Summer 7-16-2019

Placement Testing Instruments for Modality Streams in an English Language Program

Aimee Hilgers
hilgersa@mnstate.edu

Placement Testing Instruments for Modality Streams in an English Language Program

A Plan B Project Presented to
the Graduate Faculty of
Minnesota State University Moorhead

By

Aimee Elisabeth Hilgers

In Partial Fulfillment of the
Requirements for the Degree of
Masters of Arts in
Teaching English as a Second Language

June 2019

Moorhead, Minnesota

Dedication

This project is dedicated to my family, Ross, Dash, Bly, and Ives, who have supported and encouraged me throughout this process. It is also dedicated to my parents, William J. Best and M. Judith Best, for modeling lifelong learning as well as providing their support.

Acknowledgements

Special gratitude is extended to my professor and advisor, Dr. Linda Houts-Smith, for her support, experience, and expertise throughout the process of writing this project. I would also like to express my sincere appreciation to the other committee member, Dr. Brian Smith, for his support and constructive advice.

Table of Contents

List of Tables

Abstract

The purpose of this study was to determine placement testing instruments for the modality streams in the English Language Program (ELP) at Minnesota State University Moorhead (MSUM). The study involved research into and consideration of types and purposes of tests. A placement test was determined to be the appropriate test type for the program's current student placement needs. An investigation was performed regarding characteristics placement tests need to possess for our program. With these characteristics guiding research, placement test options for the modality streams were reviewed, including standardized tests and in-house assessments. This project determined that in-house placement testing instruments for the modality streams best suit our program's needs and are recommended. Future phases of the study are also suggested.

Chapter 1: Introduction

Humans long to express themselves for various reasons, and spoken and written language are the primary tools we select for that expression. English is the language of choice for many people around the world; English is used by numerous countries for professional, educational, and personal communication purposes. Therefore, people worldwide choose to learn English or improve their developing English skills to achieve one or more of these goals. Those interested in studying English, especially non-native speakers of English (NNSEs), can access educational training across the world through public and private organizations and institutions that offer English language education. In the United States, English training is offered at numerous educational institutions and corporations, so learners have many choices about where they may obtain their learning. Minnesota State University Moorhead (MSUM) is one of the educational institutions that offers English language education.

MSUM's English Language Program

The English Language Program (ELP), launched in the fall of 2017, is housed within the History, Language, Critical Race, and Women's Studies Department at MSUM, an institution in the state university system in Minnesota. The ELP is accredited through MSUM's accreditation by the Higher Learning Commission and membership in the North Central Association of Colleges and Schools ("Teaching English," 2018). The MSUM student population numbers just over 5,800 ("MSUM facts," n.d.), and within this student population, international and resident NNSEs are groups this ELP serves. Although student enrollment numbers vary each semester, the most recent published data regarding MSUM's international student population indicates a

total of 414 international students from 58 countries, Nepal and Nigeria being the countries with the highest numbers of students attending ("International student population," 2016). MSUM's website provides no data regarding numbers of resident NNSEs attending MSUM, nor does it separate NSE from NNSE international students. Students in the ELP at MSUM may be either matriculated or non-matriculated. Matriculated international students may take ELP courses in addition to another major; when federal approval for our ELP as a stand-alone English training program is gained, non-matriculated students may enroll in the program. Resident NSEs may be either matriculated or non-matriculated.

The ELP consists of thirty-five English language courses at five proficiency levels with the curriculum focusing on seven modality streams: pronunciation and speaking, listening, reading, writing, syntax, vocabulary, and culture (see Appendix A). Within the ELP, two university awards, the Certificate of Proficiency in International English (see Appendix B) and the Certificate of Academic English Proficiency, are available for students to pursue (see Appendix C). The Certificate of Proficiency in International English (CPIE) acknowledges English skills that support students in pursuit of professional careers in fields such as healthcare services, tourism, and hospitality. The Certificate of Academic English Proficiency (CAEP) certifies students' academic English language skills for academic study at higher education institutions and professional careers, such as those in business and science. Regardless of whether students wish to pursue one of the certificates, they may take courses in the program to support their English language development.

The ELP courses are housed within the Teaching English as a Second Language division of the Languages unit in the Department of History, Languages, Critical Race, and Women's Studies (HLCRWS), which is in the College of Humanities and Social Sciences. Because the

courses are tied to certificates, they went through MSUM's full curriculum approval process in the years preceding institutional approval. This process entailed the initiator consulting with the HLCRWS Department regarding the curricular proposal and revisions before the department chair consulted with affected units and ensured forms were completed. The college dean then fully vetted the proposal regarding concerns such as enrollment and financial viability before the proposal was sent to the Registrar's Office for details with issues such as cataloging entries, titles, and numbering. Relevant committees and the faculty senate reviewed the proposal before it went before a Meet and Confer meeting for consideration. Through Academic Affairs, the Registrar's Office finalized details as well as submitted paperwork to Minnesota State and notified the MSUM community when the ELP was approved ("Registrar's Office," n.d.).

Statement of the Problem

When the program was launched in the fall of 2017, the program's curriculum had been designed, admission details regarding proficiency had been designated, federal approval was in the submission process, and I had been hired full-time to both coordinate and teach in the program. During the first semester, only a few courses were scheduled, and no students were pursuing the certificates.

Although many essential matters had been finalized before launch, several details have yet to be determined two years later, including placement testing instruments for the program's modality streams that deal with language instruction. Several students are currently pursuing the Certificate of Academic English Proficiency, graduate teaching assistants seeking an MA in TESL have worked with students in the ELP courses, and approval from the federal government to enroll non-matriculated students who won't have undergone language testing as part of the admissions process is highly anticipated. Thus, the need to determine placement testing tools has

become more urgent. In my role as program coordinator, it falls to me to determine the placement instruments for the program. Therefore, I conducted this research project with the following research question in mind:

- What are appropriate placement testing instruments for the modality streams in the English Language Program at Minnesota State University Moorhead?

Scope of the Project

Because consideration was given to our program's features and needs regarding placement testing, this project applies only to MSUM's ELP. The study is not intended as a recommendation for other ELPs. However, other programs may find the process and literature relevant to their own process of inquiry into a similar question within their own context. Additionally, this report does not include the placement instruments themselves, nor does it address the efficiency or success of selected instruments; no participants were included in this study. Finally, this study only focuses on the search for testing instruments for the language modality streams in MSUM's program. Those streams include the reading, listening, writing, vocabulary, syntax, and pronunciation streams. A placement test is not needed for the culture stream since the courses in this stream focus on content rather than language learning.

Organization of the Report

This chronological report records my process of using a qualitative approach to determine appropriate placement testing instruments for the modality streams, presents the results of my research into placement testing instruments for the ELP at MSUM, and recommends appropriate instruments for our context.

My first step, described in Chapter 2, was to consider the types and purposes of tests and

select the appropriate type for our program. After I determined that placement tests were needed for all modality streams aside from the culture stream, I turned my attention to test content, forms, and sources, which I discuss in Chapter 3. Chapter 4 describes standardized test options, and Chapter 5 in-house testing instruments as sources of test items. In the conclusion in Chapter 6, I recommend MSUM construct in-house placement testing instruments and describe the next phases of the placement instrument process for our ELP's courses.

Chapter 2: Types and Purposes of Tests

Language testing is conducted for various purposes. Language testing is typically conducted in ELPs to provide information to the test takers and test givers. When considering which type of test to conduct, it is essential to evaluate the types of tests based on their purpose and determine which language tests allow test givers to find appropriate assessments for their specific purposes. Jamieson, Wang, and Church (2013) remarked that when considering purpose, we need to ask "Why are we giving a test?" and added that "purpose is related to the decisions that will be made based on the test's score. It represents a fundamental consideration before selecting or developing a test, administering it, and interpreting its scores" (p. 289). Common types of language tests that are vital to consider for our ELP include "proficiency tests, achievement tests, diagnostic tests, and placement tests" (Hughes, 2003, p.11). Examining these main assessment types and purposes will begin the process of determining the appropriate testing instruments for MSUM's ELP.

One type of language test, the achievement test, is designed to measure objectives in language courses (Hughes, 2003), to determine "student learning progress" (Kress, 2008, p. 323), as well as decisions related to student eligibility for entering and exiting programs (Kress, 2008). Emphasizing the importance of feedback from achievement testing, Brown and Hudson (1998) remarked that the purpose of feedback from achievement tests that measure course objectives will indicate "what the students have learned or learned how to do in the course" (p. 669).

Regarding the significance of achievement tests, Nunan (2015) concurred: "Providing evidence of the achievement of course goals is a fundamental purpose of assessment" (p. 173).

Achievement tests may be of two types: final achievement tests based on the course syllabus, text, or other course materials or progress achievement tests based on measuring the progress students are making in a course (Hughes, 2003; Nunan, 2015). Regardless of whether the test measures the final or an in-progress stage in learning, both types of achievement tests follow the course objectives or syllabus. Teachers typically design and implement achievement tests for use in their classrooms (Hughes, 2003).

Diagnostic tests form a second category of test types to consider. The purpose of diagnostic tests is to discover learners' strong and weak areas, in this specific case, English language skills, and detect what individuals still need to learn (Brown & Hudson, 1998; Hughes, 2003; Kress, 2008; Kim, 2015; Nunan, 2015). Brown and Hudson (1998) noted that diagnostic test feedback should relate to every objective in courses. Because of this purpose, diagnostic assessments are usually administered before instruction although they are also informally conducted during the instruction period to assess the progress of student learning and to inform instructional decisions (Kress, 2008; Nunan, 2015). Hughes (2003) noted that creating and analyzing diagnostic tests that measure broad categories of strengths and weakness, such as in speaking or reading, can be achieved somewhat easily; however, evaluating nuances within categories is more challenging. Nunan (2015) clarified this point regarding diagnostic test decisions: "you need to decide exactly what aspects of learner language you wish to diagnose, for example, the four skills (listening, speaking, reading, writing), the language systems (pronunciation, vocabulary, and grammar), or functional aspects of language use" (p. 173). Hughes (2003) added that few quality diagnostic tests exist, partly due to their unreliability,

although more are beginning to be developed with the aid of computers. Kim (2015) also observed that diagnostic tests are beginning to receive more attention in testing second languages due to their usefulness to deliver "diagnostic feedback to stakeholders and ultimately promote language learning" (p. 227).

A third test type to examine is the proficiency test. The purpose of proficiency tests is to measure language ability in test takers without consideration of their language training or educational background (Christopher, 1993; Hughes, 2003; Kokhan, 2012). Because this test type is designed to measure language proficiency, a discussion of what *proficiency* means is relevant. Hughes (2003) noted that some proficiency tests have a specific purpose, such as establishing whether a NNSE knows English well enough to enroll in a university that instructs in English. In fact, language proficiency testing, whether standardized or developed in-house, is commonplace in educational institutions (James & Templeman, 2009). An example of this type of English proficiency test frequently used for admission to institutions of higher education is the Test of English as a Foreign Language (TOEFL); in fact, Educational Testing Service (ETS), who produces the TOEFL, described it as "the leading English-language test for study, immigration and work" ("The TOEFL," n.d.). TOEFL scores are used at MSUM as one option for international students to use to indicate their English language proficiency level as a part of their admissions process ("International freshman admission," n.d.). If prospective students use the TOEFL to demonstrate this proficiency, MSUM requires a minimum score of 61 on the TOEFL iBT, 173 on the computer test, or 500 on the paper-based test ("International freshman admission," n.d.). Language tasks on the TOEFL may include written and spoken responses based on lectures test-takers have listened to and read ("The TOEFL," n.d.). TOEFL states that "in order for the TOEFL iBT test to measure how well you read, listen, speak, and write in

English, and how well you use these skills together, you will be asked to integrate these skills" ("The TOEFL," n.d.). Because integrated language tasks are included in the TOEFL test, it is clearly a test that provides results of the test-taker's overall proficiency.

While proficiency tests, such as the TOEFL, are commonly used by institutions to assess English proficiency for study, another type of proficiency test is based on internationally known criteria and "for them the concept of proficiency is more general" (Hughes, 2003, p. 12). Test takers demonstrate particular language abilities; for instance, the Cambridge First Certificate in English examination (FCE) corresponds to "levels in the ALTE (Association of Language Testers in Europe) framework, which draws heavily on the work of the Council of Europe" (Hughes, 2003, p.12). Those interested in test results, such as employers, can compare test takers' scores with the "set of specified abilities" (Hughes, 2003, p. 12).

A final assessment consideration is the placement test. Green (2018) reported that placement is one of the most common reasons language tests are constructed. The purpose of placement tests is to identify students' placement in an appropriate level or stage in an educational program based on their language abilities (Brown, 1989; Christopher 1993; Hughes, 2003; Harrington & Carey, 2009; Crusan, 2011; Banegas, 2013; Jamieson, Wang, & Church, 2013; Crusan, 2014; Ling, Wolf, Cho, & Wang, 2014; Nunan, 2015; Shin & Lidster, 2017). In other words, placement tests are designed to match students' abilities with a specific course or level in a program, and "typically they are used to assign students to classes at different levels" (Hughes, 2003, p. 16).

Test Type Selection

Evaluating the purposes of the types of tests regularly used in English language program settings leads to the conclusion that a placement test is the type needed for our program since we

need to identify the specific proficiency level in each modality that is suitable for every learner. Achievement tests and diagnostic tests should immediately be eliminated from further consideration as they are designed for use in classrooms, so they are not appropriate for course placement before students enter the classroom.

Although students' scores on the TOEFL test, a similar proficiency test, or an equivalent measure of proficiency can indicate whether students' have English proficiency to study in an academic setting in which English is the dominant language of instruction, a proficiency test is designed for universities to make admissions decisions, not to place students in specific courses in particular programs. The proficiency test scores aid in determining a learner's overall English abilities, but the test's purpose does not focus on placement concerns.

Although some educational institutions use proficiency tests to place students accurately, the research is mixed regarding the effectiveness of that method (Christopher 1993; Kokhan 2013; Ling et al., 2014). Kokhan (2013) presented research suggesting that accurate ELP course placement may occur using proficiency tests used for university admittance, but the tests do not measure what instructors "wanted their students to know or what they taught to their students" in the ELP courses (p. 471). Furthermore, Morante (as cited in Kokhan, 2013) suggested that proficiency tests used in admissions can identify levels among more proficient students, but they do not discern levels among less proficient students as do placement tests. Programs should consider that the placement test be appropriate for the range of their specific students' proficiencies (Brown, 1989; Parker, 2017).

Furthermore, unlike high-stakes proficiency tests that determine whether a student can gain admission to an educational institution, placement tests are generally considered low-stakes tests. According to Hughes (2003):

> Typically, all that is asked of language school placement tests is that they assign people to a level in that school. If people are misplaced by the test, they can usually easily be moved to a more appropriate class; provided that not too many such moves are called for, this is not a problem. (p.186)

For our ELP at MSUM, using a high-stakes test, such as TOEFL, for placement in courses in our modality streams could cause too much movement in course placement because TOEFL items integrate skills rather than separate skills by modality. Therefore, we would not be able to discern students' individual needs in each modality using the TOEFL for placement.

Kokhan (2012) explained why "too many such moves" is problematic: "the main objective of placement tests is reduction to an absolute minimum the number of students who may be at risk of failing some courses or not getting their academic degrees because of their poor language skills" (p. 292). Crusan (2011) echoed these concerns about supporting students' academic success with correct placement and later noted in 2014 the ethical aspect of placement testing, which should ensure fairness and correctness when placing students. To aid student success, James and Templeman (2009) noted that ELPs should provide accurate placement for students with the use of a placement testing instrument. ESL students are often incorrectly placed in courses without the use of an accurate assessment instrument, and these errors indicate the difficult task encountered by institutions of higher education to determine a suitable placement system (Banegas, 2013; Shin & Lidster, 2017). Thus, it is essential that programs determine correct English course placement for the success of the students' educational process (Banegas, 2013; Shin & Lidster, 2017). In light of these findings, MSUM's program must determine accurate placement instruments to ensure appropriate course placement for its students and limit

its use of TOEFL to admissions purposes. Incorrect placement could cause these same academic issues for our students.

Essential Characteristics of Useful Placement Tests

Because no placement testing methods exist that are universally determined to be the most effective, the specific educational context must drive the development of placement tools and procedures (Green, 2018). However, useful tests will include characteristics that align with principles of sound assessment: validity, reliability, practicality, authenticity, and beneficial backwash (Hughes, 2003; Brown and Abeywickrama, 2010).

Validity. In addition to using a test designed for placement rather than proficiency to achieve accurate student placement, a useful placement test must have validity. Validity or "construct validity" (Hughes, 2003, p. 26) refers to how accurately a test measures what it aims to assess (Hughes, 2003). Johnson and Riazi (2017) also stressed the need for validation so the test measures what is taught in a program's courses. For instance, a test might be designed to measure students' knowledge of information taught in an English syntax course or English writing skills in a composition course. Validity in testing can be attempted through various means, but evidence is needed to verify a test's validity in construction as well as its validity in scoring (Hughes, 2003). Hughes (2003) asserted the following regarding placement test validity:

> placement tests attempt to predict the most appropriate class for any particular student. Validation would involve an enquiry, once students were underway, into the proportion of students who were thought to be misplaced. It would then be a matter of comparing the number of misplacements (and their effect on teaching and learning) with the cost of developing and administering a test that would place students more accurately. (p. 30)

Moreover, Johnson and Riazi (2017) cited research stressing the importance of continually studying a tool's validity when it is used to evaluate test-takers. Thus, when a placement instrument is determined and developed, validity must be continually monitored by evaluating accurate student placement. Validity must also be present and evaluated in later stages in the testing instruments we use in MSUM's ELP. We will be able to ensure validity by including on the testing instruments items that connect with the course objectives and course materials, which means that separating the tests by modality streams is necessary.

Reliability. In addition to validity and accurate placement, a placement instrument must have reliability in terms of test-takers' performance as well as in scoring of the tests. Hughes (2003) discussed reliability indicators, including the concept that it is essential to "construct, administer, and score tests" (p. 36) so the same assessment could be given to similar students (regarding language proficiency) at different times yet attain similar results. As with validity, higher reliability can be attempted through various methods (Hughes, 2003), and is a characteristic of a quality placement test. MSUM's ELP must also have reliability with its placement testing instruments; reliability can be achieved through numerous means, such as including more items on the test, avoiding vague items, and writing clear directions, and training scorers, among other test-writing techniques (Hughes, 2003). Whatever test-writing methods for reliability are employed for MSUM's testing instruments, our program should continually evaluate for reliability.

Practicality. Hughes (2003) pointed out that cost and practicality are important considerations, particularly that "a test should be easy and cheap to construct, administer, score and interpret" (p. 56) even though these qualities can be challenging for programs to achieve. Furthermore, practicality includes matters such as the availability of resources needed for

developing and implementing placement testing approach: human resources, facility and equipment resources, time, and money. (Harrington & Carey, 2009; Crusan, 2011; Jamieson et al., 2013; Kokhan, 2013; Parker, 2017) At MSUM, university administration has determined that placement testing will be administered through the language program rather than through the academic unit that provides placement testing for other courses on campus. Therefore, practical considerations must be weighed against the resources available at the program level.

Human resources is an important practical consideration for our program at MSUM. Personnel in our context include one full-time faculty member who coordinates the ELP as well as provides instruction; these duties are half of the employee's assigned load.

Available facilities and equipment are another practical matter that we must consider for our program. MSUM has several large classrooms with an academic set up for projection of test materials and listening to test items ("Scheduling Services," n.d.). One classroom that holds each of the following capacities is available for scheduling: 227, 118, 105, 92, and 81 ("Scheduling Services," n.d.). Many classrooms that hold between 30 and 50 students are also available ("Scheduling Services," n.d.). Several theaters, auditoriums, and ballrooms at MSUM also have academic set up capabilities; these facilities hold 200, 282, 316, and 900 ("Scheduling Services," n.d.). Computer labs available for scheduling can hold 34, 38, 45, and 48 other computer labs available hold fewer numbers ("Scheduling Services," n.d.).

Time is an additional resource consideration for our ELP. Two factors regrading time include the time for administering the instruments and that needed to develop the tests. The first element of time regarding administering the placement tools considers when and how much time is needed. At MSUM, student orientation occurs at the beginning of each semester, which is when the placement testing must take place in order for students to register for appropriate

classes before they begin. Because orientation schedules many activities, students may have a limited time for taking placement tests. A second concern regarding time involves the time needed to determine placement testing instruments, and, if determined, develop the instruments. This matter is related to the human resources available to make this determination.

Although time to develop and administer an assessment merit practical consideration, financial constraints must also be deliberated. The university is currently operating with limited financial resources, and the English Language Program has no formal budget. However, requests can be made for funds from the HLCRWS Department by following a process which the ELP coordinator initiates. The coordinator can submit a request for funding to all faculty in the department after which the faculty votes on the proposal.

Authenticity. Authenticity, another principle of sound assessment, should be considered (Bachman, 1991; Hoekje & Linnell, 1994; Hughes, 2003; Leung & Lewkowicz, 2006). Leung and Lewkowicz (2006) stressed the importance of striving for authenticity and defined it as "the extent to which a test or assessment task relates to the context in which it would normally be performed in real life" (p. 214). Bachman (1991) also noted that authenticity has been described in various ways by language assessors, including "similarity to real life" (p. 690) as well as "direct" (p. 689) and having "face validity . . . or face appeal" (p. 690). Hughes (2003) added that the tasks and texts used for the test strive to be authentic, even though the testing context makes this impossible. Hoekje and Linnell (1994) agreed and added the role of the testing situation must not be discounted, even when the format, such as a lecture is authentic. Instead of these definitions, however, Bachman (1991) claimed that two aspects of authenticity exist, situational and interactional authenticity: "The situational authenticity of a given test task depends on the relationship between its test method characteristics and the features of a specific language use

situation" while "interactional authenticity pertains to the degree to which it invokes the test taker's language ability." (p. 691). Bachman (1991) also argued that due to the relativity of authenticity, we use the terms "'low' or 'high' authenticity, rather than 'authentic' and 'inauthentic'" (p. 692). Leung and Lewkowicz (2006) also agreed that authenticity is based on perspective, for "research suggests that what some perceive as authentic, others may view as inauthentic" (p. 215) while noting the language assessment profession's evolving perceptions of the complex nature of authenticity regarding the assessment of English used worldwide.

Our program at MSUM must also strive for authenticity whenever possible with our placement tests. These findings suggest using direct testing for productive skill tests as well as including test tasks and items as similar as possible to the real life skills we are testing. The courses in the upper-levels, which apply to the CAEP, should have an academic focus while the courses that apply to the CPIE have a more general communicative focus; therefore, the authenticity of the test items should correspond to these types of real-life skills.

Beneficial Backwash. A final characteristic of a placement test for MSUM's ELP could be beneficial backwash, or how the test affects those involved in some manner, including "learners and teachers, on educational systems in general, and on society at large" (Hughes, 2003, p. 53). Beneficial backwash can be attempted in various ways, such as several approaches that seem potentially appropriate to the context a placement test for our ELP, including using direct testing, testing skills our program views as instrumental, providing teacher training, and considering cost (Hughes, 2003). However, Hughes (2003) has also noted that backwash considerations are not a concern with placement tests since test-takers rarely do anything to prepare for them.

Scoring Considerations

A placement test has several potential scoring options. Placement tests can be norm-referenced or criterion-referenced tests, and understanding these two options will allow a program to decide which is most appropriate for their purposes. Brown (1989) clarified the difference between these types of tests, stating that a norm-referenced test (NRT) measures "global language skills or abilities," such as English proficiency in general or understanding of reading, and test-takers' scores are interpreted by comparison with other test-takers' scores (p. 68). Criterion-referenced tests (CRTs), however, attempt to assess "well-defined and specific instructional objectives," usually "unique to a particular program and serve as the basis for the curriculum" (p. 68). Rather than being evaluated on a bell-curve in relation to other test-takers' scores, each student's results of CRTs are interpreted by comparison with stated learning points. Considering our program's purpose for the placement test, an argument for CRT lies in MSUM and MnSCU's curriculum policies and perspective on institutional ownership of courses v. faculty ownership and academic freedom in courses. The need for CRT drives the choice of in-house because MSUM owns the courses and their objectives, so the test must adhere to stated objectives, which are not changeable even when a new instructor takes leadership of a course.

Considering these factors, I suggest the following decisions regarding the selection of placement testing instruments for MSUM's ELP. First, I recommend that our program use a placement test to place students in appropriate courses instead of using the TOEFL, which is a proficiency test used for admissions. Second, a criterion-reference test should be used rather than a norm-referenced test as students' proficiency should be assessed against stated course criteria. I also recommend the program connect the testing instruments to the objectives in the courses in the modality streams to ensure validity. Furthermore, our program should use a test that has

reliability by using the techniques Hughes (2003) noted, such as including more test items, not including vague items, writing clear directions, and training scorers. I have also determined that we must consider practicality issues of human resources, facility and equipment resources, time, and money for our placement instrument. Authenticity must be considered when selecting tests to strive for an instrument that best mimics how students will use the skills in their lives. Finally, our program should be cognizant of backwash regarding our placement tool; using an instrument that benefits all stakeholders as much as possible should be our goal.

Chapter 3: Test Content, Forms, and Sources

In college or university programs that provide English courses at various proficiency levels, learners' English skills are often evaluated to determine what they already know and which skills learners still need to acquire to be successful in their chosen endeavors. All ELPs must decide the ways in which they will evaluate students' language abilities for placement, but one universal placement test does not exist; instead, various placement instruments are used by educational institutions (Jamieson et al., 2013; Hughes, 2003). According to Ling, Wolf, Cho, and Wang (2014), educational institutions in the United States use a variety of means to place students, including standardized admissions test scores, commercially available standardized placement tests, and locally-developed tests, or a combination of these. MSUM's ELP needs a placement process that accurately places students in our courses. Our placement instruments must attend to correct placement for students in our English language courses, validity, reliability, and beneficial backwash, especially regarding practicality and cost. The next task after rejecting the extension of MSUM's proficiency test for admissions to the placement process is to identify the content the placement process must include and what forms it should take so that it meets the principles of validity, reliability, practicality, authenticity, and beneficial backwash. Additionally, possible sources of instruments will be identified.

Test Content

Test content, which considers not one single test, but all possible versions of a test, must be determined for placement tests (Hughes, 2003). Hughes (2003) suggested that test content be

as specific as possible, while understanding that it is impossible to measure all skills in a given area. Content will vary depending on the modality stream, which in our ELP context is the curriculum: pronunciation, listening, reading, writing, syntax, morphology, and culture (see Appendix A). Within each stream, every course has a course description and a specific set of learning objectives, which will guide the decisions regarding content. Hughes (2003) identified the following areas that test-givers must determine: "Operations; types of text; addressees of texts; length of text(s); topics, readability; structural range; vocabulary range; dialect, accent, and style; and speed of processing" (pp. 60-61). Hughes (2003) stated that options exist for different modalities in each of these dimensions. For example, regarding topics, Hughes (2003) suggested that "topics may be specified quite loosely and selected according to suitability for the candidate and type of test" (p. 60). Another dimension, operations, which are the tasks required of the test-taker, could also have alternatives, such as scanning for information and using context clues to predict definitions of unfamiliar words (Hughes, 2003). In a discussion about placement test development for a hypothetical English language training program, Hughes (2003) provided an example of a 45-minute placement test given for the purpose of determining placement in five levels. The test's topics were taken from textbooks the organization uses in all the levels while "other common lexis" were also added for the vocabulary range (Hughes, 2003, p.70). Other test specifications were provided about the remaining dimensions (Hughes, 2003).

These findings help determine decisions regarding these dimensions of content that need to be made for our ELP's placement instruments in order for them to be valid and reliable. Brown and Hudson (1998) suggested that "language testing differs from testing in other content areas because language teachers have more choices to make" (p. 653). Many of these choices deal with the test's content. As with Brown's example, I recommend MSUM's ELP select topics

from textbooks used in the courses at all levels for validity and reliability with vocabulary allowances for other lexis as determined appropriate. Therefore, the topics may include either cultural topics or language as content, depending on which modality is being tested. The other dimensions, such as operations, will need to be determined individually for each modality tested as the placement instruments are selected or developed for each modality stream.

Test Forms

Assessments come in various forms, also sometimes referred to as formats or techniques (Hughes, 2003) or types (Brown & Hudson, 1998), so placement test-givers must decide what form best suits their test type and purpose. Hughes (2003) explained that these formats "are means of eliciting behavior from candidates that will tell us about their language abilities" (p. 75). Selecting the best form for learning about these abilities falls to individual ELPs as they determine placement test instruments. These forms include selected response items, constructed response items, performance items, and self-assessment items although Shin and Lidster (2017) note that placement tools may also combine these formats as needed. One format available for assessments, including placement tests, is selected response (Brown & Hudson, 1998; Hughes, 2003; Gewertz, 2012; Agcam & Babanoglu, 2016). Brown and Hudson (1998) described selected response assessments as those in which test-takers create no language; instead, "language material" is provided, and test-takers are given items that require students to select answers that have a set number of choices (p. 658). Because test-takers usually do not generate written or spoken language, selected response items are most suitable for assessing individuals' "receptive skills like listening and reading" (p. 658). Hughes (2003) added that in addition to reading and listening, test-takers' grammar and vocabulary abilities can also be assessed with selected response items. Brown and Hudson (1998) noted the benefits of selected response items:

swift test administration and quick, objective, and simple scoring. Drawbacks of selected response items involve the lack of productive language required and the challenge of developing assessments (Brown & Hudson, 1998).

Selected response items have several techniques with which they can restrict choices. Because selected response items include a limited number of options from which test-takers can choose, these items can be multiple-choice, true-false, matching (Brown & Hudson, 1998) as well as items that ask test-takers to limit their choices, such as identifying the wrong word out of two words, identifying whether a sentence is grammatical and ungrammatical, and choosing a correct grammatical form (Agcam & Babanoglu, 2016). Of multiple-choice items, Hughes (2003) stated that they can take many forms, but the fundamental structure includes a "stem," such as "Enid has been here _____ half an hour." and several "options," including a correct choice and other "distractors" (p. 75). For this example, choices include "A. during B. for C. while D. since" (p. 75). Another example of a selected response question asked test-takers "to read an article about how scientists track bird migration and to identify two paragraphs that contain the author's opinion on the topic. The question taps key skills required . . . such as comprehending 'content rich' nonfiction and citing textual evidence for an argument" (Gewertz, 2012, "Consortia Anticipate Lengthy, Iterative Process," para. 7). Another example is the following statement, which test-takers should identify as true or false: "NY is more crowded than California" (Agcam & Babanoglu, 2016, p. 70). These examples illustrate the objective nature of selected-response items with the limited number of choices for test-takers.

These findings provide a basis for making decisions about using selected-response items for our placement instrument. As indicated by the literature, selected-response items would be appropriate for items on placement instruments for the reading, listening, syntax, and vocabulary

streams because they provide objective measurements and are easy and fast to administer and score. Although the time to develop the tests and the problem of guessing would be a drawback, the benefits outweigh the disadvantages. For our program, I have determined that selected-response items would be beneficial for these modality streams.

In addition to selected-response items, constructed-response items may be used on assessments (Brown & Hudson, 1998; Hughes, 2003; Gewertz, 2012; Agcam & Babanoglu, 2016). Constructed-response items ask students to produce language, such as writing and speech, so they are fitting for assessing these productive skills (Brown & Hudson, 1998). Brown and Hudson (1998) added that constructed-response assessments can also be used to evaluate "interactions of receptive and productive skills," such as the those between reading and writing as well as speaking and listening (p. 660). Agcam and Babanoglu (2016) noted that constructed response items are more demanding for test-takers than selected-response assessments. Advantages of constructed-response items include less of what Brown and Hudson (1998) call the "guessing factor," (p. 660) compared with selected-response items and the ability to measure productive and interactive skills. Disadvantages noted involve scoring challenges, such as the subjectivity, difficulty, and the length of time needed (Brown & Hudson, 1998).

As with selected-response assessments, common forms used for constructed-response items exist. Examples of frequently-used forms in language testing include techniques such as fill-in the blank or gap, short-answer (Brown & Hudson, 1998; Hughes, 2003), labeling, correcting mistakes, unscrambling words, and transforming sentences (Agcam & Babanoglu, 2016). Hughes (2003) added that short answer items are often used when assessing listening and reading and provided the following examples: "What does *it* in the last sentence refer to?" and "How old was Hannibal when he started eating human beings?" (p. 79). Another illustration of a

constructed-response question provided by Agcam and Babanoglu (2016) asks test-takers to fill in the blank to complete the sentence: "What _____ her name?" (p. 70) In addition to the common short answer and fill-in-the-blank forms, a constructed-response item can also ask test-takers to "Find the mistake and correct it. 'I likes travelling.'" (Agcam & Babanoglu, 2016, p. 70). These and other similar constructed-response techniques can assist in determining language abilities for placement.

These findings regarding constructed-response items allow for decisions regarding our placement instruments. The literature revealed that constructed-response items would be appropriate for placement instruments for the productive skills of writing and speaking as well as for interactions of the skills. It might be more practical to combine the assessments of skills, such as those noted in the literature, such as reading and writing or speaking and listening. The advantages of using constructed-response items include more accurate responses, which enhances validity, while the disadvantages include the time for rating and the subjectivity of scoring, which complicates reliability. Considering these factors, for our program, I have determined that using constructed-response items would be beneficial for items on placement instruments in the following modality streams: pronunciation, reading, listening, vocabulary, and syntax.

A third available placement assessment format includes performance items (Brown & Hudson, 1998; Gewertz, 2012). Performance items on tests often include three main characteristics: items should be task-based, authentic, and scored by raters who have been trained (Brown & Hudson, 1998). Performance test items can evaluate the productive skills of writing and speaking or assess a combination of skills, including reading (Brown & Hudson, 1998). Performance tasks ask the test-taker to interact with "more complex, prolonged exercises"

(Gewertz, 2012, "Consortia Anticipate Lengthy, Iterative Process," para. 9). Brown and Hudson (1998) described the advantages of performance assessments, which include their authenticity of tasks close to real-life communication, causing validity to increase, and beneficial washback of relating to a specific curriculum. Performance items have numerous disadvantages, such as those cited by Brown and Hudson (1998), including significant cost required to develop and administer. Other drawbacks relate to scoring: rater training, holding sessions for rating, and reporting score data (Brown & Hudson, 1998). Brown and Hudson (1998) also noted reliability concerns with subjectivity and unpredictable ratings, numerous validity issues related to covering content adequately, and test security due to the low number of prompts.

Performance items on English language placement tests can take various forms (Brown & Hudson, 1998; Hughes, 2003; Gewertz, 2012). Common forms include "essay writing and interviews or more recent developments like problem-solving tasks, communicative pair-work tasks, role playing and group discussions" (Brown & Hudson, 1998, p. 662). Gewertz (2012) highlighted the complexity of tasks in performance items and described a challenging performance item that asks test-takers to read an interview on a topic before directing them to related web sources. Regarding these sources, test-takers must answer constructed-response questions, investigate vocabulary, and discuss web source reliability before researching and presenting a speech using audiovisual aids (Gewertz, 2012). Another example of a performance item includes essay writing. Hughes (2003) provided several examples of writing tasks that measure only writing ability: "Write the conversation you have with a friend about the holiday you plan to have together" and "Discuss the advantages and disadvantages of being born into a wealthy family" (p. 90). As research indicates, many options for using performance items are available for determining student placement.

These findings regarding performance items also inform decisions about our ELP's placement instruments. As with constructed-response items, performance items would be suitable for use on placement instruments for the writing and speaking as well as for evaluating a combination of the skills. Similar to using constructed-response items to evaluate several skills simultaneously, combining skills on testing instruments is a possibility for our program. The advantages performance items include are a more accurate measurement of skills due to authenticity, which increases validity, and, most importantly, a placement instrument that can be directly tied to our program's curriculum. Although the disadvantages of test security and scorer subjectivity affect reliability, and time-intensiveness to develop, administer, and score is present, the two main benefits will aid in a more accurate placement, and attempts to decrease the effects of the disadvantages can be made. For our program, I have decided that using performance items on placement instruments in the writing and pronunciation modality streams would be the most appropriate.

Self-assessment is another test form, and when discussing placement tests, it would be remiss not to mention this trend in L2 testing. Although not actually a test in the traditional interpretation, this placement model or method, often termed either *directed self-placement* (DSP) or *self-assessment*, can be used instead for placement purposes. Self-assessment is a type of personal response assessment in which students evaluate their own language abilities (Brown & Hudson, 1998; Nunan, 2015; Agcam & Babanoglu, 2016). Similarly, Crusan (2011) defined self-assessment as students reporting on their writing skills and taking responsibility for their learning and performance (Hughes, 2003; Crusan, 2011). Ferris, Evans, and Kurzer (2017) noted that DSP has been popular in non-L2 university writing contexts for nearly two decades, but L2 writing programs have not embraced it as much. However, a number of intriguing studies

(Brown & Hudson, 1998; Crusan, 2011; Kolkan, 2012; Nunan, 2015; Ferris, Evans, & Kurzer, 2017; Summers, Cox, McMurry, & Dewey, 2018) have researched the use and effectiveness of DSP in second English language program contexts with mixed comments on its effectiveness for accurate placement.

Self-assessment has gained popularity, perhaps due to the practicality of low cost and lack of concern about test security (Crusan, 2011). Regardless of the interest in the practicality of DSP, it has been a subject of recent discussion regarding its accuracy, some studies supporting it, and others not (Crusan, 2011; Summers et al., 2019; Ferris et al., 2017). Ferris et al. (2017) suggested that L2 writing programs have not used DSP as much because of inaccurate self-placement due to students' lack of accurate knowledge about what writing proficiency means in an academic institution in the United States as well as either overconfidence or a lack of confidence in their language skills (Ferris et al., 2017). Summers, Cox, McMurry, and Dewey (2019) discussed a framework for self-assessment with the connection of NCSSFL-ACTFL Can-Do Statements in which learners self-assess how well they can perform the stated tasks. This framework could be helpful, and although some support for self-assessment exists, especially with regard to informing students about their learning, ultimately, Summers et al. (2019) concluded that "correlations between self-assessments and placement tests were so low that it was determined that the self-assessment was not valid as the sole measure of students' proficiency levels for placement purposes" (p. 284), especially with students in lower proficiency levels. Wall, Clapham, and Alderson (as cited in Kokhan, 2012) indicated that because students come from various educational and cultural backgrounds, asking them to evaluate their English proficiency with precision would not yield accurate measurements. Ferris et al. (2017) concurred that an in-house DSP cannot by itself accurately place multi-lingual students in courses in multi-

level composition programs in the United States but it might assist in the process. Summers et al. (2019) noted, however, that self-assessment might be used in addition to other assessment measures but is in need of refinement. DSP, especially with writing placement, has benefits, such as empowering learners with agency and letting them understand that they are valued (Crusan, 2011). Furthermore, self-assessment is tempting for universities and programs because it might help solve the problem of unpredictable needs for classes and teachers due to the typical last-minute placement testing. However, mixed research has indicated that if considered, it should be used as a complementary assessment tool rather than the sole placement instrument (Summers et al., 2019). Ferris et al. (2017) indicated that misplacement due to self-assessment can cause problems for students, who are not receiving instruction at the needed level; for instructors, who have students whose abilities vary too much; and for programs, which can have final assessment issues and low passing rates. Self-assessment might have a role in student placement to give students a voice in their educational experience and assist in their motivation, but it should not be used as the primary placement tool (Summers et al., 2019; Ferris et al., 2017).

The research provides sufficient information regarding DSP for me to make decisions regarding its use in our placement instruments. The findings suggest that too many errors can occur when students attempt to self-place. Certainly there are advantages and positive associations with DSP, including student motivation and evaluation of their skills; however, that does not necessarily equate to accurate placement. Considering the mixed results of the literature and due to the lack of research supporting DSP as a primary placement test, it will not be considered as a central placement instrument for our ELP at this time.

Test Sources

While considering these characteristics, the next stage in determining our placement process is to examine the sources of placement instruments. Often, two main types of placement testing instruments are considered for institutions: standardized and in-house assessments (Jamieson et al., 2013; Hughes, 2003). Standardized tests are written by companies or universities and sold to customers, namely various types of English language programs, while in-house instruments, also referred to as locally-developed tests, are designed by individual programs. Researchers note that both are used for placement purposes in ELPs like MSUM's (Brown, 1989; Hughes, 2003; Jamieson et al., 2013). Therefore, a closer examination of each source in general and specific options in each is warranted. Whichever option is selected for the ELP at MSUM, it is essential that the placement instrument consider accurate placement for our program's curriculum and course objectives, reliability, validity, and beneficial backwash, especially practicality and the financial situation of our program at MSUM.

Chapter 4: Standardized Test Options

Various commercial standardized placement tests are available for universities in the United States. Standardized tests can be appealing because of certain practical aspects; they alleviate stresses upon time constraints faced by universities regarding test development and scoring, especially when online tests can be taken in various locations around the world by many test takers (Jamieson et al., 2013). Language programs also have confidence in the reliability of commercial placement tests due to the extensive research and development practices of commercial test developers (College Board, 2019; Michigan Language Assessment, "Quality in Test Design," n.d.; iTEP academic, n.d.). Although commercial standardized placement tests are available and appealing, Hughes (2003) has cautioned that "placement tests can be bought, but this is to be recommended only when the institution concerned is sure that the test being considered suits its particular teaching programme" (p. 16). Careful review of tests then is in order before one of these options is selected.

Three common standardized tests that seem particularly fitting and worthy of review for our purposes were College Board's ACCUPLACER ESL, the University of Michigan's English Placement Test (EPT), and the International Test of English Proficiency (iTEP Academic). Validity needs to be addressed regarding these tests, specifically content validity needs to be determined through comparison to program content: modality streams and levels.

ACCUPLACER ESL

ACCUPLACER, a program that offers a suite of tests, including ACCUPLACER ESL, is owned and marketed by the College Board (College Board, 2019; James and Templeman, 2009). In an overview of ACCUPLACER ESL, College Board (College Board, 2019) stated that "ACCUPLACER ESL is a set of computer-adaptive, untimed tests that enables you to test quickly and efficiently across a wide range of abilities" (College Board, n.d., p. 2) that can be used to "place them [students] in English immersion programs or developmental or college-level courses" (College Board, n.d. , p. 2). This set of ESL tests includes four multiple-choice tests and one essay assessment (College Board, 2019). The ACCUPLACER ESL has been adopted at some institutions because of its practical advantages (James & Templeman, 2009). James and Templeman (2009) also noted that the ACCUPLACER ESL is easy and inexpensive to administer and tests relevant language skills with test results instantly accessible; however, they stressed that even though a commercial test can be practical for several reasons, faculty involvement is essential for accurate placement.

When reviewing placement instruments, ACCUPLACER ESL was considered because MSUM already employs ACCUPLACER for assessing NSEs and ACCUPLACER ESL for assessing NNSEs in order to place students in first-year English courses within the English Department. ACCUPLACER and ACCUPLACER ESL are the placement instruments recognized by the Minnesota State system of colleges and universities for reading, writing, and math placement (Minnesota State, 2006). Therefore, it seemed fitting to consider this test first (see Table 1).

Currently, three of ACCUPLACER ESL's five tests are used at MSUM to place students into the ESL sections of English 101, offered by the English Department, and TEFL 104:

Academic Writing, offered in the ELP. To assess this test for content validity for all of our ELP courses, I took the three tests MSUM already uses for placement: Reading Skills, Language Usage, and Sentence Meaning.

Table 1

ACCUPLACER ESL

MSUM ELP Modality Streams	**Skills Tested on ACCUPLACER ESL**	**Format**	**Content Validity for MSUM ELP**	**Used by MSUM**
Listening	Yes	Selected Response	Yes*	No
Reading	Yes	Selected Response	Yes*	Yes
Syntax	Yes	Selected Response	Yes*	Yes
Vocabulary	Yes	Selected Response	Yes*	Yes
Writing	Yes	Performance	Yes*	No
Speaking	No	NA	NA	NA

Note: *Content coverage for all levels is uncertain.

The Reading Skills test contained questions that assess literal comprehension and inferences, two of the key student learning outcomes (SLO) in our courses, such as in TEFL 103: Academic Reading I. Overall, the Reading Skills test seems appropriate for many learning outcomes in our program's reading modality stream. The Language Use test contains useful content for our syntax stream. The test contains cloze questions and sentence combining for six syntactical challenges: "Nouns, pronouns, pronoun case structure; Sentence structure; Subject-verb/agreement; Adjectives/adverbs; Verbs; [and] Subordination/coordination" (College Board, n.d., p. 5). Many of the SLOs in our syntax stream correspond with these tested features, specifically the usage of basic and complex syntactic structures in writing. Finally, the Sentence

Meaning test assesses how well the test-taker understands sentences in English with four areas addressed: "participles, phrasal verbs, and prepositions; Adverbs, adjectives, and connectives sequences; Basic nouns and verbs; [and] Basic and important idioms" (College Board, n.d., p. 4). This test could be beneficial for our vocabulary courses because the test uses one and two sentence frameworks to determine if the test-taker understands the word meanings.

The set of ACCUPLACER ESL tests also includes two tests that MSUM does not currently provide, a listening test that measures proficiency in listening for literal comprehension and implied meaning and WritePlacer ESL, an essay test that determines readiness for college writing courses. Although I was unable to take these two tests since they are not provided by our university, both of them might be useful for our listening and writing modality streams respectively, as they address several key outcomes in our program's courses. However, James and Templeman (2009) claimed that while the ACCUPLACER *WritePlacer ESL* test is efficient, it was also the least accurate placement tool in writing courses in their study and resulted in numerous misplaced students, so this raises concerns on the validity of the instrument for the writing modality stream.

Thus, from the standpoint of content validity, the ACCUPLACER ESL seems to hold promise, as it covers many of the learning outcomes in our program's modality streams. However, there is no speaking test, so one modality stream is absent, and there is cause for concern on the validity of the test for the writing modality stream. It is also difficult to determine whether the test assesses language skills across all five levels of the ELP. Nevertheless, these validity concerns may be settled later with assistance from the test developer. In *Accuplacer Placement Validity Study* (2018), College Board's Admitted Class Evaluation Service has provided information for institutions if they want to determine whether the ACCUPLACER ESL

would be a valid placement test for their individual program. That is, College Board will work with institutions to conduct validity studies for their particular program regarding the use of ACCUPLACER ESL for their program.

ACCUPLACER ESL test format corresponds to accepted test formats as stated in Chapter 3. All tests aside from the WritePlacer ESL use selected-response items in the form of multiple-choice; the format of the WritePlacer ESL is a performance item, an essay prompt (College Board, 2019).

On a practical level, the fact that some tests in the set of ACCUPLACER ESL tests are readily available at Minnesota State University Moorhead and that ACCUPLACER ESL is the state-sanctioned placement tool for use with NNSE for freshman English courses make the test an attractive choice for the ELP. MSUM already has the equipment necessary to administer this test. Other practical concerns to review as we consider extending its use to the ELP include costs and human resources. MSUM students may take the set of tests up to three times per year. The first administration is free for each student, but a charge of $10.00 for each test is assigned for each additional administration thereafter (“Accuplacer testing,” 2019). An analysis of the cost to the university provides additional considerations. ACCUPLACER ESL tests cost one unit per test for the Reading Skills, Sentence Meaning, and Language Use exams, two units for the WritePlacer ESL, and 2.5 units for the Listening test (College Board, n.d., p. 35). Universities are charged by the College Board for the units (College Board, n.d., p.35), and the administrative department on campus that conducts the ACCUPLACER for English Department and Math Department courses has stated that they have insufficient funding and personnel to provide placement testing for our ELP. They also have indicated they have insufficient funding to provide training to other university employees to administer the assessments for this purpose.

Therefore, although the ACCUPLACER ESL could be beneficial for some of our courses, the lack of funding, training, and personnel make this test impractical.

Furthermore, practical concerns are tied to future use of the instrument with uncertainty surrounding decisions currently being made at the state level. Changes for placement may be coming in the Minnesota State system of colleges and universities. The system will be using the Next-Generation ACCUPLACER beginning in January of 2020; MSUM will begin using it on that date for native English speakers' placement in writing courses. Currently, state committees are discussing various placement testing options for first-year English courses, and the future of sanctioned use of the ACCUPLACER ESL as a placement tool is unknown. If the state system chooses another placement tool, the ACCUPLACER ESL would no longer be provided to Minnesota State University.

In light of these uncertainties and budget constraints, ACCUPLACER ESL would not seem an appropriate placement test to invest our time to pilot. Fortunately, we are free to consider other options while the system office undertakes its own review of placement testing processes. According to Minnesota State's System Procedure 3.3.1 Assessment for Course Placement (2006), "a college or university may require additional assessment measures (e.g., computer literacy, study skills inventories, or occupational-related tests) for purposes other than the objectives (reading, writing, and math) of the system-endorsed placement instrument for advising and placement purposes" (Part 3: Placement Instrument, Subpart D, para. 4). Therefore, the policy allows for other placement instruments to be utilized if the program deems it necessary. An English language program is an appropriate purpose because its aims and course content often might have varied assessment needs. We took advantage of the freedom to look at other commercially available standardized tests.

Michigan English Placement Test (EPT)

A second commercial placement test to consider is the Michigan English Placement Test (EPT), the well-known Michigan Language Assessment placement test from the University of Michigan and Cambridge Assessment English (see Table 2). The EPT measures "listening comprehension, reading comprehension, grammatical knowledge, and vocabulary range" (Michigan Language Assessment, "Quality Design," n.d.). Example test questions in the sample booklet provided by the Michigan Language Assessment on their website (Michigan Language Assessment, "Michigan EPT," n.d.) indicate that this test would match course descriptions and measure many outcomes for the following modality streams in our program: listening, reading, syntax, and vocabulary. The overall format and content follow a selected-response format in the form of multiple-choice items. This exam could be beneficial for some of the language streams in our program as the EPT appears to meet content validity for the listening, reading, syntax, and vocabulary modality streams. However, the EPT does not provide speaking and writing tests, so two modality streams are not present. Moreover, whether the EPT measures language skills across all five levels of the ELP is difficult to verify.

This 60-minute timed placement test of 80 questions is expensive at nearly $6.00 per test with a minimum order of 25 tests (B. Wheeler-Floyd, personal communication, February 18, 2019). The test would also require MSUM to purchase headphones for each test-taker and provide computer access (B. Wheeler-Floyd, personal communication, February 18, 2019) as described for the ACCUPLACER ESL. Although the EPT provides reliability and possible validity for placement for some of our courses, financial considerations remain a major obstacle to adoption and results in a rejection of the EPT for our placement purposes.

Table 2

Michigan English Placemen Test (EPT)

MSUM ELP Modality Streams	**Skills Tested on EPT**	**Format**	**Content Validity for MSUM ELP**
Listening	Yes	Selected Response	Yes*
Reading	Yes	Selected Response	Yes*
Syntax	Yes	Selected Response	Yes*
Vocabulary	Yes	Selected Response	Yes*
Writing	No	NA	NA
Speaking	No	NA	NA

Note: *Content coverage for all levels is uncertain.

International Test of English Proficiency (iTEP)

Finally, the International Test of English Proficiency (iTEP), as presented in Table 3, has designed iTEP-Academic, a test that can be used for both admissions decisions for degree programs and placement within English language programs. This dual-purpose test is gaining popularity in universities (Lesho & Santoro, 2019) and could be a consideration for our program for placement purposes. The iTEP-Academic was designed as a low-stakes assessment for lower English language proficiency level students but with higher stakes for students at a higher proficiency (Lesho & Santoro, 2019). Regarding the stakes, Lesho and Santoro (2019) explained that iTEP-Academic's design, achieved through computer-adaptive means, is more difficult for high-proficiency test-takers to challenge them to perform as well on it as lower-proficiency students tend to do; this test design addresses common anecdotal concerns in IEPs about their students' lack of motivation for performing well on tests used for exit of IEPs when these students know their course performance already ensures exiting a program (Lesho, 2017,

January). Designed to be lower stakes for lower proficiency students in IEPs, (Lesho & Santoro, 2019), the iTEP-Academic "has lower level test items and tasks that give lower proficiency students an opportunity to score on the test" in each skill section (Lesho, 2017, March, para. 3).

Table 3

iTEP-Academic

MSUM ELP Modality Streams	**Skills Tested on iTEP-Academic**	**Format**	**Content Validity for our ELP**
Listening	Yes iTEP-Core	Selected Response	Yes*
Reading	Yes iTEP-Core	Selected Response	Yes*
Syntax	Yes iTEP-Core	Selected Response	Yes*
Vocabulary	No	NA	NA
Writing	Yes iTEP-Plus	Performance	Yes*
Speaking	Yes iTEP-Plus	Performance	Yes*

Note: *Content coverage for all levels is promising, yet uncertain.

Two versions of iTEP-Academic are available, iTEP-Core and iTEP-Plus. iTEP-Core measures grammar, listening, and reading, while iTEP-Plus assesses those three areas as well as speaking and writing from the beginning to mastery levels (iTEP Academic, n.d.). iTEP-Core consists of multiple choice items in a 50-minute time frame, while the iTEP-Plus requires a total of 80 minutes to allow time for the writing and speaking components. The 25-minute writing test is a two-part exam requiring test-takers to write on two prompts (iTEP Academic, n.d.). The five-minute speaking test also contains two parts in which the test-taker must speak. Descriptions of each skill area test and sample test items are available on the iTEP website (iTEP Academic, n.d.). Therefore, for consideration for MSUM's ELP, content validity appears favorable for the following modality streams: pronunciation, listening, reading, writing, and syntax. However,

iTEP-Academic does not offer a vocabulary test. iTEP addresses details concerning the iTEP-Academic tests' reliability and validity in their *Validity and Reliability Report* (Seiler, 2016). This report indicates that use of iTEP-Academic would be reliable and valid for five of our six modality streams that require a placement instrument.

The practicality of using the iTEP-Academic must also be evaluated. The iTEP-Plus requires a computer for the speaking and writing tests and headphones for the speaking component. The iTEP-Core is also available in a paper-based format (iTEP Academic, n.d.), which could alleviate issues of scheduling a computer lab. Regarding practical scoring matters, iTEP scores have been aligned for context with TOEFL, IELTS, and CEFR scores, and they are provided to programs quickly, typically within a 24-hour time frame (Lesho & Santoro, 2019). Trained graders rate the writing and speaking tests and are vigilant when detecting and flagging cut and paste cheating (Lesho & Santoro, 2019). Other security measures include online proctoring in real time as well as taking numerous pictures of test-takers while they are taking the assessment (Lesho & Santoro, 2019). Lesho and Santoro (2019) also noted that the printed test results benefit instructors since they inform them of what they need to teach while the results benefit students with agency in knowing their strengths and challenges. Financial considerations are a concern regarding the cost to purchase the test. iTEP pricing is not provided publicly, but it is available upon a phone consultation with a company representative.

These findings indicate that the iTEP delivers reliability and possible validity for placement in all the modality streams needed aside from the vocabulary stream. However, because iTEP does not contain a vocabulary test, its incompleteness causes me to reject it for consideration for our program.

Considerations

Based on a close examination of the sample test questions for the three placement tests provided on the companies' websites, it can be concluded that the three provide similar commercial standardized tests for placing students in various English language training contexts. All three companies have addressed their research regarding their tests' validity and reliability in detailed reports. On the other hand, Hughes (2003) noted that "No one placement test will work for every institution, and the initial assumption about any test that is commercially available is that it will not work well" (p. 16) with the exception of tests created for language schools that use similar textbooks, and, therefore, have similar programs. My examination has underscored Hughes' admonition for caution and has unearthed a key impractical concern; the cost for commercially produced tests is high. The alternative to high-cost standardized tests is to devise one's own in-house placement procedures.

Chapter 5: In-House Tests

In-house placement tests, sometimes referred to as locally-developed tests, are designed by individual language programs. As with commercial placement tests, their purpose is to place students into different parts in a program that best match their language abilities, and, in most cases, this means specific courses at a particular level (Hughes, 2003; Brown 1989). For instance, in MSUM's program (see Appendix A) a placement testing instrument is needed for the five levels in the following language skill areas: listening, pronunciation or speaking, reading, writing, vocabulary, and syntax. The courses in the culture stream focus more on content rather than language skills; therefore, they will not need a placement tool. Regarding placement instruments, Hughes (2003) recommended in-house tests:

> The placement tests that are most successful are those constructed for particular situations. They depend on the identification of the key features at different levels of teaching in the institution. They are tailor-made rather than bought off the peg. This usually means that they have been produced "in-house." The work that goes into their construction is rewarded by the saving in time and effort through accurate placement. (p.17)

Brown (1989) also supported creating placement tests for particular programs, emphasizing the importance of ensuring that the test addresses the skills taught in a program. Parker (2017) concurred regarding content and added the following:

> A placement test is most effective when it measures the same range of language proficiency as what is taught in the program. Test items that fall outside the proficiency

> range of the program do not accurately reflect content that would be found in the program curriculum. (p.1)

This concept provides comfort to programs like ours that need to set aside commercially available standardized tests for practical reasons. In-house tests, it turns out, are not only viable, but even recommended. Their validity levels can be high.

Validity and Reliability of In-House Tests

As with commercial placement tests, in-house versions must also address both validity and reliability. When discussing the stages of test development, Hughes (2003) noted that validity concerns may differ depending on the context, claiming that validity is critical for a "high stakes or published test" (p. 66). However, he pointed out that when considering "relatively low-stakes tests that are to be used within an institution, this may be not thought necessary, although where the test is likely to be used many times over a period of time, informal, small-scale validation is still desirable" (p. 66). When considering modality streams such as reading, listening, syntax, and vocabulary, using portions of course materials could address validity concerns by ensuring that the test measures what students are learning, provided the materials adequately address course objectives. Regarding validity, Hughes (2003) recommended to first determine the entire content and create a representative sample of test items and "whenever feasible, use direct testing" (p. 33-34). Christopher (1993) also recommended direct testing, especially for evaluating writing: "A test measuring writing ability, therefore, should be used to place students in programs where their future grades will be based upon their writing proficiency, as indirect test scores do not measure writing as a skill" (p. 16), and cited research indicating that students' indirect test scores should be disregarded when discrepancies between placement and other tests exist. Additionally, Johnson and Riazi (2017)

stated that many English language programs in the United States do not check their in-house placement tools for validity, and they strongly suggested programs make continual efforts in these regards, not only regarding the test itself, but also with creating rubrics and scoring tests, especially with writing assessments.

Hughes added that scoring should connect to the tested material and reliability be attended to as diligently as possible (2003, p. 34). Hughes (2003) also provided suggestions for increasing test reliability regarding scoring and the test itself. Several of Hughes' (2003) recommendations include not permitting too much freedom on items, such as writing tests, and making sure the test questions and instructions are clear and the format is well-organized. These researchers provide suggestions that MSUM's program can use, such as aligning test questions with curricular objectives and using direct testing when possible, as we develop placement instruments that achieve validity and reliability.

Practicality and Authenticity of In-House Tests

When considering and creating in-house placement tools, practicality must be considered. These often include time and financial constraints for "development, administration, scoring, and reporting" (Jamieson et al., 2013, p. 295). Individual programs must develop their own materials, and how a program determines which materials to use is at the discretion of the program. One way programs can save time and effort in developing an in-house placement test is to adopt or adapt course materials. The copyright principle of fair use allows for replication of passages for educational purposes with four main principles to be considered. These principles include the transformative nature of the work, the nature of the work, how much of the work is used, and how using the work would cause the copyright owner to lose a profit (Nolo, n.d.). Applying these four principles, the educational use that language programs often employ with in-house testing

would be considered fair use. For instance, when constructing a placement test for the reading stream, a program could use selected excerpts from reading textbooks used in the courses at the different levels. For these excerpts, the local test developers could create their own assessment items that correspond to the objectives of the reading courses at all levels. Materials development could be similar for testing in the other modalities in the program. For the ELP at MSUM, these findings regarding practicality and fair use can aid in alleviating some of the burden of time and other resources our program will face. Using course materials also ensures authenticity because these are the materials students will be working with in the ELP classes.

In-House Placement Testing Options for the Modality Streams

Listening. The listening modality stream in our ELP test needs a placement instrument to measure auditory skills. Kress (2008) defined auditory skills as they relate to acquiring a language as:

> the trained or culturally acquired ability to recognize the sounds that belong to a particular language, and the ability to discriminate among similar sounds within the language. Aural skills are related to the ability of the student to listen and understand spoken messages. (p. 327)

Options vary for how to assess students' listening skills, including instructor interviews with students in which students are instructed to follow directions. Many possibilities exist for the directions. For example, Kress (2008) described a method in which the interviewer reads several short written passages of 150 words or less as the student listens. Then, the interviewee reads three paraphrases of one-sentence each after which the student must select the most accurate paraphrase. Kress (2008) presented other methods for assessing various essential listening skills, such as factual and logical understanding, logical predictions, and comprehension

of main ideas. Based on the research from Brown and Hudson (1998) and Hughes (2003) regarding test forms, selected-response items have been identified as the most appropriate for listening skills because of the benefits of efficient test administration and scoring.

One way in which in-house placement tests for reading and listening can be constructed was explained by Jamieson et al. (2013). These educators have used "passages and questions from achievement tests at four different proficiency levels in the curriculum" at their four-year university (Jamieson et al., 2013, p. 289), and student performance on these two tests is measured against "a scale that consists of ordered ability levels," while "cut scores allow placement into English classes with students at relatively homogenous proficiency levels" (p. 298). In these findings, the researchers provide ideas that will cover all the listening skills in all the levels of MSUM's program (Table 4).

Table 4

Listening Stream

Suggested Forms	**Item types**	**Listening Text Sources**
Selected response items (Brown & Hudson, 1998; Hughes, 2003)	Multiple-Choice	Interview (Kress, 2008)
	True-False	Passages from achievement tests (Jamieson, et al., 2013)
	Matching	Passages from listening stream textbooks
		Passages from openly available sources

Speaking. Regarding the methods of assessment of speaking skills for placement purposes, Jamieson et al. (2013) cited a review of nearly 100 English for Academic Purpose

Programs (EAPs) in the United States published on the website *Commission for English Language Accreditation Programs.* This review determined that almost half of the EAPs interview students in-house, while 15% use a commercially-available standardized test (Jamieson, et al. 2013). Interestingly, 35% use no speaking test when determining student placement; the lack of a speaking assessment might be due to impracticality, suggested Jamieson et al. (2013).

Performance items have been found to provide the most accurate measurement of speaking skills due to their potential for higher levels of authenticity although care should be taken to ensure the items are task-based and scored by trained raters (Brown & Hudson, 1998). When testing speaking, Kress (2008, p. 329) noted that oral language "includes both one-way and interactive speech," so an oral assessment should contain both elements, if possible. In one-way speaking, a speaker presents information with little to no interaction from the listeners while a mutual exchange occurs in interactive speech (Kress, 2008). Several options exist for developing an in-house tool to measure both of these characteristics, as shown in Table 5. Often with in-house testing, speaking assessments are conducted along with the assessment of listening skills. As with the listening skills, an interview is a commonly-used placement instrument in EAPs (Kress, 2008; Jamieson et al., 2013). Kress (2008) outlined a number of options for the interview method, including structured and open-ended options. Topics for interviews are varied, such as asking students to give directions about how to perform a task, impromptu or cued storytelling, oral reports, role-playing, and extemporaneous narratives (Kress, 2008).

Jamieson et al. (2013) further discussed in-house speaking tests: "In an EAP program, the speaking portion of the placement test is an opportunity for students to demonstrate their ability to speak at length in a way that would simulate an academic speaking experience at a university"

(p. 290), contrasting it with a computerized speaking tool that could be advantageous regarding the aspect of time for administering and scoring. As an alternative to in-house face-to-face tests, programs could develop and administer a test with recorded speech using a platform such as Moodle to assess pronunciation and other speaking skills as is done in some U.S. universities (Chung, Haider, & Boyd, 2015).

Table 5

Speaking Stream

Suggested Forms	**Assessment Targets**	**Example Item Types**
Performance items (Brown & Hudson, 1998)	One way speaking (presentation)	Recorded Speech on system platform (Chung, Haider, & Boyd, 2015)
	Interactive speaking (mutual exchange) Kress (2008)	Interview (Kress, 2008; Jamieson, et al., 2013)

To evaluate a learner's speaking abilities, a set of criteria would also need to be determined. Potential rubrics that programs can adopt or adapt include the independent and integrated speaking rubrics developed by and used for the TOEFL iBT Test or other commercially produced tests. Available on the testing companies' websites, these rubrics evaluate the speaker's speech regarding main categories, such as general description, delivery, language use, and topic development (ETS, n.d.) However, an in-house rubric could be developed instead to more accurately measure a student's appropriate placement within a program's speaking courses.

Although some EAPs do not use a speaking test, I have determined that our program should provide an in-house test that contains performance items for placement in the pronunciation stream because, unlike in many programs, MSUM's curriculum separates the

listening and speaking modalities at some levels. Thus, an in-house test to assess pronunciation and other speaking skills for placement should be developed for use in our program.

Reading. As with the other modalities, a program could develop an instrument to place students in the reading stream. The test should include items relating to the learning objectives in all of the program's reading courses to address validity. When developing an in-house reading test for their English language program at the University of Hawaii, Brown (1989) indicated that "it seems logical" to include on the assessment skills and knowledge taught in their reading course (p. 80). When constructing the program's reading test, Brown (1989), explained that test items are written for each of the reading course's objectives. Hughes (2003) cautioned that while reading tests are easy to create, the tests may not measure what the test-developer intended because receptive skills cannot be directly observed as can the productive skills, so it can be difficult to know whether a test accurately measures a test-taker's reading ability.

Regardless of the challenges, Hughes (2003) provided suggestions for developers of a reading test, including considering both expeditious and careful reading operations, text selection, reading speed, and criterial levels. Other considerations for an in-house reading test include the topic choice, the text framework, passage length, and quantity of items assessed (Hughes, 2003; Kim, 2015). Once these types of test construction issues are determined, test-developers can begin to consider what reading skills to test. Keeping in mind the learning objectives of the courses, a local-developer of a reading assessment can test for elements such as comprehension of main ideas, factual and inferential understanding, and logical predictions (Kress, 2008). Reading placement tests often contain reading passages followed by multiple-choice items testing semantic and pragmatic meaning (Hughes, 2003; Kim, 2015). In addition to multiple choice, other potential techniques for a reading placement test include short answer, gap

filling or cloze; however, Hughes (2003) suggested that professional test developers have used mostly multiple choice because "few basic techniques are needed, and non-professional testers will benefit from concentrating on developing their skills within a limited range" because while other techniques, such as cloze, "involve reading to a high degree, it is not clear that reading ability is all that they measure" (p. 153). When developing a test, many decisions need to be made, such as the topics of reading passages. Hughes (2003) provided many suggestions, such as choosing interesting topics that will not likely be known to test-takers while Kim (2015) provided examples of topics on a reading placement test used at a university ELP, including a "complaint letter, Argentine ants, mysteries of smell, demise of Neanderthals" (p. 236). In addition to topic selection, when choosing readings skills to be tested, Kress (2008) underscored that when testing the reading proficiency of ELLs, it is important to "consider methods that highlight vocabulary, idiomatic expression, register, and connotation" and that ELLs might have more difficulty with these areas than native English speakers (p. 332). Kress (2008) also provided suggestions for methods to assessing reading comprehension. Kress's techniques for assessing "gist recognition" include working with passages of 100-200 words and asking the students to select the best title, main idea, or summary statement from four choices" (p. 332). Hughes (2003) has also provided many similar suggestions for constructing reading assessments. When scoring reading assessment items that are not objective, such as written and short answer items, Hughes (2003) warned that penalizing test-takers' errors in "grammar, spelling, or punctuation" is problematic because then productive skills are being tested, not simply the reading skills (p. 155). In fact, Hughes (2003) stated that testing both productive and reading skills "at the same time . . . simply makes the measurement of reading ability less valid" (p. 155).

To accurately place students in the reading stream courses, I have determined that our ELP should develop an in-house reading instrument. Researchers have provided many suggestions regarding test construction and content that can be of use for an in-house reading test (see Table 6). Based on the findings, I also recommend that our program not use productive skills in our reading test. As fair use copyright laws allow, I suggest that we select appropriate passages from the textbooks from all levels in our reading courses and construct multiple choice items that assess for the skills outlined in the course objectives.

Table 6

Reading Stream

Suggested Forms	**Item Types**	**Reading Text Sources**
*Selected response items (Brown & Hudson, 1998; Hughes, 2003)	Multiple-Choice	Passages from reading stream textbooks
Constructed response items	Gap Filling, Cloze (Hughes, 2003); Written, short answer (Hughes, 2003)	Passages from achievement tests (Jamieson, et al., 2013)
		Passages from openly available sources

Writing. In-house placement tests for determining proficiency in producing written English can be designed as well and are the most common placement assessments given in ESL programs in colleges and universities (Ling et al., 2014). James and Templeman (2009) cited research suggesting EAPs include essay-writing on placement tests. Writing tests often contain a prompt which students must write about in a set amount of time (Park, 2003; Ferris, et al., 2017; Lee & Anderson, 2007). Essay prompts are one possibility of direct placement testing. For example, Ferris et al. (2017) discussed a successful locally developed L2 writing placement test that included an essay prompt with a rubric designed to correlate with each level in the EAP and

reported that this placement instrument had resulted in few misplacements. Park (2003) indicated that a "timed impromptu essay test" (p. 3) can be an apt description of a writing test since those taking the test must immediately and quickly write an essay with no advanced knowledge of the prompt. Park (2003) added that a program might want to include several writing tasks in the prompt for reliability since some subjects pose more challenges or result in stronger writing. Yannakoudakis, Andersen, Geranpayeh, Briscoe, and Nicholls (2018) noted that when creating writing tasks for different proficiency levels on placement tests, test developers should be aware of task bias and consider using prompts of similar topics and registers to avoid biases that can occur with different topics and genres. Lee and Anderson (2007) have presented research that indicates the importance of considering several factors that affect student performance on a written placement instrument. The academic topic, the major of the test-taker, and the student's overall proficiency in the language are all necessary considerations for valid placement decisions, especially if topics are often rotated in placement testing procedures, as test outcomes vary widely depending on these test features (Lee & Anderson, 2007). As with the speaking assessment, TOEFL has independent and integrated writing rubrics available online (ETS, n.d.) that our program could use, or an in-house rubric could be created.

Table 7

Writing Stream

Suggested Forms	**Item Types**	**Example Prompt Sources**
Performance items	Timed essay prompt	Prompts from writing stream textbooks
		Prompts from openly available sources

To place students in the writing courses, I have determined that our ELP should develop and administer an in-house writing instrument. As presented in Table 7, these findings suggest

that timed writing prompts are commonly used with success and seem a viable option for our program. A scoring rubric based on the levels in our EAP, such as the successful placement essay prompt and rubric discussed by Ferris et al. (2017), is also recommended.

Vocabulary. A locally-developed placement tool for the vocabulary stream can be administered in conjunction with the reading or the writing assessment, or it can be developed as a separate test. Hughes (2003) noted that although vocabulary is indirectly tested in reading assessments, he recommended assessing vocabulary with a vocabulary test. Vocabulary is not to be overlooked as Kress (2008) has suggested that "the more extensive the students' vocabulary, the better able they will be to express their ideas," and assessors should be looking for "students who have not progressed beyond literal meanings of words; they will have difficulty with idioms, synonyms, register, and tone" (p. 333). Kress (2008) added that it is important to be vigilant for "students who overgeneralize an English language element (for example, using *childrens* as a plural) or apply a grammatical structure from their native language to English" (p. 233) when assessing students' structural vocabulary knowledge. Kress (2008) has presented many methods for developing vocabulary assessments. To assess vocabulary knowledge, an in-house test can include items that ask students to recognize analogies, use context clues, and recognize, supply, and interpret antonyms and synonyms (Kress, 2008). Harrington and Carey (2009), discussed research regarding another intriguing option for a reliable locally-developed vocabulary placement tool in the form of an online Yes/No assessment that measures knowledge of both general vocabulary and program-specific vocabulary. The results of this test were successfully used in combination with other in-house placement tests when considering student placement in the appropriate level in the institution's language program (Harrington & Carey, 2009).

Table 8

Vocabulary Stream

Suggested Forms	**Item Types**	**Target Vocabulary Sources**
Selected response items (Hughes, 2003)	Multiple-Choice	Vocabulary from vocabulary stream textbooks
	Choose one or another	Vocabulary from openly available sources
	In-house Yes/No Online tool (Harrington & Carey, 2009)	

As with the other modality streams discussed, I recommend we develop an in-house test for the vocabulary stream; furthermore, I have decided that we should strongly consider Hughes's (2003) suggestion: testing vocabulary in its own assessment. Researchers have provided several assessment item options for consideration, such constructing as multiple-choice items based on our text books that measure student performance on vocabulary objectives in our courses (see Table 8).

Syntax. *Syntax* is defined as "the rules of sentence formation; the component of the mental grammar that represents speakers' knowledge of the structure of phrases and sentences" (Fromkin, Rodman, & Hyams, 2011, p. 595) while *grammar* means "the mental representation of a speaker's linguistic competence; what a speaker knows about a language, including its phonology, morphology, syntax, semantics, and lexicon. A linguistic description of a speaker's mental grammar" (Fromkin, et al., 2011, p. 580). Even though these definitions reveal that syntax is a part of grammar, I discovered that the term *grammar* rather than *syntax* is commonly used when referring to language assessments that could apply to our syntax modality stream. Ferris et. al (2017) stated that many EAPs in the United States and around the world create and administer in-house placement tests that measure students' language skills, including grammar.

Hughes (2003) noted that testing grammar as a skill by itself has become a debated topic in language assessment; however, he stressed that "there is often good cause to include a grammar component" in placement tests "of teaching institutions," adding that a grammar test should be developed to test only grammar, meaning that preferably grammar knowledge should not be measured in a reading or writing test (p. 173). James and Templeman (2009) cited research that also suggested that English grammar items be included on EAP placement testing. Hughes (2003) added that grammar test content be related to learning materials and the objectives in courses in all levels in a program, and suggested that it can also contain items that are assumed to be known by the lowest-proficiency test-takers. Regarding test items, to assess knowledge of language structure, Kress (2008) suggested test items that include editing written passages, rewriting passages to an indicated register, and using cloze procedures. Kress (2008) advised other methods, such as writing paragraphs and essays, unscrambling sentences to assess knowledge of word order and expanding upon existing sentences with appropriate word forms. Regarding the assessment of grammar knowledge, Brown and Hudson (1998) suggested that using multiple-choice items can be used for testing receptive skills, including grammar knowledge, asserting that "multiple-choice items can provide useful information about students abilities or knowledge in those areas with relative efficiency," but may also "have limited the types of language skills that were tested in . . . grammar" due to overuse (p. 660). An example of an in-house grammar placement test for an ELP was discussed by Harrington and Carey (2009). They indicated that the program used results from the writing and speaking tests for main placement decisions while the listening and grammar tests were used for predicting levels. The paper-based, 45-minute, three-part grammar test consisted of

> assessing knowledge of English syntax (word order). Part one consists of eight grammatically ill-formed sentences that must be reordered; part two requires the identification and correction of one grammar error in each of eight sentences, and part three requires the completion of a gapped descriptive passage in which 18 grammatical features have been removed. (p. 616).

This example of an in-house grammar test illustrates the varied techniques available for grammar tests. Hughes (2003) asserted that "gap filling, paraphrase, completion, and multiple choice" can meet most grammar assessment needs and whatever techniques are selected for grammar tests, "the text of the item [should] be written in grammatically correct and natural language" that are "corpus based examples" (p. 174).

Table 9

Syntax Stream

Suggested Forms	**Item Types**	**Target Syntax Sources**
Selected response items (Brown & Hudson, 1998; Hughes, 2003)	Multiple-Choice	Construct similar items used in syntax stream textbook
Constructed response items	Editing written passages, cloze (Kress, 2008), completion, gap filling (Hughes, 2003)	Construct similar items from openly available sources
Performance items	Written paragraphs, paraphrase (Hughes, 2003; Kress, 2008)	

Based on the researchers' suggestions as presented in Table 9, I have determined that MSUM's ELP develop an in-house placement test for the syntax modality stream; specifically, we should, as Hughes (2003) recommended, test students' grammar abilities with one grammar instrument rather than measuring them in combination with reading or writing. We should base

the test items on objectives in the syntax stream courses. I recommend constructing multiple-choice items similar to items in our course textbooks at all levels and considering using gap filling and completion as well.

Chapter 6: Conclusion

One research question guided the overall project: What type of placement instruments should be used for the modality streams in our ELP? University administration has decided that responsibility for placement testing rests on the ELP. The ultimate goal of the ELP is to best match students' abilities with specific courses in the program. The program is structured with seven different modality streams, six of which address language skills. The placement tests need to provide a match between students' abilities with listening, speaking, reading, writing, vocabulary and syntax and the courses offered in these six streams. The seventh stream, cultural knowledge and adjustment, does not enroll students strictly on a language ability basis, so the placement test need not address it.

Proficiency tests, although used in some university contexts for placement, have been determined to be invalid for MSUM. The test MSUM uses for admissions purposes, the TOEFL, integrates language skills on test items, whereas MSUM ELP separates skills in modality streams. Therefore, MSUM will not attempt to extend the use of its admissions test, the TOEFL, to the placement of students in the ELP courses.

Ling et al., (2014) have discussed research confirming that "both standardized language tests (e.g., ACCUPLACER ESL) and in-house tests could be used for ESL placement" (p. 2), and both have beneficial features. Practicality considerations are one area where the two sources of tests differ from each other. Time and cost in particular are relevant in determining whether to purchase a commercial test or develop and use a test in-house. Commercial tests increase costs on an initial and perhaps ongoing basis, while in-house tests require large amounts of time to

create, adapt, pilot, and validate. As budget constraints continue at our institution, cost is currently a more significant factor than time. At this point, printing locally-developed placement assessments at our university would cost far less than purchasing commercial versions. As a result, developing an in-house placement test seems the most prudent option, and based on the research, non-DSP in-house placement tests are recommended for the modality streams in our program.

Table 10

Decisions: Placement Testing Instruments

Modality Stream	**Form**	**Text and Item Sources**
Listening	Selected-response items	Lectures and recordings from all levels in course textbooks
Pronunciation	Performance items	Interview
Reading	Selected-response items	Passages from all levels in course textbooks
Writing	Performance items	Course textbooks and other openly available sources
Vocabulary	Selected-response items	Vocabulary from course textbooks and other openly available sources
Syntax	Selected-response items	Items from course textbooks and other openly available sources

In order to expedite the development of in-house tests, MSUM will adopt and adapt existing course materials and other openly available materials. Items on all placement instruments will correspond with course objectives in the streams. Table 10 indicates decisions for each modality stream. For the listening stream, I have decided that an in-house test be

developed using lectures and recordings from all levels in course textbooks and writing corresponding selected-response test items that align with course objectives. For the pronunciation, or speaking, stream, I have determined that an interview format will be used and a rubric developed based on the objectives of each course. Constructed-response and performance items will be used on this instrument. The placement instrument for the reading stream will contain passages from textbooks at all levels; selected-response items in the form of multiple-choice will be written for these passages. For the writing stream, I have decided that the test items should be performance based, a prompt or set of prompts with multiple writing tasks should be developed along with a rubric that corresponds with writing course objectives. For the vocabulary stream, I have determined that an in-house instrument that contains selected-response items be developed following course objectives. An instrument with selected-response items will also be developed and implemented for the syntax stream adapting items from course textbooks and other openly available sources words.

Recommendations for Future Phases of the Project

Further recommendations for our program include the next stages of developing tests as suggested by Hughes (2003), which include details such as writing test items, trialing the tests, making needed changes, writing handbooks for stakeholders, and training raters and interviewers. Furthermore, cut scores or standard setting will also need to be determined for the placement tests in the streams once the tests and rubrics are developed (Shin & Lidster, 2017). The cut scores should be based on the student learning outcomes for the courses (Shin & Lidster, 2017; Parker, 2017). All of these tasks should involve all the faculty of the ELP as well as the program coordinator.

Finally, further phases of this research also include ongoing evaluation of and improvements to the effectiveness of the placement tools for accurate placement (Green, 2018). If, in the future, the program grows and printing costs become a concern, cost comparisons with commercial tests could be performed. Alternately, an in-house computer-based test could be developed.

With accurate placement in the streams guiding the path for developing and continually evaluating the placement instruments for the ELP at MSUM, instructors and students will have confidence in the decisions made to aid each student in their goal to learn English in the United States.

References

Accuplacer testing. (2019). Retrieved February 9, 2019 from https://www.mnstate.edu/asc/placement-testing.aspx

Agcam, R., & Babanoglu, M. P. (2016). Students' perceptions of language testing and assessment in higher education. *Universitepark bulten, 5*(1-2), 66-77. doi:10.22521/unibulletin.2016.512.6

Bachman, L. (1991, Winter). What does language testing have to offer? *TESOL Quarterly*, *25*(4), 671–704. doi: 10.2307/3587082

Banegas, M. (2013). *Placement of English as a second language (ESL) student learners in higher education (Doctoral dissertation)*. Retrieved from *ProQuest LLC* (3561743).

Brown, H.D. & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practice (*2nd ed.). White Plains, NY: Pearson Education.

Brown, J. D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly*, *23*(1), 65–84. https://doi.org/10.2307/3587508

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, *32*(4), 653–675. https://doi.org/10.2307/3587999

Christopher, V. L. (1993). *Direct and indirect placement test scores as measures of language proficiency and predictors of academic success for ESL students* (master's thesis). Retrieved from https://open.library.ubc.ca/collections/ubctheses/831/items/1.0078128

Chung, S., Haider, I., & Boyd, R. (2015, March). The English placement test at the University of Illinois at Urbana-Champaign. *Language Teaching 48*(2), 284–287.

College Board. (n.d.). Retrieved June 7, 2019, from https://accuplacer.collegeboard.org/

College Board. (n.d.). *ACCUPLACER ESL* [PDF of Powerpoint]. Retrieved from https://accuplacer.collegeboard.org/pdf/accuplacer-esl.pdf

College Board, Admitted Class Evaluation Service. (2018, February). *ACCUPLACER Placement Validity Study* [PDF]. Retrieved February 4, 2019, from https://aces.collegeboard.org/aces-studies/accuplacer-placement-validity-study

Crusan, D. (2011). The promise of directed self-placement for second language writers. *TESOL Quarterly: A journal for teachers of English to speakers of other languages and of standard English as a second dialect, 45*(4), 774–780. https://doi.org/10.5054/tq.2010.272524

Crusan, D. (2014). Placement testing. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. doi:10.1002/9781405198431.wbeal1421

ETS. (n.d.). TOEFL® Scoring Guides. Retrieved June 7, 2019, from https://www.ets.org/toefl/institutions/scores/guides/

Ferris, D. R., Evans, K., & Kurzer, K. (2017, April). Placement of multilingual writers: Is there a role for student voices? *Assessing Writing*, *32*, 1-11. doi:10.1016/j.asw.2016.10.001

Fromkin, V., Rodman, R., & Hyams, N. (2011). Glossary. In *An introduction to language* (9th ed., pp. 569-599). Boston, MA: Wadsworth/Cengage.

Green, A. (2018). Placement Testing. In J. I. Liontas, TESOL International Association, & M. DelliCarpini (Eds.), *The TESOL Encyclopedia of English Language Teaching*. doi:10.1002/9781118784235.eelt0377

Gewertz, C. (2012, Aug. 14). Consortia provide preview of common tests. *Education Week*, *32*(1), 1.

Harrington, M., & Carey, M. (2009). The on-line yes/no test as a placement tool. *System: An International Journal of Educational Technology and Applied Linguistics 37*(4), 614–626).

Hoekje, B., & Linnell, K. (1994, Spring). "Authenticity" in language testing: Evaluating spoken language tests for international teaching assistants. *TESOL Quarterly*, *28*(1), 103-126. doi:10.2307/3587201

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, England: Cambridge University Press.

International freshman admission & application requirements. (n.d.). Retrieved June 4, 2019, from https://www.mnstate.edu/admissions/international/freshman.aspx

International student population. (2016). Retrieved February 9, 2019, from https://www.mnstate.edu/admissions/international/student-population.aspx

iTEP Academic. (n.d.). Retrieved June 7, 2019, from https://www.itepexam.com/products/itepacademic/

James, C., & Templeman, E. (2009). A case for faculty involvement in EAP placement testing. *TESL Canada Journal*, *26*(2), 82–99. https://doi.org/10.18806/tesl.v26i2.416

Jamieson, J., Wang, L., & Church, J. (2013). In-house or commercial speaking tests: Evaluating strengths for EAP placement. *Journal of English for Academic Purposes*, *12*(4), 288-298. doi: 10.1016/j.jeap.2013.09.003

Johnson, R., & Riazi, A. (2017, April). Validation of a locally created and rated writing test used for placement in a higher education EFL program. *Assessing Writing*, *32*, 85–104. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S107529351630068X doi: 10.1016/j.asw.2016.09.002

Kim, A. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing 32*(2), 227–259).

Kokhan, K. (2012). Investigating the possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing*, *29*(2), 291–308. https://doi.org/10.1177/0265532211429403

Kokhan, K. (2013). An argument against using standardized test scores for placement of international undergraduate students in English as a second language (ESL) courses. *Language Testing 30*(4), pp. 467–489). doi: 10.1177/0265532213475782

Kress, J. E. (2008). Assessment. *The ESL/ELL teacher's book of lists* (2nd ed.). San Francisco, CA: Jossey-Bass.

Lee, H.-K., & Anderson, C. (2007). Validity and topic generality of a writing performance test. *Language Testing*, *24*(3), 307–330. https://doi.org/10.1177/0265532207077200

Lesho, D. (2017a, January 24). The right way to use test data to improve an English language program. Retrieved June 7, 2019, from https://www.linkedin.com/pulse/right-way-use-test-data-improve-english-language-program-dan-lesho?trk=prof-post

Lesho, D. (2017b, March 9). The norming effect of using a standardized English proficiency test within an IEP. Retrieved June 7, 2019, from https://www.linkedin.com/pulse/norming-effect-using-standardized-english-proficiency-dan-lesho/

Lesho, D. & Santoro, C. (2019, March). *Using iTEP as a gate keeping test*. Paper presented at the international TESOL conference, Atlanta, GA.

Leung, C., & Lewkowicz, J. (2006, March). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly, 40*(1), 211-234. doi:10.2307/40264517

Ling, G., Wolf, M., Cho, Y., & Wang, Y. (2014). *English-as-a-second-language programs for matriculated students in the United States: An exploratory survey and some issues.* (Report No. ETS RR-14-11). Princeton, NJ: Educational Testing Service. Retrieved from https://files.eric.ed.gov/fulltext/EJ1109340.pdf

Michigan Language Assessment. (n.d.). Michigan EPT. Retrieved June 7, 2019, from https://michiganassessment.org/institutions/products-services/michigan-ept/

Michigan Language Assessment. (n.d.). Michigan EPT: Quality in test design. Retrieved June 7, 2019 from https://michiganassessment.org/institutions/products-services/michigan-ept/camla-ept-quality-in-test-design/

Minnesota State. (2006). System procedure 3.3.1 assessment for course placement. Retrieved October 12, 2018, from http://www.minnstate.edu/board/procedure/303p1.html

MSUM facts. (n.d.). Retrieved February 9, 2019, from https://www.mnstate.edu/about/facts.aspx

Nolo. (n.d.). Measuring fair use: The four factors. Retrieved February 14, 2019, from https://fairuse.stanford.edu/overview/fair-use/four-factors/

Nunan, D. (2015). *Teaching English to speakers of other languages: An introduction*. New York, NY: Routledge. Retrieved from http://search.ebscohost.com.trmproxy.mnpals.net/login.aspx?direct=true&db=nlebk&AN=955887&site=ehost-live

Park, T. (2003). An investigation of an ESL placement test of writing using many-facet Rasch measurement. *Columbia University Working Papers in TESOL & Applied Linguistics 4*(1), 1-21. https://doi.org/10.7916/D8DR2V1W

Parker, D. (2017, Fall). Guiding questions for evaluating placement tests. *College ESL Quarterly*. Retrieved February 3, from http://www.languageartspress.com/community

Registrar's office. (n.d.). Retrieved June 4, 2019, from https://www.mnstate.edu/registrar/navigator/

Scheduling services. (n.d.). Retrieved June 5, 2019, from https://msum.bookitadmin.minnstate.edu/BrowseFacilities.aspx

Seiler, S. (2016). Validity & reliability report. No location: iTEP International. Retrieved March 2, 2019, from https://www.itepexam.com/wp-content/uploads/2019/01/iTEP-Academic-Reliability-Validity_25JAN19.pdf

Shin, S., & Lidster, R. (2017). Evaluating different standard-setting methods in an ESL placement testing context. *Language Testing 34*(3), pp. 357–381). doi: 10.1177/0265532216646605

Summers, M. M., Cox, T. L., McMurry, B. L., & Dewey, D. P. (2019, February). Investigating the use of the ACTFL can-do statements in a self-assessment for student placement in an intensive English program. *System, 80*, 269-287. doi:10.1016/j.system.2018.12.012

Teaching English as a foreign language (TEFL) accreditation. (2018). Retrieved February 5, 2019, from https://www.mnstate.edu/teaching-english-foreign-language/accreditation.aspx

The TOEFL® Test. (n.d.). Retrieved February 2, 2019, from https://www.ets.org/toefl

Yannakoudakis, H., Andersen, Ø., Geranpayeh, A., Briscoe, T., & Nicholls, D. (2018).

Appendix A

English Language Program (ELP) & Interface with Degree Programs

Level	Pronunciation	Listening	Reading	Writing	Syntax	Vocab	Culture
09x and 1xx ELP Oral Skills	ELP 101 Pronunciation Lab	ELP 102 Conversation I	ELP 093 Reading I	ELP 094 Intro to Writing Conventions	ELP 095 Basic English Syntax	ELP 096 Conversational Vocabulary	ELP 107 Orientation to Campus & Community
2xx ELP Oral Skills	ELP 201 Pronunciation Lab	ELP 202 Conversation II	ELP 103 Reading II	ELP 104 Composition I	ELP 105 Intermediate English Syntax I	ELP 106 Foundational Vocab	ELP 207 Intro to America
3xx ELP Reading	ELP 301 Pronunciation Lab	ELP 302 Discussion & Debate	ELP 203 Reading III	ELP 204 Composition II	ELP 205 Intermediate English Syntax II	ELP 206 General academic vocab	ELP 307 Contemporary America & World Events
1xx TEFL Reading	TEFL 101 Oral Presentations I	TEFL 102 Lectures & Note-taking I	TEFL 103 Academic Reading I	TEFL 104 Academic Writing I	TEFL 105 Advanced Syntax I	TEFL 106 Subject Studies	TEFL 107 American Childhood Classics
2xx TEFL Writing	TEFL 201 Oral Presentations II	TEFL 202 Lectures & Note-taking II	TEFL 203 Academic Reading II	TEFL 204 Academic Writing II	TEFL 205 Advanced Syntax II	TEFL 206 Subject Studies	TEFL 207 American Cultural Classics
B.A./B.S. TOEFL 61	COMM 100 (TSE 22)	COMM 100 (TSE 22)		ENGL 101 & 201 (TWE 17)			
M.A. TOEFL 78							

Appendix B

Certificate of Proficiency in International English

Description
The Certificate of Proficiency in International English recognizes non-native speakers of English who have demonstrated a sufficient level of knowledge of English, literacy skills, and communication skills for the purposes of interacting personally and professionally in English with other English speakers around the world.

Admission Requirements

- Completion of secondary school, verified by a transcript evaluation service
- 1 year of demonstrated English language study
- Financial Self-sufficiency Statement
- US $20.00 application fee

Student Learning Outcomes

- Demonstrate proficiency in English in all four language domains
- Demonstrate knowledge of English morphology and syntax
- Apply knowledge of English to personal and professional tasks in all four language domains, such as listening and speaking in discussions and debates, reading sources of current events, or writing for technical or business purposes
- Demonstrate interpersonal communication skills and active learning strategies
- Develop familiarity with contemporary world events across a variety of topics
- Apply knowledge of contemporary world events to interpersonal communications in appropriate ways with a variety of people

Core Requirements (19 credits)

Skill/Content Area	Courses	Credits
Pronunciation Accuracy	ELP 101 or (ELP 201) or (ELP 301)	1
Oral Skills	ELP 302 or (TEFL 101 and TEFL 201)	3
Reading	ELP 203 or (TEFL 103)	3
Composition	ELP 204 or (TEFL 104)	3
Syntax	ELP 205 or (TEFL 105)	3
Vocabulary	ELP 206 or (TEFL 106)	3
Contemporary World Events	ELP 307	3
		19

*Course titles are listed on the back.
*Note: All courses have prerequisites or placement test score requirements. Students with lower proficiency levels at arrival may need to take additional credits before they are able to enroll in the certificate courses.

Certificate of Proficiency in International English

Core Requirements (19 credits)

Skill/Content Area

Pronunciation Accuracy
ELP 101 Pronunciation Lab I (1) **or**
ELP 201 Pronunciation Lab II (1) **or**
ELP 301 Pronunciation Lab III (1)

Oral Skills
ELP 302 Discussions and Debates (3) **or**
TEFL 101 Oral Presentations I (3) **and**
TEFL 201 Oral Presentations II (3)

Reading
ELP 203 Reading III (3) **or**
TEFL 103 Academic Reading I (3)

Composition
ELP 204 English Writing II (3) **or**
TEFL 104 Academic Writing I (3)

Syntax
ELP 205 Intermediate English Syntax II (3) **or**
TEFL 105 Advanced English Syntax I (3)

Vocabulary
ELP 206 General Academic Vocabulary (3) **or**
TEFL 106 Vocabulary for Subject Studies I (3)

Contemporary World Events
ELP 307 Contemporary America and World Events (3)

Appendix C

Certificate of Academic English Proficiency

Description
The Certificate of Academic English Proficiency recognizes non-native speakers of English who have demonstrated a high level of knowledge of English, literacy skills, and communication skills for academic purposes and recognizes that they are capable of engaging in higher academic study or professional activities in English.

Admission Requirements

- Completion of secondary school, verified by a transcript evaluation service if outside of U.S.
- TOEFL composite score of 54 or higher on the iBT, or completion of the MSUM Certificate of Proficiency in International English
- Financial Self-sufficiency Statement
- US $20.00 application fee

Student Learning Outcomes

- Demonstrate advanced proficiency in English in all four language domains
- Demonstrate advanced knowledge of English morphology and syntax
- Apply knowledge of English to academic tasks in all four language domains, such as essay writing, public speaking, and close reading of discipline-specific texts
- Demonstrate critical thinking skills, independent learning abilities, and active learning strategies for college success
- Develop familiarity with touchstones from a variety of American eras and sociocultural groups
- Apply knowledge of touchstones from American sociocultural groups and history to academic tasks, such as reading literature, engaging in classroom discussions, or writing essays.

Core Requirements (21 credits)

Skill/Content Area	Courses	Credits
Public Speaking	TEFL 101 or TEFL 201 or COMM 100	3
Lectures & Note-taking	TEFL 102 or TEFL 202	3
Academic Reading	TEFL 103 or TEFL 203	3
Academic Writing	TEFL 104 or TEFL 204 or ENGL 101 or ENGL 201	3
English Syntax	TEFL 105 or TEFL 205	3
English Morphology	TEFL 106 or TEFL 206	3
American Culture	TEFL 107 or TEFL 207	3
		21

*Course titles are listed on the back.
*Note: All courses have prerequisites or placement test score requirements. Students with lower proficiency levels at arrival may need to take additional credits before they are able to enroll in the certificate courses.

Certificate of Academic English Proficiency

Core Requirements (21 credits)

Skill/Content Area

Public Speaking
TEFL 101 Oral Presentations I (3) **or**
TEFL 201 Oral Presentations II (3) **or**
COMM 100 Speech Communication (3)

Lectures & Note-taking
TEFL 102 Lectures & Note-taking I (3) **or**
TEFL 202 Lectures & Note-taking II (3)

Academic Reading
TEFL 103 Academic Reading I (3) **or**
TEFL 203 Academic Reading II (3)

Academic Writing
TEFL 104 Academic Writing I (3) **or**
TEFL 204 Academic Writing II (3) **or**
ENGL 101 English Composition I (3) **or**
ENGL 201 English Composition II (3)

English Syntax
TEFL 105 Advanced English Syntax I (3) **or**
TEFL 205 Advanced English Syntax II (3)

English Morphology
TEFL 106 Vocabulary for Subject Studies I (3) **or**
TEFL 206 Vocabulary for Subject Studies II (3)

American Culture
TEFL 107 American Childhood Classics (3) **or**
TEFL 207 American Cultural Classics (3)