erased, more score can be added to voice criterion, which the rater viewed it as important as the test taker is non-native speakers, who might employ voice to help them in communication. Next, silence should be defined more clearly as what kind of silence is positive and what is negative. A list of positive and negative silence can be provided to both the test takers and the raters. The posture is another criterion to be considered since the test taker sits down for the whole test in order for the recording to be done. So not much judgment might be able to obtain from that or it can be considered a free score given. To eliminate this criterion or not needs to be consulted with the whole group of the raters. Lastly, individual may perform different professional manner to suit different types of patients. Such claim is true when the situation is uncontrolled, but not in this controlled test administration. The problem of this manner concern may stem from not enough examples of the role-play to show to the rater. Appropriate and inappropriate manner are needed to show to the rater to get the clear cut.

Adequacy and ability to assess

All raters agree that the detail on the rubric is enough and that no other description is needed to be added. In terms of ability of the test to differentiate the student's dispensing skills in English, the raters believe that the rubric with guided criteria has the capability to do so. The question though

is on the pharmaceutical science knowledge that the raters do not have knowledge on. Their concern came from different response the students produced especially when the response is totally different from the suggested answers. The students may have the ability to dispense drugs in English, but the capability to give the possible impression and dispense the correct medication is questionable.

Time constraint

The raters complained about the time constraint in this test administration. The way the administration was held needs to be changed. More staff might be needed both for assessing the test takers and for administering the test. Applying more raters can assist not only the shorter time frame for the whole administration but can also decrease the weariness of the raters. However, such administration would be able to apply if the schedule of the raters is available at the same time.

Recommendation

A concern from a rater is whether the task can be totally unprepared versus the current version where some part of information is given to the test takers to study a week before the test. The rater believes that the students might get to rehearse for the test when in fact rehearsing can be a good technique for them to remember the process for dispensing. However, if unprepared task can be applied, the task can appear more authentic as well.

One criterion is suggested from the rater, which is an overall or general impression. This is the criterion being used in a test for medical students and is considered as a part of communication skill. That medical test, however, do not contain any criteria on professional content. Since this criterion can be judged by language instructors as it is a part of the service assessment, adding this criterion in needs to be discussed with the rest of the raters, but not the informant specialists.

Objective number three

To establish the extent of this dispensing task's test-usefulness.

Following L. Bachman and Damböck (2018)'s AUA, it consists of four claims: the intended consequences of using the assessment, the intended decisions to be made, the intended interpretations, and the intended assessment records. Each claim is explained with intended outcome and one or more of its qualities. In addition to that, backing is provided to strongly support the AUA.

Claim 1: Consequences

The intended consequences of using dispensing-skill assessment in the English proficiency will be beneficial to the stakeholders, which are test takers, language instructors, and school as indicated in the table below.

Table 50: Intended consequences

Intended consequences	Intended stakeholders
1a. Students who mastered the task will be able to dispense drugs in English successfully. 1b. Student who do not mastered the task will receive a recommendation on what to improve.	Students (test takers) Teacher (language instructor)
2. Students' listening and speaking skills will be improved.	Students (test takers) Teacher (language instructor)

For the first intended consequence 1a, the students who mastered the task will be able to dispense drugs in English successfully. This will benefit the students as they can employ this skill in their real life when they do an internship or in work at the drug store when they graduate. It can also benefit the teacher as the aim of the course is achieved. Otherwise, the 1b intended course is applied. The student who does not master the task will receive a recommendation on what to improve. Students are given recommendation on what skills to improve with the focus on linguistic and communication skills stated in the rubric. Teachers learn how to guide the students and specify the areas important to the performance assessment. The second intended consequence is improving

students' listening and speaking skills since the assessment and the content brought in is a combination of various accents and styles in speaking. Students are exposed to various accents starting from the content taught to prepare students for the test task. Students have a chance to prepare the test task with their instructor and their peers, which can increase their listening and speaking skills. During the test task, students are matched with an instructor who never trains them the task preparing them to be accustomed to new accents and speaking style, which finally improve their listening and speaking skills.

Table 51: Possible consequences

Other possible consequences	Stakeholders who might be affected
1. The faculty will receive fewer complaints on students' performance when training at the drug stores and when working after they graduate.	Students (test takers) Teacher (language instructor) Faculty
2. Students get to practice for their license exam.	Students (test takers) Faculty, University

Two other possible consequences are listed for this test task. If students can master the dispensing skills in English, the faculty will receive fewer complaints on students' performance when training at the drug stores or even when working after they graduate. Fewer complaints would have a direct effect on the students and the faculty as the faculty always asks for reflection from the drug stores about the internship session they send students to every year. The effect of this possible consequence can also be beneficial to language instructors as it is an indicator that the course content and objective are achieved. The next possible consequence is the opportunity for the students to practice for their license exam. Since the task is developed partly based on the Thai rubric, which is in accordance with the license exam test in Thai, preparing for and mastering this dispensing test can be a practice test for the students. This will also be beneficial to the faculty, who spends some time training the students for the exam in their sixth year. When the students are ready for the exam and can achieve the test,

the university can benefit from this as they earn reputation when a high number of students pass the test.

Backing for Claim 1

The backing for claim 1 is the students' feedback. A questionnaire was given to students after the test to collect their feedback. The questions focus on the rubric usage with a section available for students to provide comment One of the questions students are asked is whether the detail on the rubric improve their performance. Out of 147 students who took the test and answered the question, 141 students or 95.9 % answered yes while 6 students or 4.1% do not think that it helps improve their performance.

Apart from the question, space was provided for students to write down their comments on this test task. Not all of the students filled in the comments part, some comments related to the intended consequences gathered are found to support claim 1 as shown below.

"I felt like I was in the real situation. At least, it makes me realize what I lack of and what I should improve on." (Translated from Thai)

"This activity is very good and fit to our professional work. Thank you."

"It's a good assessment already. Easy to understand a rule. Not take a long time during a test."

"This kind of test administration can improve my communication skills and I can use it in my daily life." (Translated from Thai)

"This role-play will be very helpful for my future. Thank you."

"I would love to practice this before the internship program to get me ready for it. In general, I would say that this test administration is excellent. It helps encouraging me to improve my English proficiency." (Translated from Thai)

Claim 2: Decisions

The intended decisions are usually linked to the intended consequences. The first decision is that the teacher decides to give scores on the basis of the assessment. This decision is made two times both immediately after the assessment. The scores from two raters are summed and divided by two if the difference between the raters is not over 3.5 out of 50. When the score difference is over 3.5, the decision on the scores cannot occur immediately right after the assessment, since a third rater needed to rate the assessment. The process of equaling the scores might take up to two weeks depending on the schedule of the second rater, who will rate the performance through video observation.

Table 52: Decisions

Decisions to be made	**Individuals who make the decisions**	**When the decisions will be made**	**Stakeholders who might be affected**
1. Decide which scores to give based on the assessment	Teacher	Immediately after the assessment and after two raters have turned in the scores	Students Teacher
2. Provide students with feedback on their dispensing performance	Teacher	When the score is announced	Students Teachers

The second decision is that the teacher provides feedback to the students. This process occurs two weeks after the assessment is completed. The instructor of the class is the person who provided the feedback to their students in the class together with the announced score. The students are given detail on what they have done well and what they should improve on their future dispensing task, which can be either for their Thai dispensing test or the real dispensing in the drug stores.

Backing for Claim 2

For value-sensitivity, rubric for task performance was given to students a week earlier for them to study. Their class instructor was responsible to explain the rubric usage to students and how the scores are given. In addition, the rules on test administration were given to students in hardcopy a week earlier

as well to prepare them for the real test administration. The rules of the test administration for both the instructor who administered the test and for students as test takers can be found in the Appendix.

For equitability, the students' performance is rated by two raters and if the differences between the raters exceed 3.5, a third rater is used to counterbalance the score. Apart from that, the raters are trained on how to rate the students in a similar aspect. The training session was provided to all raters before the content is taught and a week before the exam. The raters are allowed a week to study the content for the role-play and a chance to ask any questions concerning the role-play. The MFRM is used to investigate whether any students need to adjust to their fair score as explained in the previous section.

Claim 3: Interpretations

The intended interpretation about three aspects of students' dispensing ability, which are pharmaceutical knowledge, language use, and strategic competence, are relevant to the summative decision. The interpretations are sufficient for the low-stakes summative decisions, which will be made on the basis of this assessment. The interpretations are meaningful with respect to the course content on dispensing skills, and generalizable to the tasks taught in the current classroom.

Backing for Claim 3

The backing for intended interpretations can assure the stakeholders that the interpretations are relevant, meaningful, and generalizable. First, relevance can be proved with the evidence of joint effort with pharmacy informant specialists and pharmacy experts through interview, consultation hours and questionnaire as described in the previous section. For generalizability, the IOC index was utilized to ask a language expert for agreement and consultation hours with pharmacy informant specialists. In addition, the SPEAKING grid and the manual provide an analysis of the task characteristics in classroom and the administrative procedures in detail for stakeholders involved, which are administrator, raters and students.

Claim 4: Assessment records

The scores from the dispensing assessment are consistent across different times of administration to different groups of students. Students' scores are consistent across different tasks and different role-play instructors. Students' performances are scored consistently by the teacher, according to the rating scale with the aid of video record for second observation and MFRM to calculate the fair scores in case the students are assessed by harsh rater.

Administrative procedures for the assessment task

The assessment happens on a day where normal classes are held. A total of five classes is held on for three hours each, three classes in the morning from 9:00 am to 12:00 pm and two classes in the afternoon from 13:00 to 16:00 pm. The instruction is provided to students a week earlier that their normal class would be utilized as an assessment time. A total of four rooms are used for the morning session and three rooms for the afternoon session. In other words, one big room is added to the number of the normal classroom used. All of the students are gathered in a big room, while the instructors are in the normal classroom. That means students from three sections in the morning are kept in one room from 9:00 am to 12:00 pm on the test day. Although students will perform the test for only five minutes roughly, they are not allowed to wait outside the waiting room.

No cell phone usage is allowed during the waiting time. Students are allowed to study with their friends and on the textbook. Every student is given a drug label form, which they are asked to fill in their name and section before entering the test room. Students will turn in this form to the rater once they finish the test. Random student name lists were given to instructors in the testing room. A coordinator, who has the name list, informed the students their number for the test and the test room they are supposed to go to. While one student is taking the test, two students in the queue are asked to leave the big room to wait in front of the test room. Each student performs with the rater who is not their classroom instructor. Both the students and the instructors are

informed that when they happen to meet either one of the other from their own class, the students are to leave the room to seek the coordinator.

Students enter the test room with the drug label form and a pen or a pencil of their choice. Their seat is arranged to be across from the raters who have a device for recording the performance. The students inform their name to confirm their identity according to the list of the instructors. Students start the task when they are ready. The person who pressed start and stop the device is the students if the device used is laptop and the instructors if the devices are iPad and camera. Each student is given five minutes to finish the task, which is asking the rater who acts as a patient, dispensing appropriate drugs, and giving instruction and caution. Students who finish the test and the questionnaire can collect their belongings and leave the room without talking to the students who have not taken the test.

Scoring method

Three main criteria are divided in the rubric, which are pharmaceutical science knowledge, language use, and strategic competence, each composed of 30%, 40% and 30% respectively. The total raw score of three sections at 50 is calculated to 20. The score system for pharmaceutical science knowledge is dichotomous between 1 and 0 or 2 and 0. The students receive the score when they perform the criteria listed. The score system for language use is partial credit from 2 as need improvement, 3 as fair, 4 as good, and 5 as excellent. The raters assess the students based on the level they demonstrate. The strategic competence is rated dichotomous and partial credit system with same label as language use. The score ranges differently according to the sub-criteria. The detail of the rubric can be found in the Appendix.

The first rater acts as a patient and rated the students during and immediately after the performance. The first rater adds up the total score and turned it in to the coordinator. Any feedback can be written on the form. The second rater observes the performance through video, rates the students on the rubric form, adds up the score, and turns the form in to the coordinator.

Once all the forms are turned in, the coordinator checks the score difference between the raters and assigns the third rater if the difference is over 3.5. An adjustment from MFRM might be applied in case of fair scores. The detailed scores and feedback are given to students approximately three weeks after the test.

Possible sources of inconsistency and backing for claim 4

The possible sources of inconsistency are divided into three types: inconsistencies in the administration of assessment, inconsistencies across different assessment task, and inconsistencies in how students' performances are scored. These inconsistencies are assured by the possible backings as shown in table 52. Firstly, the inconsistencies in the administration of assessment can stem from the different times of administration, which in this task performance is the morning and the afternoon. In addition to that, different administration to different sections of students might occur as the numbers of sections are not the same. These possible inconsistencies can be prevented by following the administrative procedure restrictively so that what time of the day or what sections students are in do not matter.

Table 53: Possible backing to assure the consistency of the score

Possible sources of inconsistency	**Possible backing to assure consistency**
1. Inconsistencies in the administration of assessment - Different times of administration - Different administration to different sections of students	- Documentation: administrative procedure to be followed
2. Inconsistencies across different assessment task - Different task situation - Different role-play instructors	- Documentation: the detail of the task specification to be used in the assessment - Teacher's notes on the difference response other than stated in the task specification
3. Inconsistencies in how students'	- Documentation: Rating scale, video

performances are scored - Different applications of the rating scale from one student to the next - Different applications of the rating scale among the raters	records, training session, and instructions for scoring - Teacher's notes on how the rating scale may have been applied differently from one student to the next - Teacher revision of the scores through video observation

The inconsistencies across different assessment task may happen due to different task situation and different role-play instructors. The possible backings are the documentation on the detail of the task specification used in the assessment. The raters are supposed to study the task beforehand and have the specification with them during the assessment. They can make notes if the students' responses are different from the suggestions on the task specification. Lastly, the Inconsistencies in how students' performances are scored can be different applications of the rating scale from one student to the next and different applications of the rating scale among the raters. The first thing to overcome the issue is applying the training session for the use of rating scale with video records of sample performance and instructions for scoring. Apart from that, MFRM can be used to calculate the differences in scores that may be too harsh for the students from one rater to another and apply the fair scores to students. Teachers can also take notes how the rating scale may have been applied differently from one student to the next. Since the performance is video recorded, the score revision can be applied as well as using the fair scores from MFRM.

Chapter 5

Discussions and Conclusion

This study described the process in developing and validating the tasks and rubric for dispensing drugs in English through quantitative and qualitative methods. The ultimate outcome of this study is the validated task and rubric to use with pharmacy students when dispensing drugs in a Thai context. It also determined to establish the extent of this dispensing task's test-usefulness. Apart from developing the task according to Douglas (2000), the framework that this study follows is SPEAKING grid (Hymes, 1964) and the Assessment Use Argument (AUA) (L. Bachman & Damböck, 2018).

The participants of this study were divided into four groups covering informant specialists, pharmacy experts, pharmaceutical science students, and language instructors. Informant specialists were three pharmacy instructors from Faculty of Pharmaceutical of Science, Chulalongkorn University. Twenty-one pharmacy experts retrieved from the list of drug stores that work with the faculty accepting the student trainees. The next group of participants was one hundred and forty-seven fifth-year students studying at the Faculty of Pharmaceutical Science of Chulalongkorn University in the academic year of 2016, while three students from the group were volunteered for semi-structure interviews. The last group of participants was six language instructors who assessed the dispensing performance while the interviews were conducted on four raters.

Three research instruments in this study were questionnaire for students and pharmacy experts, a set of questions for interviews for students, and a set of questions for raters. The questionnaire focusing on the criteria was validated by a Thai language expert in ESP and two Thai pharmacy informant specialists. The question set for interviewing students was reviewed by a language expert in assessment. The interview aimed at generating the general idea of dispensing assessment in Thai and the students' perspectives in dispensing assessment in English in order to derive a SPEAKING grid for dispensing skill assessment in English.

The question set for raters was reviewed by another language expert in assessment. They meant to specify the difficulties and ease of rubric use as a whole and in detail.

Developing the tasks and the rubric was first conducted by interviewing two informant specialists and three pharmaceutical science students with different levels of English proficiency. The data were analyzed and compared to literature review of the task and the rubric used in previous studies. Based on the data from the interview and the questionnaire, the SPEAKING grid for dispensing skills was developed first as a frameset for the task and the rubric to follow. Each task was developed according to the SPEAKING grid and information as in symptoms was provided by informant specialists. For the task validating process, the task was revised several times by informant specialists to investigate content and language instructors to review skills and language use.

The interview with informant specialists and students, and the Thai dispensing rubric, a questionnaire on the criteria in the rubric were developed and distributed to twenty-one pharmacy experts and one hundred and forty-seven pharmaceutical students. The questionnaire of thirty-seven items distributed to pharmacy students and pharmacy experts was analyzed quantitatively through ANOVA and post hoc test analysis. The rubric was first reviewed by informant specialists including the score ratio for each criterion. It was used on one hundred and forty-seven students leading to the scores to be analyzed in MFRM. Since the students can be divided according to their major, the data were compared to find agreements among three groups, which were pharmacy experts (PE), pharmaceutical care students (PC), and pharmaceutical science students (PS). Next, the scores on the task performance were analyzed through MFRM to identify the fair scores of students, the raters and the criterions. Finally, a content analysis was conducted on the interview with raters to validate the rubric and also to support the AUA.

Summary of the findings

The findings in this study are based on five sources of detailed information gathering from interview with informant specialists, interview with students, analysis on questionnaire, MFRM, and interview with raters. First is the interview with

informant specialists on the courses and the background of the students. Before students reach the second semester of their fifth year, they have taken internship courses for three semesters accumulating for four hundred and twenty-two hours of practicing at the drug stores. All of the internship happened under the supervision of the pharmacists in charge at the drug stores. For the last internship course, the students were assessed in a real situation with real patients for their dispensing skills in Thai by their instructors, which occurred two times, one at the beginning of the course and the other one at the end of the course. The students were prompted to spend no more than five minutes with each patient. A problem was reported from the internship drugstores around Thailand concerning the students' ability to communicate and dispense drugs in English proficiency.

Second is the interview with three students from different levels of English proficiency. As far as the time on task is concerned, two students believed that the same amount of time should be used for dispensing skills in English as in Thai. The weak student viewed the way the assessment works as the same as the strong student, while the student whose English proficiency was in the middle thought differently and needed more time. All of the students were fine with one rater when rating live, but would prefer two raters to counter balance the score if possible. However, they would rather not have two raters rating presented at the same time due to the anxiety. This goes in line with the number that IELTS employed when the test taker is interviewed with their speaking task. When being questioned about authenticity, students preferred to have unprepared and unrehearsed tasks where no preparation or prompt is given. They also preferred the process and the rubric similar to their Thai dispensing assessment since they were familiar with them. The preference in an impromptu task and the process is parallel to the real-world situation and Thai dispensing task, which is the practice for the students to take the pharmacy license examination. Such preference can promote the students' self-confidence as Lundberg (2008) suggested that simulation is useful for students to apply their theory on simulation.

On the whole, PE and PS shared similar views in terms of keeping the distant relationship, such as avoiding the name use, while PC students seemed to show

empathy in this dimension and did not mind sharing such information. One compelling item is “Thank patients”, which is the only item being viewed different among groups. The important level the group rated is PE as the highest to PC and PS as the lowest. It seems that the students had not achieved an important part of the service care to patient as in saying thank you.

In terms of rubric, the three easiest criteria to rate, which all belong to the pharmaceutical science knowledge, are *allergy, underlying disease*, and *instruction*. The three hardest criteria, which are all categorized as language use section, are *grammar, pronunciation*, and *question type*. Four criteria out of seventeen needed to be adjusted are *verifying understanding, caution, pronunciation*, and *initiating communication*. *Verifying understanding* and *caution* are found to not assess the same dimension as the other criteria and need adjustment. For pronunciation and initiating communication, the problem was found on the step calibration between the score level of three and four for both criteria.

Last is the interview with four raters. Although the majority of the raters felt at ease using the rubric, the hardest section for two raters is strategic competence and language use for one rater and the pharmaceutical science knowledge section as the easiest section. Some suggestions on the criteria score weight are on instruction and caution as the detail can vary. One criterion was viewed as unnatural is verifying understanding. And the criterion the raters have questions with is the impression as some students gave different answers than the suggested detail on the rater handouts. The time constraint was also an issue to the raters as some felt the rush from the test takers.

Discussions and Conclusions

This study aimed at developing tasks and a rubric for classroom performance-based assessment in dispensing skills of pharmaceutical science students. Apart from the interviews with informant specialists and students, the questionnaire was used to obtain a majority view toward the dispensing examination tasks. The study also created and validated the rubric for the dispensing tasks through MFRM. Some of the following findings agree with previous studies as seen below.

Creating the task and the rubric for assessing pharmacy students on dispensing drugs in English

In order to create the task and the rubric successfully, this study found that using several sources of data were needed together with the review to the context use. Relying on literature review only cannot lead to the local context involvement the study aims for. A needs analysis can decrease the researcher's bias by including various stakeholders in the needs analysis (Huhta, 2010). Although some studies (Aliakbari & Boghayeri, 2014; Edwards, 2000; Tsou, 2009) preferred to analyze the needs on only students and some (Basturkmen & Shackleford, 2015) focused on the students and the language instructors, the needs analysis through the interview with informant specialists and the students can help generate various insights and verify the literature review to initiate the tasks. Likewise, the task development tended to include at least three groups of stakeholders including professionals in the field, professionals in the educational factor, the language experts, and the test-takers (Grice et al., 2017; Johnson & Riazi, 2017; Luka, 2008; Macqueen et al., 2016). The finer needs analysis was conducted, the better language teaching program there is to serve the students with a more favorable outcome (Long, 2005). Employing both the students and the professionals in the work field and the educational field aimed at generating more precise content for the task.

In order to communicate successfully, a person needs to apply their language capability their cognitive and non-cognitive skills for a certain context (Elder, McNamara, Kim, Pill, & Sato, 2017). To achieve such result, a joint cooperation between the fields is necessary. Implementing the topical knowledge that involves knowledge in the specific field needs cooperation from the specialist informant (Douglas & Myers, 2000; T. McNamara, 1996). Yet, it is challenging to balance the content and the language. Although the study avoided the work sample approach which is not attentive to the linguistic factors, pharmacy experts did not view using correct English language and pronunciation correctly as important as both groups of the students. This result from the questionnaire matches what Elder (2016) reported about the experts' view toward the

assessment that it tends to be differ from the linguists. Similarly, Macqueen et al. (2016) stated that the experts do not have the ability to specify the needed linguistic criteria for the assessment. As a matter of fact, the experts tend to use their knowledge and professionalism as their baseline while the linguists rely on the language and communication aspects (Douglas & Myers, 2000). The professionals have a tendency to award more score ratio to their content knowledge than the linguistic detail (Byrnes, 2008). Thus, using the expert's judgment alone in assessing ESP assessment do not fit the construct of assessing linguistic ability in the content it belongs to. The development of ESP assessment with proof of validity needs collaboration between content teachers and language teachers in order to balance the joint interest. However, such collaboration may be hard to establish as not much interaction between disciplines were not originally formed (Arnó-Macià & Mancho-Barés, 2015). It is vital for a course developer to extend the relationship to the professionals in the fields. In fact, a decent educational practice can occur when the collaboration and the understanding between the content and the language instructor can be compromised (Brennan & Naerssen, 1989). One method to ease the process is that the university and the faculties engage in and assign voluntary persons who could be responsible of such consultation.

Various classification among academic disciplines have long been acknowledged (Jones, 2011). Based on the attitudes of the two groups of the students, this study concluded that they held different degrees of attitudes. This is in line with Linnenluecke, Russell, and Griffiths (2009) who suggested that a multi-culture can exist within the same organization. It is asserted by Lee (2007) who mentioned that the institution and the discipline can result in sub-culture. This can be found in significantly different views of PS and PC on servicing the patient although the pharmaceutical science content is the same for both departments until the fifth year, which was the time the data was collected. It is speculated that the students' work attitude was affected by the department they are in regardless of the same content facilitated. Pharmaceutical care students rated the items higher than the other groups since they were groomed to the role

of dispensing drugs in the drug stores. In contrast, pharmaceutical science students are to work in the drug industries which seem not to require them to perform much of service-minded skills. Considering the service-mindedness, it is worth noting the training for pharmacy students may want to pay more attention to the quality of service-mindedness, especially pharmaceutical science students, who did not view offering warm greeting as high as pharmaceutical care students and pharmacy experts. This result agreed with Austin (1990) who stated that the culture of the discipline can mainly contribute to the identity of the member. The culture of the field the students are in might reflect their thoughts on how important the criteria are in performing the task.

Validating the task and the rubric for assessing pharmacy students on dispensing drugs in English

Validating the task and the rubric needed several consultation hours to fill in the information that may be missing. Since the task depended on the rubric, the use of MFRM and the interview with raters greatly contribute to the validation and the detail on what and how to improve the rubric. The MFRM result can be used to explain the validity of the three facets candidates, raters, and criteria. When the candidates are misfit, they can be fairly adjusted according to the fair score average calculated by the MFRM program. Apart from that it can determine the criterion that does not belong to the same dimension. This led to the adjustment and scores of the rubric, which is similar to Johnson and Riazi (2017) that suggested the revision of their rubric after MFRM result.

Several changes were applied to the adjusted rubric (see Appendix). Two criteria categorizing as *verifying understanding* and *caution* were in need of adjustment as the consultation with the raters suggested the reasons of the problem. First, the ways of fulfilling *verifying understanding* criterion should vary and students should recognize various ways of doing so as to perform it in a more natural way. The raters never had an experience being asked by a pharmacist to repeat the information. This criterion is stated in the Thai dispensing rubric as ‘rechecking to reduce the medication error’. According to the specialist who

taught the course to students, completing this task in Thai can be done indirectly without asking the patients to repeat the information. The students can perform an act of checking the medication as well. However, this criterion will not cause problem as it was eliminated. It was agreed among the language instructors to eliminate *verifying understanding* since both qualitative data from the interview with the raters and quantitative data from MFRM suggested that it does not assess the same construct.

In order to avoid the judgment from language instructors on *caution*, *content appropriateness* was added to allow the rater to assess their satisfaction as patients. The 'caution' criterion needed the expert review on the score ratio. This is due to the comments from the raters explaining the detail that they (as non-experts in pharmaceutical skills) believe the item could be judged by. The problem of this criterion was reported by the raters as a questionable dichotomous decision. Some of them believe that linear scale can be applied due to the amount of detail that might be able to justify by the language instructors following the information provided in the task, which was designed by informant specialists. The detail dividing into different levels of satisfaction might allow language instructors to assess the level of student's professional skills. Such interference showed the attempt of involvement the language instructors wish to apply. This is in line with other studies (Elder et al., 2017) that strived to include the experts' judgment in their assessment, which responded to Davies' concern of on the judgment of language instructors whether the assessment on the language use in particular context focuses on linguistic detail too much and might ignore the norms of that context (Elder et al., 2017). Future studies on the similarities and differences in scoring system and reasons comparing to the professional judgment can answer the adjustment to this issue.

The hardest and easiest criteria to assess match the result from the interview with the raters. The top three easiest criteria to rate, which all belong to the pharmaceutical science knowledge, indicating a good sign of adopting the background knowledge of the test takers as a part of this ESP assessment since the researcher do not want the content criteria to appear too hard for language

instructors to use. The top three hardest criteria, which are all categorized as language use section, reflecting the strong ability in assessing this field, which is the expertise the language raters belong to. This also confirms their confidence in assessing the students in these areas. In fact, the criteria in the harsher area all belong to language use and strategic competence while the more lenient criteria belong to pharmaceutical knowledge, which are all dichotomous system. The MFRM result coordinated with the interview with the raters. Despite the difficulties in some criteria, the raters believed that this rubric can be used to assess pharmacy student's dispensing performance. In addition, the result has shown that the dichotomous system is beneficial in deciding on the students' professional knowledge. Similar to dichotomous decision on performance tree decision (Fulcher, Davidson, & Kemp, 2011), this study has shown that the dichotomous system is beneficial for language instructors in making easier judgment on the students' applied professional knowledge. This method is similar to what the pharmacists utilized in assessing pharmacy student communication as Gillette, Rudolph, Rockich-Winston, Stanton, and Anderson Jr (2017) applied the dichotomous system on their pharmacist-patient counseling rubric.

Several key findings are recognized from this study. The faculty and the subject instructors have prepared the students ready for the real world dispensing experience at this stage by following the guideline of training students to meet enough hours of requirement. This fact helps assuring that the students have experience in dispensing skills in Thai. With at least three hundred hours of internship over three semesters, it is thus better to follow the same process and use similar criteria to what they have been trained for and practiced with, which coordinates with the preference of the process and the rubric the students stated in the interview. In addition, no matter how well a validated framework was developed for a context, it might not meet the needs of the students in another context (Grice et al., 2017). The criteria adopted from the Thai dispensing rubric that are related to the pharmaceutical science knowledge were specified as dichotomous system and the score weight of the criteria was designated by informant specialists. Apart from that, the criteria were gathered and divided it

according to the genres they belong to. The sections are similar to the six phrases in Woodward-Kron and Elder (2016)'s studies related to the OSCE and OET examinations, which can reflect the authenticity of this test tasks.

The time for the assessment is crucial to both test takers and the raters as interlocutors. As a matter of fact, the length of encounter may affect the interpretation of the test takers' performance (Grice et al., 2017). The limited time of assessment should be set the same for all. Based on the information about the Thai dispensing skills examination in the courses from both informant specialist and students, the time limit for each test taker is set at five minutes, which is considered not too long or too short. In fact, the group of this study spent an average of 5:14 minutes per person according to the time recorded, which included the time of writing the label. In terms of the number of the rater, the study decided to use one rater for the face-to-face conversation following the Thai dispensing skill examination but video record for the second raters to assure the scores reliability. According to the student, the video record can also decrease the students' stress of having two raters listening and rating them while receiving the scores from two raters. Students also stated that they are familiar with the process and the Thai rubric which they have to prepare for the license examination. Thus, adopting a similar process of Thai dispensing skills and the criteria can essentially boost the washback for the students since they can employ the skills obtained from the test in their real life as some of them expressed their opinions in the AUA.

The last aim is to prove the extent of this drug dispensing task's test usefulness, which is provided through AUA. Indeed, various best practices on language assessment were mostly based on AUA (Avineri et al., 2010). Such framework to prove the test-usefulness can address and circumvent the common issue that may occur in quantitative research (Purpura, Brown, & Schoonen, 2015). The four claims were explained to justify the use of this test task supported by the interviews and the MFRM, which is similar to how Johnson and Riazi (2017) validated their local writing test using an amalgam validation framework. In fact,

many classroom researches (Llosa, 2008) have applied different ways as a framework to investigate their test-usefulness and among them is AUA.

This study also proposed intended consequences, decisions, interpretations, and assessment record with evidences for backing the claims. While negative consequences could be used to weaken the consequences (Papageorgiou & Tannenbaum, 2016), this study presented the intended consequences with a positive agreement of the test takers through questionnaire and comments with over 95% believed that the test improved their performance. Papageorgiou and Tannenbaum (2016) stated that the intended decisions, such as the cut score, can be affected by the relevant stakeholders, which is similar to this study that the scores of each criterion was adjusted by informant specialists and language instructors. Although Papageorgiou and Tannenbaum (2016) opted out the meaningfulness to their standard setting, the intended interpretations in this classroom assessment were sufficiently and meaningfully linked to the course content through the description of SPEAKING grid and the manual offered to all stakeholders. The assessment records this study provided documentation on administration and rating procedures.

Relying on the scores only cannot lead to the rating consistency (Purpura et al., 2015), training rater is a must and the use of MFRM is recommended. The result that one of the rater who has been teaching for more than thirty years is the harshest rater matches what Eckes (2009) suggested that the senior rater may tend to rate harsher as they may want to “set the standard” to the others. In like manner, a study by Bonk and Ockey (2003) found that the more experience the raters obtain, the higher their standards are raised. In addition, the experienced raters tend to have more bias than inexperienced raters even after the training (Bijani & Khabiri, 2017). However, the harsh rater in this study conformed to the group agreement and is not considered a threat to the assessment. Following such a result, training is needed to ensure that all raters are in the conformity, especially the senior or experienced raters. Furthermore, the rating bias might not be completely banished, the training and norming session can help the rater obtain self-consistency (Avineri et al., 2010). Likewise, Johnson and Riazi (2017)

recommended that the rater training and the norming session should be held at a different time in order to aid rater consistency. The rater who misfit can receive more training sessions to align the judgment to the group, while the harsh rater can be matched to the more lenient rater to balance the scores. Following suggestions from Stahl and Lunz (1992), Lumley and McNamara (1993) endorsed providing feedback to raters on how they assess each item together with the interview or the protocol in order to decrease the rating inconsistency. Similarly, utilizing the result from the MFRM to inform each rater's performance in assessing can lessen the raters' judgment effect (Bonk & Ockey, 2003). Based on previous studies (Bijani & Khabiri, 2017; Bonk & Ockey, 2003; Eckes, 2009), good training sessions should offer the raters the detail of their rating performance, which can be obtained from MFRM.

Implications

Based on the result relating to literature review, two main implications stemming from this study are LSP assessment and reliability.

LSP assessment

The authenticity this rubric based on is the information from informant specialists and the Thai dispensing rubric the school used to train students. The task and the rubric were designed according to the Thai dispensing examination that the students have taken and the license examination they will take when they finish the course. This fact aid in the high level of washback as several students' comment mentioned this benefit in preparing them for the test and their future work. The level of high authenticity of this test limits the application of the rubric to other contexts since it was designed based on content for this dispensing test task. However, collaborating informant specialist can greatly promote the high level of washback adding the authenticity to the test. Several consultation sessions and more than one informant specialist are essential in generating useful information to the rubric and the test tasks.

The indigenous criteria and task can be developed with professional and students' collaboration. Although lots of previous research can supply suggestions to criteria, adopting the criteria from the literature review alone do not lead to the

criteria that fits the context the study aims for. Some criteria from literature review that seems relevant to other context such as scheduling doses or name introduction are not much relevant to Thai context. According to the informant specialist, the name introduction included is relevant in other context because some consultation fee can be charged and the formal name introduction is needed. Scheduling doses and name introduction may be considered a further service in Thai context, which does not exist in normal drug store visit.

Combining the content knowledge into the rubric as a dichotomous scale can be ease of use for language instructors and raters as the result was shown on the leniency the scores of these criteria are and also the majority of the raters agreed on. Since the content knowledge is the area we do not have the ability to judge it on a linear scale, a dichotomous choice aids the language instructors in assessment and can also avoid the issue of reliability of the rater's judgment as a non-expert. In fact, the same dichotomous scale should be applied to the content knowledge to avoid the difference in difficulty level as found in this study.

The medical jargon is viewed as necessary starting from the development stage, which is included as one of the criteria. In fact, some health professionals cannot differentiate the communication pattern with medical professionals from the patients (Macqueen et al., 2016). However, due to the fact that raters verbally reported some obscure cases of the word use, the current content might not be enough in guiding the differences between the medical jargon and the layman terms. A clear guideline based on the real language use should be established to define the differences for both the language instructors and the test takers.

Reliability

The assumption that raters who have been assessing performance in a related field like medical would be comfortable judging the rubric in related field cannot be made. This goes in line with Grice who mentioned that employing the assessment in medical field might not evaluate the skills of pharmacist (Grice et al., 2017). In fact, the experience in other assessment the raters bring can lead to

the obscure judgment on some criteria, especially the new criteria that the rater may feel uncertain of. The differences in scores among the new and old raters need to be evaluated through MFRM to avoid such effect on the reliability issue.

For performance-based assessment, rater training sessions are essential not only for conforming the raters' agreement to a similar alignment, but also on their performance of being an interlocutor. Unlike other performance-based assessment that do not require contribution from the raters, such as group oral discussion, the rater in oral interview might be tired from rating and responding to the test takers (Bonk & Ockey, 2003).

For performance-based assessment with later observation to occur smoothly, manuals for raters, students, and administrator are needed including an assistant to help administer the test. In a single classroom assessment with one rater and one group of students, this might not be the case. In order to achieve the goal of reliability as the performance can be assessed by more than one rater, an administrator should be in charge in addition to the raters. One extra person can aid in guiding students to enter correct examination room according to the order assigned, collect the document, and supervise the whole test. This study suggested that when the decision of classroom assessment involves more than one rater and one section, systematic training, administrator and manual must be provided.

Focusing on task administration, a number of the task, students and time need to be balanced. One thing to keep in mind if the tasks are limited in number is that the test should be administered and ended within half-a-day for the same group of the students taking the same tasks. The number of the students per one rater should not exceed twenty-five per three hours of the examination. Students can get exhausted from waiting too long in the preparation room, while the raters can be fatigued from simulating the case and rate thirty students. The limit in number of the task, students and the time is to make sure that the assessment can be administered fairly to every student and no cheating can happen or any students will benefit from taking the test after the others. The detail in administration can enhance the reliability in terms of score collection, and also

reflects the practicality issue in how the test should be held. It is noted that reliability and practicality are not separated dimension as they can overlap as explained in this study.

Limitations

1. This study applied simulation technique by training the language raters to act as patients. Although the use of extra persons could increase the reliability of the test, employing actors was not feasible to the study due to the cost concern.

2. Because of the time constraint, the first rater who acted as patients did not have enough time to write comments for each test taker. The test takers received the comments from only their classroom instructor, who observed their performance as the second rater.

Due to these limitations, the students could not receive comments from the first rater, who can provide essential comments on how they can improve their performance, since the first rater conducted the face-to-face conversation with them and may have a better sense on how their information was received and taken.

Recommendations for future research

1. If the performance-based assessment involves the use of technology like video recording, a small training session should be offered to everyone. As a couple video was missing from this study due to the human error, a training session should be lounged for people in the administration, which in this case are raters and students.

2. Deciding the ESP test task with authenticity in mind might need some discourse study to better support the scale development. Although collaborating with informant specialists has drawn out data enough for the task and criteria development, this study does not investigate in-depth study on the discourse of language use in Thai context. The future study can explore the way the language is used in this particular context, which might lead to the unique this group of professional holds.

3. In case of performance assessment where raters are required to perform as well as rating, training session should be offered not only on the rating instruction, but also on ways to perform the role. The training on the performance can ensure the reliability of the task to different students.

4. A corpus in related field can add the detail to the language use of this specific group and also the speaking characteristics of Thai students. The future research can develop the learner corpus in order to categorize learners into their appropriate level.

5. Non-verbal communication can be described in more detail with the video-recording; however, this issue is not the focus of this research. Future research could aim on examining non-verbal communication as a part of assessment, such as gesture, facial expression, and eye contact.

APPENDICE

Appendix I: Questionnaire (English)

Directions: This questionnaire is divided into three sections: 1) personal information 2) overall communication skills and 3) pharmaceutical science content. Please fill in the answers in the space provided and put a √in ☐ .

Part I: Personal information

1.1 Name _____

1.2 Field of the study ☐ Pharmacy Care ☐ Pharmacy Science

1.3 Gender ☐ male ☐ female

1.4 Grade of Experiential English I _____

1.5 Grade of Experiential English II _____

1.6 Grade of English for Pharmaceutical Profession II _____

1.7 GPA _____

Part II: Overall communication skills

Directions: Based on your pharmacy internship, please indicate the degree of importance of the following skills for pharmacist-patient communication at a drug store, on a scale of 1-4.

4 represents VERY IMPORTANT **3 represents IMPORTANT**

1 represents NOT AT ALL **2 represents SOMEWHAT**

Also, please write down your answer in the space provided.

2.1 Initiate communication

How important are these skills when you practice as a pharmacist?		1 Not at all	2 Somewhat	3 Important	4 Very
2.1.1	Greet patients.				
2.1.2	Identify yourself as a pharmacist to the patients.				
2.1.3	Introduce your name to patients.				
2.1.4	Confirm patient's identity (whose medication is for?).				
2.1.5	Ask for patient’s name.				
2.1.6	Call patients by name.				
2.1.7	Offer warm greeting				
Other: (please fill in)_____					

2.2 Use effective verbal communication

How important are these speaking/communication skills when you practice as a pharmacist?		1 Not at all	2 Somewhat	3 Important	4 Very
2.2.1	Use pace and silence appropriately.				
2.2.2	Speak loudly enough to be easily heard.				
2.2.3	Use appropriate tone of voice.				
2.2.4	Use correct English language.				
2.2.5	Use correct English pronunciation.				
2.2.6	Use words the patient will likely understand and avoid medical jargon.				
2.2.7	Modify communication to meet special needs of patients (e.g., elderly, low health literacy, cultural differences).				
2.2.8	Use open-ended questions and closed-ended questions appropriately to avoid bias (e.g. This drug is very strong. Did you throw up when you took it? – Inappropriate What's your symptom? - appropriate).				
2.2.9	Use written information to emphasize and help oral communication (Write the name of medication and the instruction on the drug label).				
Other: (please fill in) _____					

2.3: Conclude the encounter

How important are these skills when you practice as a pharmacist?		1 Not at all	2 Somewhat	3 Important	4 Very
2.3.1	Summarize information (the medication and the instruction).				
2.3.2	Ask if there is anything else patients would like to discuss.				
2.3.3	Invite patients to contact if questions or concerns arise.				
2.3.4	Thank patients.				
2.3.5	End the conversation politely.				
Other: (please fill in)_____					

2.4: Use effective non-verbal communication

How important are these skills when you practice as a pharmacist?		1 Not at all	2 Somewhat	3 Important	4 Very
2.4.1	Demonstrate appropriate eye contact.				
2.4.2	Demonstrate appropriate posture and body language.				
2.4.3	Wear appropriate attire.				
2.4.4	Display appropriate health-professional manner.				
Other: (please fill in)_____					

Part III: Pharmaceutical science content

3.1 Elicit information from patients

How important are these skills when you practice as a pharmacist?		1 Not at all	2 Somewhat	3 Important	4 Very
3.1.1	Ask patients about their concerns or reasons for visit.				
3.1.2	Give patients opportunity and time to talk.				
3.1.3	Ask for a complete record of patients' current health conditions and therapies.				
3.1.4	Ask questions to assess patients' understanding of key information about medications.				
3.1.5	Ask questions to assess patients' experience with medications currently being taken.				
Other: (please fill in)_____					

3.2 Initiate educational interventions

How important are these skills when you practice as a pharmacist?		1 Not at all	2 Somewhat	3 Important	4 Very
3.2.1	Emphasize key information.				
3.2.2	Provide reasons for advice and options for treatment.				
3.2.3	Provide appropriate recommendations based on IEASC[1].				
3.2.4	Discuss one drug or therapeutic regimen at a time.				
3.2.5	Provide complete and clear instructions on medication.				
3.2.6	Verify patient understanding of new information provided.				
3.2.7	Work with patients to schedule the doses.				
Other: (please fill in)_____					

Would you be willing for the researcher to contact you at some time in the future to discuss your answers in more detail?

☐ Yes ☐ No

If yes, please provide your email address and phone number.

Email _____

Tel. _____

Thank you for your participation. Your answers are highly valued.

[1] IEASC refers to Indication Efficacy Safety Adherence Cost-effectiveness

Appendix II: Questionnaire (Thai)

แบบสอบถาม

คำชี้แจง: แบบสอบถามฉบับนี้มีสามตอนคือ 1) ข้อมูลส่วนตัว 2) ทักษะการสื่อสารโดยรวม และ 3) เนื้อหาด้านเภสัชกรรม โปรดเติมข้อความและเครื่องหมาย √ ในช่อง ☐

ตอนที่ 1: ข้อมูลส่วนตัว

1.1 ชื่อ นามสกุล- _____

1.2 เพศ ☐ ชาย ☐ หญิง

1.3 สาขาวิชา ☐ การบริบาลทางเภสัชกรรม ☐ เภสัชกรรมอุตสาหการ

1.4 เกรดวิชาภาษาอังกฤษเพื่อการเรียนรู้ในชีวิตจริง 1 (Experiential English I) _____

5.1เกรดวิชาภาษาอังกฤษเพื่อการเรียนรู้ในชีวิตจริง 2 (Experiential English II) _____

1. 6เกรดวิชาภาษาอังกฤษสำหรับวิชาชีพทางเภสัชศาสตร์ 1 (English for Pharmaceutical Profession I) _____

ตอนที่ 2: ทักษะการสื่อสารโดยรวม

คำชี้แจง: ในฐานะเภสัชกรโปรดประเมินระดับความสำคัญของทักษะที่ใช้ในการสื่อสารกับคนไข้ที่มาซื้อยา<u>ที่ร้าน</u> โดยการเติมเครื่องหมาย √ในช่องที่ตรงกับความคิดเห็นของคุณตามระดับ 1-4 และเติมคำตอบลงในช่องว่างโดย

4 = สำคัญมาก **3 = สำคัญ**

2 = สำคัญบ้าง **1 = ไม่สำคัญเลย**

2.1: การเริ่มต้นการสนทนา

ทักษะเหล่านี้สำคัญระดับไหนเวลาคุณเป็นเภสัชกร		**1 ไม่สำคัญเลย**	**2 สำคัญบ้าง**	**3 สำคัญ**	**4 สำคัญมาก**
2.1.1	ทักทายผู้ป่วย				
2.1.2	บอกผู้ป่วยว่าคุณเป็นเภสัชกร				
2.1.3	บอกชื่อตนเองแก่ผู้ป่วย				
2.1.4	ยืนยันตัวผู้ป่วย (ซื้อยาให้ใคร)				
2.1.5	ถามชื่อของผู้ป่วย				
2.1.6	เรียกผู้ป่วยด้วยชื่อของผู้ป่วย				
2.1.7	ต้อนรับทักทายอย่างอบอุ่น				
อื่น ๆ: (โปรดระบุ)_____					

2.2 การใช้คำพูดในการสื่อสาร

ทักษะการสื่อสารเหล่านี้สำคัญระดับไหนเวลาคุณเป็นเภสัชกร		1 ไม่สำคัญเลย	2 สำคัญบ้าง	3 สำคัญ	4 สำคัญมาก
2.2.1	มีจังหวะในการพูดและหยุดวรรคตอนได้อย่างเหมาะสม				
2.2.2	พูดเสียงดังฟังชัด				
2.2.3	ใช้น้ำเสียงเหมาะสม				
2.2.4	ใช้ภาษาอังกฤษได้ถูกต้อง				
2.2.5	ออกเสียงภาษาอังกฤษได้ถูกต้อง				
2.2.6	ใช้คำพูดที่เข้าใจง่ายและหลีกเลี่ยงคำศัพท์เทคนิคทางการแพทย์ยาก ๆ				
2.2.7	ปรับบทสนทนาให้เหมาะสมกับผู้ป่วยที่มีความต้องการพิเศษ (เช่น คนชรา ผู้มีความเข้าใจเรื่องการใช้ยาน้อย และ ความแตกต่างทางวัฒนธรรม)				
2.2.8	ใช้คำถามปลายเปิดและปลายปิดอย่างเหมาะสมเพื่อหลีกเลี่ยงการชักจูง และอคติ (เช่น ยาตัวนี้ค่อนข้างแรง คุณอาเจียนหรือไม่เมื่อทานยาตัวนี้ –คำถามปลายปิดที่ไม่เหมาะสม คุณมีประวัติแพ้ยาอะไรหรือไม่คะ/ครับ – คำถามปลายเปิดที่เหมาะสม)				
2.2.9	ใช้ภาษาเขียนเพื่อเน้นย้ำและช่วยการสื่อสาร				
อื่น ๆ: (โปรดระบุ)_____					

3.2: การจบการสนทนา

ทักษะการสื่อสารเหล่านี้สำคัญระดับไหนเวลาคุณเป็นเภสัชกร		1 ไม่สำคัญเลย	2 สำคัญบ้าง	3 สำคัญ	4 สำคัญมาก
2.3.1	สรุปข้อมูล (เช่น ยาที่จ่าย วิธีการใช้ยา)				
2.3.2	ถามผู้ป่วยว่าต้องการปรึกษาเรื่องอื่นอีกหรือไม่				
2.3.3	เชิญชวนให้คนไข้ติดต่อกลับมาอีกครั้งหากมีข้อสงสัย				
2.3.4	ขอบคุณคนไข้				
2.3.5	จบการสนทนาอย่างสุภาพ				
อื่น ๆ: (โปรดระบุ)_____					

4.2: การสื่อสารโดยไม่ใช้คำพูด

ทักษะการสื่อสารเหล่านี้สำคัญระดับไหนเวลาคุณเป็นเภสัชกร		1 ไม่สำคัญเลย	2 สำคัญบ้าง	3 สำคัญ	4 สำคัญมาก
2.4.1	การสบสายตาอย่างเหมาะสม				
2.4.2	แสดงท่าทาง และใช้ภาษากายอย่างเหมาะสม				
2.4.3	แต่งตัวเหมาะสม				
2.4.4	วางตัวเหมาะสมในฐานะเภสัชกร				
อื่น ๆ: (โปรดระบุ)_____					

ตอนที่ 3: เนื้อหาด้านเภสัชกรรม

3.1: การซักถามข้อมูลจากผู้ป่วย

ทักษะเนื้อหาด้านเภสัชกรรมเหล่านี้สำคัญระดับไหนเวลาคุณเป็นเภสัชกร		1 ไม่สำคัญเลย	2 สำคัญบ้าง	3 สำคัญ	4 สำคัญมาก
3.1.1	ถามผู้ป่วยถึงสาเหตุของการมาร้านขายยา /การมาซื้อยา				
3.1.2	เปิดโอกาสและให้เวลาผู้ป่วยพูด				
3.1.3	ซักถามประวัติการเจ็บป่วยครั้งนี้ และการรักษา				
3.1.4	ซักถามผู้ป่วยเพื่อประเมินความเข้าใจในข้อมูลสำคัญของยา				
3.1.5	ซักถามผู้ป่วยเพื่อประเมินประสบการณ์ของผู้ป่วยในการใช้ยาตัวปัจจุบัน (หากมี)				
อื่น ๆ: (โปรดระบุ)_____					

3.2: การให้ข้อมูลเกี่ยวกับยาแก่ผู้ป่วย

ทักษะเนื้อหาด้านเภสัชกรรมเหล่านี้สำคัญระดับไหนเวลาคุณเป็นเภสัชกร		1 ไม่สำคัญเลย	2 สำคัญบ้าง	3 สำคัญ	4 สำคัญมาก
3.2.1	เน้นย้ำข้อมูลสำคัญ				
3.2.2	บอกเหตุผลของการแนะนำยา หรือแนวทางการรักษาชนิดนั้น ๆ				
3.2.3	ให้คำแนะนำที่เหมาะสมโดยยึดหลัก IEASC[1]				

3.2.4	อธิบายการใช้ยา หรือแนวทางการรักษาทีละตัว				
3.2.5	แนะนำวิธีการใช้ยาอย่างชัดเจนและครบถ้วน				
3.2.6	ตรวจสอบความเข้าใจข้อมูลใหม่ที่ให้กับผู้ป่วย				
3.2.7	ช่วยผู้ป่วยจัดตารางการใช้ยา				
อื่น ๆ: (โปรดระบุ)_____					

คุณยินยอมให้ผู้วิจัยติดต่อคุณในอนาคตเพื่อซักถามคำตอบของคุณเพิ่มเติมหรือไม่

☐ ยินยอม ☐ ไม่ยินยอม

หากคุณยินยอม กรุณากรอกข้อมูลเพื่อการติดต่อ

อีเมล์ _____

โทรศัพท์ _____

ขอขอบคุณสำหรับความร่วมมือ

[1] IEASC คือ การใช้ยาต้องมีข้อบ่งใช้ (Indication เป็นยาที่มีประสิทธิภาพดี ((Efficacy) มีความปลอดภัยสูง)Safety(มีมาตราการส่งเสริมให้เกิดความร่วมมือในการใช้ยา)Adherence) และเหมาะสมคุ้มค่า (Cost-effectiveness)

Appendix III: Consent form

วิทยานิพนธ์เรื่องการพัฒนาแบบประเมินความสามารถในการจ่ายยาเป็นภาษาอังกฤษของนิสิตเภสัชศาสตร์ไทย (A development of performance-based assessment on dispensing drugs in English of Thai pharmacy students)

จุดประสงค์ของการวิจัย:

1. เพื่อพัฒนาแบบประเมินความสามารถในการจ่ายยาเป็นภาษาอังกฤษสำหรับนิสิตเภสัชศาสตร์ไทย
2. เพื่อพัฒนาเกณฑ์การตรวจแบบประเมินความสามารถในการจ่ายยาเป็นภาษาอังกฤษสำหรับนิสิตเภสัชศาสตร์ไทย

วันที่ให้คำยินยอม วันที่.....เดือน.....พ.ศ.....

1. ก่อนที่จะลงนามในใบยินยอมให้ทำการวิจัยนี้ ผู้วิจัยได้อธิบายถึงวัตถุประสงค์ของการวิจัย วิธีการวิจัย ตลอดจนรายละเอียดสำคัญต่าง ๆของการวิจัยนี้ จนข้าพเจ้ามีความเข้าใจดีแล้ว
2. ผู้วิจัยรับรองว่าจะตอบคำถามต่าง ๆ ที่ข้าพเจ้าสงสัยด้วยความเต็มใจ ไม่ปิดบังซ่อนเร้นจนข้าพเจ้าพอใจ
3. ข้าพเจ้ามีสิทธิ์ที่จะบอกเลิกการเข้าร่วมโครงการวิจัยนี้เมื่อใดก็ได้ และเข้าร่วมโครงการวิจัยนี้โดยสมัครใจ และการบอกเลิกการเข้าร่วมการวิจัยนั้นไม่มีผลต่อคะแนนหรือเกรดของรายวิชา 5500210 ภาษาอังกฤษสำหรับวิชาชีพทางเภสัชศาสตร์ 2 (English for Pharmaceutical Profession II) ที่จะพึงได้รับต่อไป
4. ผู้วิจัยรับรองว่าจะเก็บข้อมูลของตัวข้าพเจ้าเป็นความลับ จะเปิดเผยได้เฉพาะในรูปที่เป็นข้อสรุปผลการวิจัย การเปิดเผยข้อมูลของตัวข้าพเจ้าต่อหน่วยงานต่าง ๆ ที่เกี่ยวข้องต้องได้รับอนุญาตจากข้าพเจ้า และจะกระทำได้เฉพาะกรณีจำเป็นด้วยเหตุผลทางวิชาการเท่านั้น
5. ข้าพเจ้าได้อ่านข้อความข้างต้นแล้ว มีความเข้าใจดีทุกประการ และได้ลงนามในใบยินยอมนี้ด้วยความเต็มใจ

ผู้ยินยอม.....ลงนาม

)……(

Appendix IV: Manual for dispensing assessment

Role-play assessment guidelines for administrator

Preparation

1. **Three rooms** are used for the assessment plus **one preparation room**. The rooms should be on the same floor for quick access and possibly the same room for morning and afternoon session. Contact staff at pharmacy faculty for this regard at least two weeks before the exam. Request **one staff for assistance**
2. At least three numbers of **equipment** are needed for recording the performance. Laptop, camera or iPad can be used to record the students' performance. One type of equipment with extra battery or chargeable station is set up in each exam room.
3. A **random list of students** can be prepared on Excel. Keep the original number of the students for record and **code them** starting with the section followed by their number in the section (This is the code to be used to name the video). Print out at least two copies of the students' random list, one for the raters to check and one for the administrator to call out and check the students after their assessment is done.
4. The students will always **role-play with the instructors who are not from their sections**. They will be assessed by their own instructor in the second observation. Below is the sample match of two and three observations.

	Instructors		
Students	First round	Second round	Third round
Section 1	Section 2	Section 1	Section 3
Section 2	Section 3	Section 2	Section 1
Section 3	Section 1	Section 3	Section 2

5. Three **bags** are needed to keep students' cellphone separately according to their sections.
6. Each rater is given two sets of documents, which are **case details** with sign stating initial symptoms to show to students and **copies of rubric** equal to number of students to be rated.
7. At least a week before the exam, show the video of how the assessment will be held and schedule a day to **show the students how the assessment will be done**, especially for those students who will have to press the record button on their own.

8. The administrator needs to have **a copy of the students' random list for each room, extra copies of the rubric, and the copies of self-assessment** (enough number for every student).

At the beginning

1. Arrive at least **20 minutes before the class time** to arrange the room and set up the equipment.
2. Once the students arrive, **show the video of how to record the performance** to students to make sure they fully understand the process and know how to do it correctly.
3. Ask students to **put their cellphones in the bag** according to their section.
4. **Distribute the drug label** to students and ask them to write their name, ID, section and their number in the section.
5. **Announce the first three persons** on the list of each section and inform them the room number and their rater's name.

During

1. **Let the students know of their turn at least ten minutes before** their schedule.
2. **Check the students** who finished their assessment and let them **fill in the self-assessment** with their section and their number in the section (the code for the video).
3. After the students finish the exam and the self-assessment, **let the students collect their cellphone and belongings** and leave the room without talking to students who have not taken the exam.

After

1. **Collect the rubrics** from the raters to put in Excel score sheet for first rating.
2. Make **copies of the videos** for the second rating in order to distribute copies of the video and rubric for the raters in the second round.
3. Each rater is given **two weeks to rate the second observation**.
4. **Fill in the scores** from the second rating in Excel sheet. If the score **differences between two raters are more than 3.5, find a third rater**.
5. Once all the scores are collected, **calculate the mean, minimum and maximum**. **Distribute the scores** in detail to the instructors.

Role-play assessment guidelines for raters

Rating explanation

1. Pharmaceutical science skills section is a dichotomous system as in yes or no. Please check whether the student fulfills the criteria or not. Although the suggested answers are provided, the rater does not need to concern the matching information of *impression, dispensing, instruction*, and *caution* to the suggested answer.
2. Language use and strategic competence sections are partial credit. Please rate the students according to the ability as trained in the workshop.
3. The rater can assess the amount and quality of detail on *instruction* and *caution* in *content appropriateness*, which is the criterion in the strategic competence section
4. When rating *initiating communication*, please assess the criterion on the level of welcoming patient to talk and while the student investigates the symptoms.

First observation (role-play)

1. Each rater will be given a list of students and rubric sheets. Please note the name of the student and their section if you assess the student who is not in your list. Instructors do role-play with students who are not in their section. If a student of your section shows up at the exam room, please inform him or her to seek the administrator.
2. Each student is allowed five minutes for role-play assessment. Any performance that goes over the limit of 5 minutes will not be considered. The rater can stop the performance once it is over time. Rater has 1-2 minutes to rate a student before the next student come in. Rater can signal the student to wait or to come in to the exam room when the rater is ready.
3. The sentence to state the initial symptom must not directly indicate that you (the rater) have the symptom, such as I have a headache. This is to assure that the students check for the patient's identity. The sentence should be neutral as in the following examples;
    - Do you have medicine for headache?
    - I want to buy drug for headache.
    - Can you suggest medication for headache?

4. Please bring your stapler and enough staples (around 30).
5. When the rater finished each student, please put the rubric form on the first page followed by the drug label before stapling them together.

Second observation

1. Rate the students in your section and return the scores to the administrator within two weeks.
2. Please write down the number of the students according to the number listed on the video and comment on the students' performance.
3. Contact the administrator if you have problem with the video or the rating.

Role-play assessment guidelines for students

1. Students are given **rubric to study at least two weeks before the exam**. The document can be downloaded from the website. Please discuss with your instructor if you have any questions regarding the detail of the rubric.
2. The students are required to **arrive 15 minutes earlier** than their usual class time. All students are required to be in the same preparation room until being informed otherwise to leave the room for the assessment.
3. All **cellphones must be turned in** to the administrator and can be collected when the students completed the exam.
4. Students are **randomly called out to the exam room**. Each student receives at least a ten-minute warning for their turn.
5. Students can **study on their materials in the room** while waiting for their turn.
6. Apart from the **drug label, only a pen or a pencil is allowed in the exam room**. All of your belongings can be left in the preparation room.
7. Students must **write their name, section, and their number in the section on the drug label** and submit it to their rater before leaving the exam room.
8. When students enter the exam room, students are given information about the patient, which are **sex, age and initial symptom**.
9. Once the student finishes the exam, students are required to **fill in the questionnaire** to assess their own performance and the administration. Please write down ID, the section and number in the section.
10. Students who **complete their self-assessment will receive the scores detail** from the raters comparing to their own assessment.
11. Important procedures to keep in mind regarding the pharmaceutical skills:
    - Elicit information about the patient's history (Who's the medication for? What are the symptoms? Any previous drug(s) use? Any family or social history that might contribute to the symptom?)
    - Ask about any drug allergies or underlying diseases
    - Give impression on the possible disease
    - Dispense drug(s) for the possible disease
    - Provide reasons for dispensing the medication
    - Explain information about the medication (What is it? Why this drug is offered? What's the dose? How is it being used? What's the caution?)
    - Suggestions on how to get better and or how to avoid the disease

Appendix V: Dispensing rubric for Thai pharmacy students

I: Pharmaceutical science skills (30% -- 15 points) Part I score:_____

Topic		Yes	No
Patient awareness	*Investigate and respond to patient's concerns and needs* (*Elicit patient questions, concerns, reasons for visit, current health condition and medications currently being taken)*	2	0
Allergy	*Ask patient about their allergies*	1	0
Underlying disease	*Ask patient about their underlying diseases*	1	0
Impression	*State the possible disease*	1	0
Dispensing	*Provide the name of the medication(s)*	2	0
Reason(s) for dispensing	*Provide reason(s) for dispensing the medication*	2	0
Instruction	*Provide instruction of the medication(s) dispensed*	2	0
Caution	*Provide detail on caution*	2	0
Verifying understanding	*Verify patient understanding of the medication usage*	2	0

II: Language Use (40% -- 20 points) Part II score: _____

Topic	Need improvement	Fair	Good	Excellent
Grammar *(intelligible)*	2	3	4	5
Pronunciation *(intelligible)*	2	3	4	5
Question type *(use appropriate open-ended and close-ended questions and not leading the patients)*	2	3	4	5
Word choice *(layman terms: easy-to-understand word choice for patient and clarify medical words if needed)*	2	3	4	5

III: Strategic Competence (30% -- 15 points) Part III score: _____

Topic	Need improvement	Fair	Good	Excellent
Voice: *tone, volume, pace, silence*	1		2	
Initiate communication: *greet patient warmly, identify patient's identity*	1	2	3	4
Conclude the encounter: *summarize information, ask if any questions arise, thank patient*	2	3	4	5
Non-verbal communication: *eye contact, gesture, posture, professional manner*	1	2	3	4

Total score: _____

Comments: _____

_____

Appendix VI: Drug label

Name ……..…… ID …… Section ……

Choose the appropriate drug label and complete the information in medication and direction section (bolded).

Drug label for oral medication

Date.....n/a.....	Date.....n/a.....
Patient's name.....n/a.....	Patient's name.....n/a.....
Medicine's name	**Medicine's name .....**
Indication.....n/a.....	Indication.....n/a.....
Take.....tab(s)/ cap(s)/ tsp(s)/ tbsp(s)	**Take..... tab(s)/ cap(s)/ tsp(s)/ tbsp(s)**
……. time(s) a day	**……. time(s) a day**
❒ before meals	❒ before meals
breakfast lunch dinner at bedtime	breakfast lunch dinner at bedtime
❒ after meals	❒ after meals
every.....hours as needed	every.....hours as needed .....

Drug label for topical medication

Date.....n/a.....	Date.....n/a.....
Patient's name.....n/a.....	Patient's name.....n/a.....
Medicine's name	**Medicine's name .....**
Indication.....n/a.....	Indication.....n/a.....
Directions.....	**Directions.....**
.....	**.....**
Dosage.....n/a.....	Dosage.....n/a.....

Appendix VII: Questions for raters

Overall rubric usage

- Did you feel comfortable using the rubric?
- Which section was the easiest to rate?
- Which section was the hardest to rate?
- Does the rubric specify adequate information for the rating? If not, what should be added in?
- Can this rubric be used to assess and differentiate the student's performance in dispensing drugs in English?
- Are there any criteria you think should be added in the rubric?
- Are there any criteria you think it's unnecessary?

Topic		**Yes**	**No**
Patient awareness	*Investigate and respond to patient's concerns and needs* (*Elicit patient questions, concerns, reasons for visit, current health condition and medications currently being taken)*	2	0
Allergy	*Ask patient about their allergies*	1	0
Underlying disease	*Ask patient about their underlying diseases*	1	0
Impression	*State the possible disease*	1	0
Dispensing	*Provide the name of the medication(s)*	2	0
Reason(s) for dispensing	*Provide reason(s) for dispensing the medication*	2	0
Instruction	*Provide instruction of the medication(s) dispensed*	2	0
Caution	*Provide detail on caution*	2	0
Verifying understanding	*Verify patient understanding of the medication usage*	2	0

Pharmaceutical science knowledge

- Were you confident rating the students in this section?
- Do you think the scale for each criterion is appropriate?

- Are there any criteria you found it hard to rate the student?
- Any recommendation?
- Do you think the score level is appropriate? If not, how would you like it to be?

For each criterion, please specify your confidence rating it as a non-pharmacist.

Language use

- Were you confident rating the students in this section?
- Do you think the scale for each criterion is appropriate?
- Are there any criteria you found it hard to rate the student?
- Do you think the score level is appropriate? If not, how would you like it to be?
- Any recommendation?

Strategic competence

- Were you confident rating the students in this section?
- Do you think the scale for each criterion is appropriate?
- Are there any criteria you found it hard to rate the student?
- Do you think the score level is appropriate? If not, how would you like it to be?
- Any recommendation?

Appendix VIII: MFRM result

Criteria	Patient awareness		Allergy		Underlying disease	
Response	No	Yes	No	Yes	No	Yes
Score	0	2	0	1	0	1
Total	4	292	3	293	9	287
%	1	99	1	99	3	97
Avg Measure	2.26	3.04	4.30	5.13	3.01	4.01
Exp Measure	1.56	3.05	4.15	5.13	3.03	4.01
Mean square	1.3	1.1	.9	1.0	.7	1.0

Criteria	Impression		Dispensing		Reason for dispensing	
Response	No	Yes	No	Yes	No	Yes
Score	0	1	0	2	0	2
Total	44	252	6	296	16	280
%	15	85	2	98	5	95
Avg Measure	1.57	2.25	1.43	2.84	1.03	2.30
Exp Measure	1.31	2.30	1.36	2.84	.86	2.31
Mean square	1.2	1.1	.5	1.0	.9	1.1

Criteria	Instruction		Caution		Verify understanding	
Response	No	Yes	No	Yes	No	Yes
Score	0	2	0	2	0	2
Total	1	295	22	274	67	229
%	0	100	7	93	23	77
Avg Measure	1.92	3.76	1.37	2.08	.70	1.25
Exp Measure	2.26	3.76	.69	2.13	.03	1.45
Mean square	.2	1.0	3.8	1.2	6.8	1.8

Appendix IV: The adjusted dispensing rubric for Thai pharmacy students

Role-play assessment rubric (50 points)

I: Pharmaceutical Skills (20%: 10 points) **Part I score: _____**

Topic		Yes	No
Chief Complaint	*Elicit patient questions, concerns, and reasons for visit*	1	0
History & Present illness	*Ask patient about their current health condition and medications currently being taken*	1	0
Allergy	*Ask patient about their allergies*	1	0
Underlying disease	*Ask patient about their underlying diseases*	1	0
Family & Social history	*Ask patient about their family and social history*	1	0
Impression	*State the possible disease*	1	0
Dispensing	*Provide the name of the medication(s)*	1	0
Reason(s) for dispensing	*Provide reason(s) for dispensing the medication*	1	0
Instruction	*Provide instruction of the medication(s) dispensed*	1	0
Caution	*Provide caution(s)*	1	0

II: Language Use (40%: 20 points) **Part II score: _____**

Topic	Need improvement	Fair	Good	Excellent
Grammar *(intelligible)*	2	3	4	5
Pronunciation *(intelligible)*	2	3	4	5
Question type *(use appropriate open-ended and close-ended questions and not leading the patients)*	2	3	4	5
Word choice *(layman terms: easy-to-understand word choice for patient and clarify medical words if needed)*	2	3	4	5

III: Strategic Competence (40%: 20 points) **Part III score: _____**

Topic	Need improvement	Fair	Good	Excellent
Voice: *tone (1), volume (1), pace (1), silence (1)*	1	2	3	4
Initiate communication: *greet warmly and welcoming the patient (2), identify patient's identity (2)*	1	2	3	4
Conclude the encounter: *summarize information (2), ask if any questions arise (1), thank patient (1)*	1	2	3	4
Non-verbal communication: *eye contact (1), gesture (1), posture (1), professional manner (1)*	1	2	3	4
Content Appropriateness: *enough information provided on instruction (2) and caution (2)*	1	2	3	4

Total score: _____→_____

Comments:

_____

REFERENCES

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*: Ernst Klett Sprachen.

Aliakbari, M., & Boghayeri, M. (2014). A Needs Analysis Approach to ESP Design in Iranian Context. *Procedia - Social and Behavioral Sciences, 98*, 175-181. doi:https://doi.org/10.1016/j.sbspro.2014.03.404

Arnó-Macià, E., & Mancho-Barés, G. (2015). The role of content and language in content and language integrated learning (CLIL) at university: Challenges and implications for ESP. *English for Specific Purposes, 37*, 63-73.

Austin, A. E. (1990). Faculty cultures, faculty values. *New directions for institutional research, 1990*(68), 61-74.

Avineri, N., Londe, Z., Hardacre, B., Carris, L., So, Y., & Majidpour, M. (2010). Language Assessment as a System: Best Practices, Stakeholders, Models, and Testimonials. *Issues in Applied Linguistics, 18*(2).

Bachman, L., & Damböck, B. (2018). *Language Assessment for Classroom Teachers*: Oxford University Press.

Bachman, L., & Palmer, A. (2012). Language assessment in practice. In: Oxford: Oxford University Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing*: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1): Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2012). *Language assessment in practice*: Oxford: Oxford University Press.

Basturkmen, H., & Shackleford, N. (2015). How content lecturers help students with language: An observational study of language-related episodes in interaction in first year accounting classrooms. *English for Specific Purposes, 37*, 87-97.

Bijani, H., & Khabiri, M. (2017). Investigating the effect of training on raters' bias toward test takers in oral proficiency assessment: A FACETS analysis. *Journal of Asia TEFL, 14*(4), 687-702. doi:10.18823/asiatefl.2017.14.4.7.687

Bolton, K. (2008). English in Asia, Asian Englishes, and the issue of proficiency. *English Today, 24*(2), 3.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*: Lawrence Erlbaum Associates Publishers.

Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing, 20*(1), 89-110. doi:10.1191/0265532203lt245oa

Brennan, M., & Naerssen, M. v. (1989). Language and content in ESP. *ELT journal, 43*(3), 196-205.

Brown, H. D., & Abeywickrama, P. (2010). *Language Assessment Principles and Classroom Practices* (2nd ed.). New York: Pearson Education.

Brown, J. D. (1996). *Testing in Language Programs*. Upper Saddle River, New Jersey: Prentic Hall Regents.

Brunfaut, T. (2014). Language for Specific Purposes: Current and future issues. *Language Assessment Quarterly, 11*(2), 216-225.

Butler, S. M., & McMunn, N. D. (2006). *A Teacher's Guide to Classroom Assessment: Understanding and Using Assessment to Improve Student Learning*. San Francisco, CA: Jossey-Bass.

Byrnes, H. (2008). Assessing content and language. In *Encyclopedia of language and education* (pp. 2182-2197): Springer.

Cameron, D. (2012). *Working with Spoken Discourse*. London: Sage Publications.

Chalhoub-Deville, M. (2003). Second Language Interaction: Current Perspectives and Future Trends. *Language Testing, 20*(4), 369-383.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting : a guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.

Collett, J., Rees, J., Mylrea, S., & Crowther, I. (1994). Performance based assessment in pharmacy education. *International Journal of Pharmacy Practice, 3*(1), 38-41.

Douglas, D. (2000). *Assessing Languages for Specific Purposes*: Cambridge University Press.

Douglas, D. (2001). Three problems in testing language for specific purposes: Authenticity, specificity and inseparability. In *Studies in Language Testing:*

Experimenting with uncertainty: essays in Honour of Alan Davies. Cambridge: Cambridge University Press.

Douglas, D. (2013). *ESP and Assessment* (B. Paltridge & S. Starfield Eds. First ed.): John Wiley & Son, Inc.

Douglas, D., & Myers, R. (2000). Assessing the communication skills of veterinary students: Whose criteria? In A. J. Kunnan (Ed.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida* (pp. 60-81). Cambridge: Cambridge University Press

Eckes, T. (2009). Many-facet Rasch measurement. *Reference supplement to the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment*, 1-52.

Edwards, N. (2000). Language for business: effective needs assessment, syllabus design and materials preparation in a practical ESP case study. *English for Specific Purposes, 19*(3), 291-296. doi:https://doi.org/10.1016/S0889-4906(98)00029-5

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.

Elder, C., & McNamara, T. (2016). The hunt for “indigenous criteria” in assessing communication in the physiotherapy workplace. *Language Testing, 33*(2), 153-174. doi:10.1177/0265532215607398

Elder, C., McNamara, T., Kim, H., Pill, J., & Sato, T. (2017). Interrogating the construct of communicative competence in language assessment contexts: What the non-language specialist can tell us. *Language & Communication, 57*, 14-21. doi:https://doi.org/10.1016/j.langcom.2016.12.005

Elder, C., Pill, J., Woodward‐Kron, R., Mcnamara, T., Manias, E., Webb, G., & Mccoll, G. (2012). Health Professionals' Views of Communication: Implications for Assessing Performance on a Health‐Specific English Language Test. *TESOL quarterly, 46*(2), 409-419.

Flowerdew, J. (2002). Ethnographically inspired approaches to the study of academic discourse. In J. Flowerdew (Ed.), *Academic discourse* (pp. 235-252): Pearson Education.

Frey, B. (2014). *Modern classroom assessment*. Thousand Oaks, CA: SAGE Publications.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Oxon: Routledge.

Fulcher, G., Davidson, F., & Kemp, J. (2011). Effective Rating Scale Development for Speaking Tests: Performance Decision Trees. *Language Testing, 28*(1), 5-29.

Gillette, C. c. g. w. e., Rudolph, M., Rockich-Winston, N., Stanton, R., & Anderson Jr, H. G. (2017). Improving Pharmacy Student Communication Outcomes Using Standardized Patients. *American journal of pharmaceutical education, 81*(6), 1-8. Retrieved from http://ezproxy.car.chula.ac.th/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=125158848&site=eds-live

Graham, J. G., & Beardsley, R. S. (1986). English for specific purposes: Content, language, and communication in a pharmacy course model. *TESOL quarterly, 20*(2), 227-245.

Gray, D. E. (2013). *Doing research in the real world*: Sage.

Green, A. (2014). *Exploring language assessment and testing: Language in action*. New York: Routledge.

Grice, G. R. g. s. e., Gattas, N. M., Prosser, T., Voorhees, M., Kebodeaux, C., Tiemeier, A., . . . Juang, P. (2017). Design and Validation of Patient-Centered Communication Tools (PaCT) to Measure Students' Communication Skills. *American journal of pharmaceutical education, 81*(8), 33-48. doi:10.5688/ajpe5927

Henry, G. T. (2008). Practical Samping. In L. Bickman & D. J. Rog (Eds.), *The SAGE handbook of applied social research methods*: Sage Publications.

Hill, K., & McNamara, T. (2012). Developing a Comprehensive, Empirically Based Research Framework for Classroom-Based Assessment. *Language Testing, 29*(3), 395-420.

Huhta, M. (2010). Language and communication for professional purposes needs analysis methods in industry and business and their yield to stakeholders.

Hymes, D. (1964). Introduction: Toward Ethnographies of Communication. *American Anthropologist, 66*(6_PART2), 1-34.

Hyvärinen, M.-L., Tanskanen, P., Katajavuori, N., & Isotalus, P. (2012). Evaluating the use of criteria for assessing profession-specific communication skills in pharmacy. *Studies in Higher Education, 37*(3), 291-308.

Jacoby, S. (1998). *Science as performance: Socializing scientific discourse through the conference talk rehersal.* (Doctor of Philosophy). University of California, Los Angeles,

Jacoby, S., & McNamara, T. (1999). Locating competence. *English for Specific Purposes, 18*(3), 213-241.

Johnson, R. C., & Riazi, A. M. (2017). Validation of a locally created and rated writing test used for placement in a higher education EFL program. *Assessing Writing, 32*, 85-104.

Jones, W. A. (2011). Variation among academic disciplines: An update on analytical frameworks and research. *Journal of the Professoriate, 6*(1), 9-27.

Kane, M. (2006). Content-related validity evidence in test development. *Handbook of test development, 1*, 131-153.

Kimberlin, C. L. (2006). Communicating with patients: skills assessment in US colleges of pharmacy. *American journal of pharmaceutical education, 70*(3), 67.

Layder, D. (2012). *Doing excellent small-scale research*: Sage.

Lee, J. J. (2007). The shaping of the departmental culture: Measuring the relative influences of the institution and discipline. *Journal of Higher Education Policy and Management, 29*(1), 41-55.

Lim, J. M. A. (2013). Rubric-referenced oral production assessments: perceptions on the use and actual use of rubrics in oral production assessments of high school students of St. Scholastica's College, Manila. *Language Testing in Asia, 3*(1), 4.

Linnenluecke, M. K., Russell, S. V., & Griffiths, A. (2009). Subcultures and sustainability practices: The impact on understanding corporate sustainability. *Business Strategy and the Environment, 18*(7), 432-452.

Llosa, L. (2008). Building and Supporting a Validity Argument for a Standards-Based Classroom Assessment of English Proficiency Based on Teacher Judgments. *Educational Measurement: Issues and Practice, 27*(3), 32-42.

Long, M. H. (2005). *Second language needs analysis*: Cambridge University Press.

Luka, I. (2008). Students and the educator's co-operation as a means of development of students' ESP competence.

Lumley, T., & McNamara, T. F. (1993). *Rater Characteristics and Rater Bias: Implications for Training*. Retrieved from https://ezp.lib.unimelb.edu.au/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED365091&site=eds-live&scope=site

Lundberg, K. M. (2008). Promoting self-confidence in clinical nursing students. *Nurse educator, 33*(2), 86-89.

Lynch, B. K., & McNamara, T. F. (1998). Using G-theory and Many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing, 15*(2), 158-180. doi:10.1177/026553229801500202

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.

Manias, E., & McNamara, T. (2016). Standard setting in specific-purpose language testing: What can a qualitative study add? *Language Testing, 33*(2), 235-249. doi:10.1177/0265532215608411

Maxwell, J. A. (2008). Designing a Qualitative Study. In L. Bickman & D. J. Rog (Eds.), *The Sage handbook of applied social research methods*.

McNamara, T. (1996). *Measuring second language performance*. London ;: Longman.

McNamara, T., & Knoch, U. (2012). The Rasch wars: The emergence of Rasch measurement in language testing. *Language Testing, 29*(4), 555-576. doi:10.1177/0265532211430367

McNamara, T., Knoch, U. & Fan, J. (2019). *Fairness, Justice and Language Assessment*. Oxford: Oxford University Press.

McNamara, T. F., & Roever, C. (2006). *Language testing : the social dimension*. Malden, MA :: Blackwell Pub.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.

O'Sullivan, B. (2012a). Assessment issues in languages for specific purposes. *The Modern Language Journal, 96*(s1), 71-88.

O'Sullivan, B. (2012b). A Brief History of Language Testing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment*. New York: Cambridge University Press.

O'Hagan, S., Pill, J., & Zhang, Y. (2016). Extending the scope of speaking assessment criteria in a specific-purpose language test: Operationalizing a health professional perspective. *Language Testing, 33*(2), 195-216. doi:10.1177/0265532215607920

Oosterhof, A. (2003). *Developing and using classroom assessments* (3rd ed. ed.). Upper Saddle River, N.J. :: Merrill/Prentice Hall.

Paltridge, B. (2006). *Discourse analysis : an introduction*. London Continuum.

Papageorgiou, S., & Tannenbaum, R. J. (2016). Situating standard setting within argument-based validity. *Language Assessment Quarterly, 13*(2), 109-123.

Parkhurst, C. (1994). Assessing and Improving Students' Verbal Communication Abilities in Pharmacy Courses. *American journal of pharmaceutical education, 58*(1), 50-55.

Pill, J. (2013). *What doctors value in consultations and the implications for specific-purpose language testing.* (Doctor of Philosophy). The University of Melbourne.,

Pill, J. (2016). Drawing on indigenous criteria for more authentic assessment in a specific-purpose language test: Health professionals interacting with patients. *Language Testing, 33*(2), 175-193.

Purpura, J. E., Brown, J. D., & Schoonen, R. (2015). Improving the Validity of Quantitative Measures in Applied Linguistics Research1. *Language Learning, 65*(S1), 37-75. doi:10.1111/lang.12112

Ryan, J. J., & Lopez, S. J. (2001). Wechsler Adult Intelligence Scale-III. In I. D. William & M. Hersen (Eds.), *Understanding Psychological Assessment*: Springer.

Schwartzman, E., Chung, E. P., Sakharkar, P., & Law, A. V. (2013). Instruction and assessment of student communication skills in US and Canadian pharmacy curricula. *Currents in Pharmacy Teaching and Learning, 5*(6), 508-517.

Sibbald, D. (1998). Oral clinical skills examination: An innovative reinforcing strategy for nonprescription medication courses. *American journal of pharmaceutical education, 62*(4), 458.

Simon, M., Ercikan, K., & Rousseau, M. (2012). *Improving large-scale assessment in education: Theory, issues, and practice*: Routledge.

Stahl, J., & Lunz, M. (1992). *A comparison of generalizability theory and multi-faceted Rasch measurement.* Paper presented at the Midwest Objective Measurement Seminar, University of Chicago.

Tsou, W. (2009). Needs-based curriculum development: A case study of NCKU's ESP program. *Taiwan International ESP Journal, 1*(1), 77-95.

Weir, C. J. (2005). *Language Testing and Validation: An Evidence-based Approach*. Basingstoke: Palgrave Macmillan UK.

Wiriyachitra, A. (2002). English-language teaching and learning in Thailand in this decade. *Thai TESOL Focus, 15*(1), 4-9.

Woodward-Kron, R., & Elder, C. (2016). A comparative discourse study of simulated clinical roleplays in two assessment contexts: Validating a specific-purpose language test. *Language Testing, 33*(2), 251-270.

VITA

NAME	Sasithorn Limgomolvilas
PLACE OF BIRTH	Bangkok
INSTITUTIONS ATTENDED	Chulalongkorn University San Francisco State University
PUBLICATION	Limgomolvilas, S. (2016). Idea Sharing: Facilitating Presentation Skills through Integrated Face-to-Face and Distance Learning, as well as Forum Instruction. NIDA Journal of Language and Communication, 21 (28), 29-42. Limgomolvilas, S., Janpugdee, P., Wudthayagorn, J., & Pugsee, P. (2015). The Roles and Availability of Specialized Word Lists in EAP Learning and Teaching for Thai Engineering Students. Language Education and Acquisition Research Network (LEARN) Journal, 8 (2), 30-42. Limgomolvilas, S. (2015). Review: Assessment Myths. PASAA, 50, 177-180.