Creating a new level-based composition placement test at a growing IEP

Jason A. Dunstan
University of Toledo
A Thesis

entitled

Creating a New, Level-based Composition Placement Test at a Growing IEP

by

Jason A. Dunstan

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Master of Arts Degree in English

Dr. Douglas W. Coleman, Committee Chair

Dr. Anthony Edgington, Committee Member

Mr. Alexander Wrege, Committee Member

Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo

May 2014
An Abstract of

Creating a New, Level-based Composition Placement Test at a Growing IEP

by

Jason A. Dunstan

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Master of Arts Degree in English with a concentration in ESL

The University of Toledo
May 2014

This paper discusses the research and design of a new composition placement test to better fit the rapidly growing population at an intensive English program. The previous placement test no longer fit because it was designed for a much smaller student population. After reviewing the research, new versions of prompts and instructions were designed to use the curriculum directly as criteria, against which to measure a student’s placement. This will decrease the occurrences of misplacement due on discrepancies between language skills, reading ability and writing ability, for example. The new prompts and instructions consider variables of content, linguistic accessibility, task difficulty/variety, and transparency regarding writing evaluation. This can inform any IEPs that are going through similar growing pains with the dramatic increases in international students in the US, year after year.
This paper is dedicated to Kaleena Spivey, without your help, I would not have made it this far without you. Thank you for all your love and support through this, Mocha.

Love you,

Nilla
Acknowledgements

I would like to thank Dr. Coleman for his guidance in this process, and the valuable feedback.

I would like to extend this same thank you to Dr. Anthony Edgington and Mr. Alexander Wrege.

Without my parents, Ken and Sue Dunstan, and their endless patience with my process, I would not have gotten to this point either. I thank you both for this opportunity.
Table of Contents

Abstract iii
Acknowledgements v
Table of Contents vi
List of Tables viii
List of Figures ix
Preface x

I. Placement Testing and the ALI 1
   A. Trends at the American Language Institute 1
   B. Placement rigidity by program size 2
   C. Description of previous placement test 4
   D. Indirect vs. direct assessment 5
   E. Drawbacks to the new CALT 6
   F. Correlation between reading and writing ability 9

II. Prompt Design 11
   A. How students respond 11
   B. Prompt variables 14
   C. Choice in prompts 19

III. The Current Project 21
   A. Goals of the current project 21
   B. Level ranges and curricular expectations 24
   C. Prompt and instruction modification 25
   D. Scoring of writing placement samples 32
E. Looking into the future

References

Appendices

A. ALI Writing levels with curricular goals
B. Prompt Handout for previous composition test
List of Tables

Table 1.1  Student movement after initial placement (Source: ALI Accuplacer pilot study, 2013). ........................................................................................................................................7

Table 3.1  Level exit goals. This table indicates the ALI writing abilities by level based on ALI writing curriculum and the goals needed to proceed to the next level. ........................................................................................................................................23

Table 3.2  Low-Level Prompt Set Examples. This table indicates the original prompt used by the ALI and a suggested modified prompt. .....................................................26

Table 3.3  Low-level Instructions. This table shows the original low-level instructions and suggested modified instructions for writing assessment. .........................27

Table 3.4  Mid-Level Prompt Set Examples. This table describes the original mid-level writing prompts and suggested modified prompts. ........................................28

Table 3.5  Mid-level Instructions. This table explains the original mid-level writing instructions and potential modified instructions. .......................................................30

Table 3.6  High-Level Prompt Set Examples. This table describes the original high-level prompt examples with suggested modified prompts. ..............................31
List of Figures

Figure 1.1 Comparison of IEP student population growth. Proportional comparison of ALI population growth compared to nation-wide enrollment at IEPs (National value divided by 1000 to proportionalize data; based on reports from IIE (2001-2012)). .................................................................1

Figure 1.2 Past, present, and projected enrollment at the ALI. This figure presents the enrollment of students at the ALI from 2000 to the projected Fall 2014 (Source: Enrollment data from the ALI, 2000-2014). ...............................................2

Figure 1.3 First to second term advancement of students in the pilot study data (Source: ALI Accuplacer pilot study, 2013). .................................................................8

Figure 1.4 Second to third term advancement of students in the pilot study data (Source: ALI Accuplacer pilot study, 2013). .................................................................8
Preface

New student placement testing at an intensive English program (IEP) is often an anxious time for both international students and administration. This process shows a student how long they will be in the program and shows the IEP how many new students are entering the different levels. When it is done correctly, placement testing is a valuable step, but if there are problems with the method of testing, it can be, at best, a waste of time, and at worst, additional labor and student frustration.

Proficiency levels in an IEP are often institute-specific, which means that a student at a basic level at one institute may not have sufficient language proficiency for the basic level somewhere else. Each institute sets its own expectations for the levels, therefore there can be some utility in standardized test scores, for example a score on Educational Testing Service’s Test of English as a Foreign Language (TOEFL), but the scores from a test like this are overall, rather than a granular look at the student’s knowledge and ability. For example, a score of 50 on a skill test doesn’t show you what the student knows exactly, but implies a general level of competence. The general level of competence might be adequate as a measure for admission to a university, but at an IEP it is necessary that administrators and faculty are able to see more clearly what students know and what they do not, because there is a curriculum to follow at the IEP.

The curriculum informs placement testing and any associated scores, but when one single test is insufficient to understand the varying skill areas of a student, it requires another measure. For this reason, programs will have a standardized test and a composition test that new students take. The composition test offers a student the opportunity to demonstrate skills that may not have been apparent from the standardized
test, and when the composition test is carefully designed, it can get an IEP targeted
information about a student regarding their fit for particular proficiency levels.

It is this last point that motivates this thesis. A composition test can be designed to
yield level-specific and, therefore, more valuable information to the IEP about each
individual student. This thesis begins by discussing the population trends at the American
Language Institute, its former and current placement testing process, and proposes a new
composition test to supplement the standardized placement test. I review research on the
comparisons between indirect and direct tests, computer adaptive language tests, the
correlation between language skills (e.g. reading and writing), and prompt design
concerns. I finish by comparing the former composition test, highlighting its
shortcomings, and demonstrating how it can be improved to support the current
placement test.
Chapter One

Placement Testing and the ALI

1.1 Trends at the American Language Institute.

The American Language Institute (ALI), an Intensive English Program (IEP) at the University of Toledo, has grown immensely since 2010. This is likely due to an overall increase in international student enrollment at U.S. institutions, specifically IEPs, which has increased by almost 120% to over 110,000 students in 2012 from only slightly over 50,000 students in 2010 (Institute of International Education (IIE), 2001-2012). The data for the ALI shows growth as well, from 84 to 217 students, over a 250% growth. See Figure 1.1 for proportional comparison.

![Figure 1.1: Comparison of IEP student population growth. Proportional comparison of ALI population growth compared to nation-wide enrollment at IEPs (National value divided by 1000 to proportionalize data; based on reports from IIE (2001-2012)).](image)

Although the cause for this nationwide increase is not the focus in the current project, the result of that growth is important to understanding the circumstances at the
ALI. While the data previously mentioned available from IIE only reaches to 2012, the ALI has shown an additional increase of 34% from 2012-2013 and again within the academic year 2013-2014, with a 29% increase. Because of that increase, the current Spring 2014 population sits at 375. With a list of applicants for the following academic year potentially causing another rise of 33%, this project is at once both reactive and proactive. These past increases, the current academic year, and the projected increase is visible in Figure 1. These increases in the past and current academic years coupled with the projected increase is visible below in Figure 1.2.

Figure 1.2: Past, present, and projected enrollment at the ALI. This figure presents the enrollment of students at the ALI from 2000 to the projected Fall 2014 (Source: Enrollment data from the ALI, 2000-2014).

1.2 Placement rigidity by program size.

These projections create a strong desire in the administration to have greater precision in initial placement of students. In a survey done by Kahn, Butler, Weigle, and Sato (1994) in California, they found that with smaller institutions (<130 students) placement rules were more fluid and flexible. The survey included qualitative questions
for each program, which allowed them to elaborate on responses to the survey. The surveys asked for information about the various agencies’ placement, levels taught, and curricular adherence to the statewide core, but the responses on placement are the most interesting.

One of the smaller programs stated that, as a program with a smaller student body, it is difficult to adhere to strict placement rules on a constant basis. The smaller population practically requires that the program be flexible, lest a program suffer a level or course that has unsustainably low student counts, which would lead to fiscal irresponsibility and could have a swift negative impact. The ALI was in this position for the large majority of its 37-year history, but this has certainly changed. It should be noted that this survey was not limited to IEPs, but also included community programs that in some cases did not teach classes above the intermediate level.

With smaller programs, misplacement is a tolerable mistake. If 5% of the students are initially misplaced, the administration must move the students at the beginning of a term. Five to ten misplaced students are not a difficult proposition, especially when there is a small total population, though it is still frustrating to those students. However, if misplacement remains at 5%, this number of students becomes quickly unmanageable as a program grows. For example, at the ALI a mere 5% misplacement rate would require 20 students to be moved, at various levels, involving various instructors, not to mention the labor which goes into the placement of these students, which is seemingly wasted when misplacement occurs.

It is for this reason that a computer adaptive language test (CALT) was adopted. Even with this test, however, there is some degree of error, because it is an indirect test of
skills such as writing and speaking (see Patterson (2000) for a better understanding of CALTs). The shortcomings of indirect measure tests are covered in more detail in a later section. This new placement test uses reading samples and multiple choice answers to determine a reading score, audio conversations with multiple choice answers to determine a listening score, and multiple choice questions regarding various grammar concepts to determine a grammar score. After implementation of this testing measure, students’ scores were tracked for expected improvement over terms and showed to be a reliable metric, that is, scores continued to improve while attending ALI courses.

1.3 Description of previous placement test.

A description of the previous placement test is in order, as it paints a fuller picture of the transition the ALI is going through. The past placement test was entirely hand-scored. New incoming students were given a placement test with the same sections as the new CALT replacement and would answer to the best of their ability. There was one advantage to this system that is lost with the new CALT system, namely a finer look at particular areas where a student had serious problems, and students would be placed where these problems began. The tests were designed in a scaled fashion, where for example there might be 36 questions and each set of six to eight were more difficult than the set preceding them, with the focus of the question relating to a particular concept in the corresponding level at the ALI. It could be seen then in their score, that a student began having difficulty with questions categorized at a Basic 2 proficiency level. See Table A.1 (see appendix A) for a better understanding of the order of ALI’s levels of instruction.

The next placement test for students was a composition to directly assess their
writing ability. After completing the indirect (multiple choice) placement test portion, they were given a sheet of paper that had directions to complete the composition portion. An example of this sheet is available in Table B.1 & B.2 (see appendix B). The instructions are simple, directing the student to write as well as he/she can in the 30 minutes. The student chooses between one of three topics to write about two that were judged as more difficult by the committee when these were originally created, and one that was deemed simpler. There is concern with this choice because it removes some control in the placement test, however this was used in an effort to offer a student multiple topics to ensure their best writing sample (discussed in section 2.3). Students are offered only the paper and a pencil, with no access to any reference material for this test. It was assumed that this offered students the opportunity to demonstrate their command of English through open-ended prompts, or the converse, would allow gaps in student knowledge to appear, in spite of what the indirect CALT yielded.

1.4 **Indirect vs. direct assessment.**

Research shows direct measures are more accurate at assessing students’ abilities, particularly in writing ability, because as Haswell (1998) states, “Indirect testing may be less valid for international students since they often show a larger gap between the skill and habits needed to write extended essays and the facility with the kind of surface features measured in test items” (as cited in Crusan, 2002, p. 21). In other words, this is simply the difference between recognition and recall (Perkins, 1983). In cognitive psychology terms, recognition is the result of having seen something in the past and the ability to choose the correct answer based on that memory, and this contrasts recall, which is the ability to bring something from memory without a visual cue or other
assistance.

Recognition does not demonstrate a thorough understanding of the concept, and certainly it does not demonstrate the student’s ability to use it when appropriate. There is a great deal of faith in direct measures for their accuracy, but there is also the balancing of labor and productivity that must be controlled with them. Direct measures typically require human action, where an indirect measure usually requires very little, yet yields many scores for many students simultaneously. It feels unfair to not offer students the best options with regard to assessment and placement, the balance must be maintained. This is not for the worst, however, as Breland (1983) has done extensive study on the topic, to find that, while direct measures are believed to be a more accurate assessment of a student, a combination of the two is the most effective (Breland, Camp, Jones, Morris, & Rock, 1987).

1.5 Drawbacks to the new CALT.

There are two shortcomings of the CALT employed by the ALI. One of which was mentioned in a previous section, and that is the lack of transparency. If a student scored a 20 out of 36 on the old placement test, the administration would be able to look at each level of question (corresponding to levels in the curriculum content) and determine if a student was weak at a certain level, but perhaps stronger at some higher concepts. With this transparency, the ALI could place a student into the level where they demonstrated weakness. The CALT is opaque. The questions get noticeably more difficult in all sections of the test, but there is no institutional control over which concepts are tested and which concepts yield higher or lower scores.

The ALI has done correlation studies that compare CALT scores to previous
placement test scores (overall), and has also studied student progress in classes. Although
the instruction appears to be working, judged on increasing CALT scores of students,
there was a disconcerting trend with the pilot study group of students. While their reading
scores improved on the CALT from the first administration to the second at the end of
their first term, they did not advance to the next level, and this was most pronounced in
reading and writing courses. See Table 1.1 and Figures 1.3 and 1.4 for a representation of
this data. Table 1.1 shows students’ movements with color. Light red signifies that the
student did not advance in that term, white cells signify that the student did advance a
level into the next term, and, for additional clarity, dark red shows clearly the students
who did not advance in either the first or second terms.

<table>
<thead>
<tr>
<th>ID</th>
<th>First term level</th>
<th>Second term level</th>
<th>Third term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Basic 1</td>
<td>Basic 2</td>
<td>Basic 2</td>
</tr>
<tr>
<td>2</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 2</td>
</tr>
<tr>
<td>3</td>
<td>Intermediate 1</td>
<td>Intermediate 2</td>
<td>Advanced 1</td>
</tr>
<tr>
<td>4</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
</tr>
<tr>
<td>5</td>
<td>Basic 2</td>
<td>Basic 2</td>
<td>Intermediate 1</td>
</tr>
<tr>
<td>6</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
</tr>
<tr>
<td>7</td>
<td>Intermediate 2</td>
<td>Intermediate 2</td>
<td>Advanced 1</td>
</tr>
<tr>
<td>8</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 2</td>
</tr>
<tr>
<td>9</td>
<td>Intermediate 2</td>
<td>Intermediate 2</td>
<td>Advanced 1</td>
</tr>
<tr>
<td>10</td>
<td>Basic 2</td>
<td>Basic 2</td>
<td>Basic 2</td>
</tr>
<tr>
<td>11</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 2</td>
</tr>
<tr>
<td>12</td>
<td>Intermediate 2</td>
<td>Intermediate 2</td>
<td>Advanced 1</td>
</tr>
<tr>
<td>13</td>
<td>Basic 1</td>
<td>Basic 1</td>
<td>Basic 2</td>
</tr>
<tr>
<td>14</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 2</td>
</tr>
<tr>
<td>15</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
</tr>
<tr>
<td>16</td>
<td>Basic 2</td>
<td>Intermediate 1</td>
<td>Intermediate 2</td>
</tr>
<tr>
<td>17</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 2</td>
</tr>
<tr>
<td>18</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
</tr>
<tr>
<td>19</td>
<td>Advanced 1</td>
<td>Advanced 1</td>
<td>Advanced 2</td>
</tr>
<tr>
<td>20</td>
<td>Advanced 1</td>
<td>Advanced 1</td>
<td>Advanced 2</td>
</tr>
<tr>
<td>21</td>
<td>Basic 1</td>
<td>Basic 1</td>
<td>Basic 1</td>
</tr>
<tr>
<td>22</td>
<td>Intermediate 2</td>
<td>Intermediate 2</td>
<td>Advanced 1</td>
</tr>
<tr>
<td>23</td>
<td>Basic 1</td>
<td>Basic 1</td>
<td>Intermediate 1</td>
</tr>
<tr>
<td>24</td>
<td>Intermediate 1</td>
<td>Intermediate 1</td>
<td>Intermediate 2</td>
</tr>
</tbody>
</table>

Table 1.1 Student movement after initial placement (Source: Krull, J.M., 2013)
The underlying cause for this discrepancy is assumed to be a misplacement, not based on the reading comprehension, as it is objectively measured for all students, but that there is either a disconnect between their reading score and their true writing ability or issues with inter-rater reliability.
1.6 Correlation between reading and writing ability.

There has been research done on the correlation between reading and writing development in an individual’s literacy, regarding both first language acquisition (Fitzgerald & Shanahan, 2000; Shanahan, 1984, 2006; Shanahan & Lomax, 1986) and second language acquisition (Carson, Carrell, Silberstein, Kroll, & Kuehn, 1990; Esmaeili, 2002). There is agreement in the field that there is an interconnection between reading and writing ability, but to what extent, and in what way exactly is still debated.

Three concepts that exist are uni-directional, bi-directional, and non-directional (Eisterhold, 1990; Esmaeili, 2002). Uni-directional, which is the most intuitive, in that it is the development of skills in reading, and improvement thereof, improves an individual’s writing ability. It is predicated on the process of learning one’s first language, where a child first learns to read before he/she is able to write and improves writing through reading. The writing-through-reading concept is intuitive because it also follows the listening and speaking pattern, where the more passive act of listening precedes speech (Shanahan, 2006).

The bi-directional and non-directional are quite similar, but the following explanation will attempt to distinguish them. A bi-directional concept is the idea that reading leads to improved writing and that writing leads to improved reading (Shanahan & Lomax, 1986). The bi-directional idea is based on the assumption that shared brain systems are used for the interpretation of meaning in reading and the creation of meaning in writing. Gains in reading improve writing and gains in writing improve reading.

The third concept, a non-directional transfer, is similar to the bi-directional, except there is no transfer of skills between two systems, but rather the two systems are
one and the same. They each originate from the same area of the brain, so one would expect there to be a near 1:1 correlation (which there is not).

All of these concepts are based on a cognitive psychology thought that systems are often shared by different mental processes, especially when there are similarities in the medium (e.g. listening and speaking, reading and writing). The precise way in which the two modalities relate to one another, as mentioned, is not of great importance to this composition test. It is only important that the systems are related, regardless of direction: writing preceding reading, reading preceding writing, or a synchronous progression. It is, however, necessary that we understand there is some degree of connection, because it is this connection that creates the basis for the current project. A composition test must be tailored to the varying degrees of a student’s abilities. As a student’s reading comprehension improves, so do the questions this student receives in testing (using the CALT). Therefore, it logically follows that this student should also see an approximate increase in the difficulty of a composition prompt.
Chapter Two

Prompt Design

2.1 How students respond.

Prompt design has received increasing attention in the last twenty years, especially in ESL. This is credited to the increasing number of international students who arrive to the US to begin or continue their education, no doubt. Prompt design follows a shift from indirect to direct testing measures given to native speakers and domestic students in the last 50 years (for an in-depth review of this shift, see Huot (1990)). Williamson (1993) refers directly to IQ tests from over 50 years ago (an indirect test of one’s aptitude), and the same inherent invalidity is found when a comparison between indirect and direct tests is investigated (Heck & Crislip, 2001). It is exactly that invalidity that the current composition test attempts not to repeat. The test should measure the writing ability for every student, just as IQ tests showed advantages for certain groups. Through diligent review and recreation of the prompts from the previous test, and the creation of new prompts, this will be avoided.

A review of what informs prompt design follows. Kroll and Reid (1994) investigated the varying output by students depending on the prompts received compared to the level of their writing. Smith et al. (1985) is oft-cited because of the close look and comparisons of the type of prompt a student receives and many different factors of the student’s writing sample (e.g. total words, error rate, holistic score, etc.) Three different topic structures were offered to the participants in Smith’s study; an open-ended structure, a response-based prompt with one passage, and a response-based prompt with three passages. All of the writing focused on an instance of creativity, but the source of
material shifted. The open structure simply instructed the writer to describe a time when they felt they were creative and what is involved in an average act of creativity. The second prompt offered the student a single passage about a person doing something creative. The student had to explain how that person was creative and then generalize from that single instance to describe acts of creativity. The third prompt gave three passages, including the 206 word passage (the longest of the three), and asked the writer to describe the actions of each of those three people, and then describe what they felt were average parts of acts of creativity. The participants were split into three levels of writing proficiency; a low level, a mid level, and an upper level, and this three-level split will be revisited in the current project. Smith et al. (1985) found that each proficiency level performed best with a particular topic structure.

Upper proficiency students did best without a passage, the first prompt. Low proficiency students scored higher with one passage than none, and higher still with three passages. The mid-level students had a pattern similar to low proficiency writers, but gained a more significant improvement when they had three passages as source material, instead of one. Smith et al. guesses this is because there is more material for a weaker writer to crib from, and this also would explain why the mid-level students did best with additional material also. Although other variables were studied by Smith et al., the overall holistic score is of most importance. Other values, such as words written, are not as useful unless we know the actual proficiency of the students in the study. Just as proficiency levels vary between institutes, it is difficult to gauge what exactly low, mid, and upper level mean to Smith et al. This study demonstrates two reasons to avoid reading passages in composition tests, one directly and one indirectly. If you offer
reading passages, you are giving students additional material to use and inform their writing, which explains why more passages correlated with higher scores for weaker students. The composition test should measure only writing, so offering as little source material as possible is a more accurate way of assessing it. The implication of this study is that you are also inadvertently assessing their reading ability when you incorporate a passage.

Hamp-Lyons and Mathias (1994) carried out a meta-analysis using historic essays between 1985 and 1989 (a total of 8,583 essays). They took the concept of difficulty, (similar to Smith et al.’s concept of difficulty) and compared an expert’s judgment of prompt difficulty to the average score earned on the prompts. The title of expert was given to ESL writing instructors with at least two years of experience. Hamp-Lyons and Mathias (1994) created categories of prompts that ranged from a private/expository essay topic to public/argumentative. The experts judged the private/expository writing as easiest and the public/argumentative as most difficult. The difference between private and public writing is the source of the information that the author will use. Private writing refers to writing where an author will write about something personal, like opinion, the creativity prompt from Smith et al. (1985) with no passages. Public writing is the opposite, where a student must use external information, such as the 206 word passage from either the second or third prompts from Smith’s study. Both public and private could fall inside the expository category because according to Hamp-Lyons & Mathias (1994), expository writing “[means] factual writing, which also includes description and narration”.

13
There is not an analog for argumentative in the Smith et al.'s study, but this means the author chooses an opinion, and must defend it in their essay. A hypothetical example would be using the three passages from the Smith et al. study and having the writer decide which was the most creative, and convince the reader. The expert judges in Hamp-Lyons and Mathias study came to the same conclusions that their review of ESL writing textbooks and previous studies yielded; private/expository writing is taught first, and progresses into the opposite end, just as they expected, however scores did not move inversely with difficulty.

In the design of prompts for this proposed composition test, it is important to remain acutely aware of these categories. The entire prompt set at any level should remain constant, because without this there is an undeniable validity problem. It is not limited to the student’s part in this, but as Hamp-Lyons and Mathias (1994) suggested, readers might compensate for the writer and think, “This is a hard prompt, the writer is doing well to find this much to say (or, to say it this well)” (p. 60).

2.2 Prompt variables.

Kroll and Reid (1994) identify multiple factors in prompt design; context, content, linguistics, task, rhetoric, and evaluation. After these are outlined, the focus will narrow specifically to the factors relevant at the ALI.

Contextual variables refer to the setting where the writing will actually take place. It is concerned with a situational validity of sorts. If a student has spent a large majority of time writing on a computer through earlier education and will write most of the papers in the future using a computer, there would not be high contextual validity if they wrote a placement exam with paper and pencil, for example. The ALI incorporates, both, in-class
writing on paper and larger written assignments where a computer is used, so at least for the placement purpose in the ALI, there is not concern with this kind of validity. There is research, in fact, that shows allowing the use of a computer (and longer time in this specific study) does not result in a significant improvement in a student’s score (Kenworthy, 2006), only fewer spelling and minor grammar errors.

Content variables are always of high importance during the prompt design stages because it refers to the accessibility of a topic to the writer. While the topics must be general enough that all students can access information from their experiences for the lower and, potentially, the mid-level proficiencies, as the proficiency increases, the topics can and should become more difficult. There is research by Lim (2010) that suggests accessibility only begins to be especially important into the graduate levels, where a student would be more likely to have a wealth of knowledge on one specific subject, but that for undergraduate students, everyday topics are effective. Everyday is a subjective, but through careful consideration this variable can be mitigated. Using the ALI population as an example, there are large numbers of students from Saudi Arabia and China, and these students make up over 80% of the total population combined. There is not a great deal of overlap, but one inoffensive topic that can be discussed is cars. If a prompt asks a student to consider the impact of electric or hybrid cars on the environment, this may be wholly inaccessible to the student, because they have no experience or knowledge of these technologies. This also excludes students from poorer countries who may not even know much about traditional gasoline cars! It is important to consider seriously any prompts before releasing them into regular testing rotation.
Linguistic variables refer to the language used in the prompt and is also concerned with accessibility, but not with the topic itself (i.e. content variable). The wording of the prompt and instructions is the primary concern here. Kroll and Reid (1994) demonstrate how wording can influence the student responses by citing an instance where an ESL student discussed cement when instructed to give “concrete examples”. Although this may seem a humorous response, it highlights the importance of clarity in the instructions and to be sure that each prompt is understood by the target proficiency level. Lim (2010) mentions clarity in instruction with regard to the task variable (covered in the next paragraph), but Lim also considers the effect of simple word count of a prompt on a writer. A prompt with very detailed instructions could possibly be lost on a low-level writer, and so prompt length is something that must be considered alongside word choice.

Task variables, as mentioned above, are concerned with the instructions given within the prompt. It refers to the expected function of the essay. Is the writer simply describing a personal event or familiar people (a low-level task), are they comparing and contrasting two things (a mid to mid-high level task), or are they given a list of tasks with varied functions? An example from Kroll and Reid (1994) describes a prompt that may appear to be quite simple, but in their evaluation is a list of 13 individual tasks. It asks the writer to identify advantages and disadvantages of two choices (4 tasks), compare and contrast these advantages and disadvantages (4 tasks) of each choice (2 tasks), and, finally, to choose one (1 task) of the two options and give reasons and examples for that choice (2 task). A strong writer might be able to encompass well all of these tasks in a single essay, but a prompt with such disregard for task variables would certainly unfairly disadvantage an ESL student, especially when a time-limit is included. The tasks set
before a writer must be reasonable within any other artificial constraints that are placed on him/her (e.g. time), and that is clearly not a concern with the task mentioned above. Connor and Carrell (1993) suggest that the tasks laid before a writer do not affect the score, whether all tasks were completed or not (as cited by Lim (2010) in reference to the 13-task prompt). This is something to be aware of in prompt design and scoring. The raters in these examples may not notice nor care if all the tasks are completed, a teacher in a class, ALI or otherwise, would certainly notice if parts of a question are left unanswered.

Rhetorical variables relate to task variables, in that it concerns the function of the essay itself. Kroll and Reid (1994) point to the delicate balancing in this variable between an effective amount of instruction, where a prompt that states: “Life. Discuss.” is not sufficient for any writer to craft a solid paper. A writer would have difficulty knowing precisely what is expected of them. The opposite side of this scale is a prompt with too many instructions, which could result in multiple essays looking very similar and, thus, create difficulties in scoring. The freedom of a prompt will begin as very open with the low level students, allowing the writer to demonstrate in almost any way their command of written English. The freedom will narrow as students move up in proficiency, as certain criteria are expected at certain levels in the ALI (this will be mentioned in more detail in evaluation variables next), and reaching the top of the range, the freedom will return, but the tasks in the prompts will become more difficult, to correspond with Hamp-Lyons and Mathias (1994) difficulty research.

Evaluation variables are variables that should be presented to the writers, but are much more the responsibility of the program to manage. These can be managed through a
number of different methods, including rater calibration, well outlined rubrics, prototypical essay models for varying scores, and more. There is a responsibility on the student that can be made clear, especially with regard to the expectations of the writing. The previous placement test was quite opaque in that respect, as you can see from the example in Table B.1 & B.2 (in appendix B), there is only instructions, which are read aloud to the students before they begin, but not indication as to the criteria that will be assessed in their writing. Further instruction could be given to students based on the prompt they receive, because these essays are criterion-referenced, which means they are measured against the expectations in levels at ALI. They are not compared to the other students (Brown, 1989).

There are specific skills or competencies a student must have to qualify for a level, and for that reason, it might help students who take a mid-level prompt to know that their paragraph structure will receive attention by the readers. This mirrors the reality of a classroom setting, and would therefore offer more contextual validity. If a student chooses a prompt and then doesn’t demonstrate competencies at their level, there placement might be appropriate, but only if the prompt called for it and they did not deliver. They should not be penalized if they did not know they were scored for the quality of their paragraph structure, or essay organization. This is a danger of a direct assessment, and one that the old placement test did not do a good job of handling. This is why it is important to ensure the students are given prompts that are level-specific that also make clear the criteria against which they will be scored.
2.3 Choice in prompts.

An additional factor to be considered on the periphery of prompt design is the idea of prompt choice. It is something that is, more often than not, included in composition placement tests, and therefore warrants some discussion in the current project. Polio and Glew (1996) researched this exact question: *should students have a choice?* They discuss the underlying assumptions to the concept, mainly should a student have a choice between two or more prompts. By offering choice, it lowers the probability that a student is unfairly disadvantaged. Although this point was not clarified in any of the reviewed literature, Polio and Glew (1996) studied writer choice, and found that familiarity with the topic was the most motivating factor in a decision, and although Polio and Glew did not review scores earned, they cite Freedman (1983), who states that any differences in student scores on prompts they found “easy and interesting” and prompts they found “dull and difficult” were not statistically significant. Polio and Glew also cite a study of Canadian high school students by Evans (1979) who found that the easier prompts received lower scores, though he theorized that it was simply a case of the weaker students selecting weaker prompts. Although Evans may or may not have been right, this lends credence to the point that offering students a choice might skew the data.

The international students interviewed in Polio and Glew’s study express a preference of having choice, but this may simply be because they are accustomed to having it. The available research certainly does not point to an improvement of writing scores when choice is present. These data suggest that the ALI’s previous placement test offered unnecessary choice, and ultimately may have disadvantaged students. If they chose a prompt that was too hard or too easy for them, they would not demonstrate their
knowledge well. If there are multiple prompts presented, it is difficult to ensure that students write something that can be measured against the ALI’s curricular criteria.
Chapter Three

The Current Project

3.1 Goals of the current project.

In this chapter, I will use prompts that were already in use for the ALI composition placement test and modify them, where possible, to fit the goals of each level. Then I will discuss briefly the placement process at the ALI and how this new system will work well with the computer adaptive language test (CALT).

All new students at the ALI are given the CALT for placement in each of the courses, which are divided into a reading and writing course, a listening and speaking course, and a grammar course. The results of this CALT are available immediately after the student finishes, but it only yields scores in reading, listening, and grammar, which means the CALT has no direct measure of writing ability. This is where the problem of our placement system occurs.

A student’s reading score may suggest that they fit well in a particular level. As an example, we can imagine a student scores inside the Basic 2 reading range, but as described in Shanahan (2006), reading and writing scores are rarely perfectly in sync. This is the assumed cause of students having to repeat their first reading and writing course at the ALI, as described in chapter 1, and is likely the result of writing ability lagging behind reading ability. If the example student is placed into Basic 2 by means of the reading score, but their writing ability is at a Basic 1 level, they will spend their first term catching up their writing ability to their reading ability. It should also be noted that a student who is not in the appropriate writing course is certainly not receiving the most effective writing instruction, so while they may still improve even in the wrong level, you
can expect that they would not improve as quickly as if they were in the appropriate level.

The CALT has an additional benefit of offering immediate scores for each test-taker, and until this point, this data was not utilized until the following day or sometimes the next week. With this immediate score we can determine the general range of a student’s reading ability, and give them an appropriate composition prompt on the same day as the CALT (see Table 3.1 for level exit goals at the ALI). The level-appropriate prompts minimize the problems that can arise from students choosing inappropriate prompts for their ability (Evans, 1979; Freedman, 1983), thereby increasing reliability between students and validity in the test itself.

The right column lists the concepts and/or skills that must be mastered through the writing curriculum. To be clear, what is listed is the requirements for exit, so for a Basic 2 student to advance to Intermediate 1, they must demonstrate the full paragraphs
<table>
<thead>
<tr>
<th>Level</th>
<th>Prompt set</th>
<th>Exit goals per level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced 2</td>
<td>High</td>
<td>- Academic reading with paraphrasing and/or quotation  - 2-3 page paper - Introduction (4-5 sentences) - Hook, background information, and thesis statement - Conclusion (4-5 sentences)</td>
</tr>
<tr>
<td>Advanced 1</td>
<td></td>
<td>- Academic reading with paraphrasing and/or quotation  - Short essay - Introduction (4-5 sentences) - Hook, background information, and thesis statement - Conclusion (4-5 sentences)</td>
</tr>
<tr>
<td>Intermediate 2</td>
<td>Mid</td>
<td>- Short essay - Introduction (&gt;= 2 sentences) - Hook, background information, and thesis statement - Conclusion (&gt;= 2 sentences)</td>
</tr>
<tr>
<td>Intermediate 1</td>
<td></td>
<td>- Longer paragraph (11-14 - Clear expression of ideas - Effective topic and concluding sentences</td>
</tr>
<tr>
<td>Basic 2</td>
<td>Low</td>
<td>- Full paragraph (&gt;= 8 sentences) - With topic and concluding sentences</td>
</tr>
<tr>
<td>Basic 1</td>
<td></td>
<td>- Paragraph (5-8 sentences) - General to specific organization - Present and past of <em>be</em></td>
</tr>
<tr>
<td>Foundations</td>
<td></td>
<td>- Simple sentences with subjects and verbs - Few/if any, adjectives</td>
</tr>
</tbody>
</table>

Table 3.1. Level exit goals. This table indicates the ALI writing abilities by level based on ALI writing curriculum and the goals needed to proceed to the next level.

The prompt’s appropriateness can be measured against the exit goals in each level of the writing curriculum, which can be viewed on the right of Table 3.1. This table includes only the parts of the writing curriculum that can be written into instructions, as task specificity is a concern for this new placement test as well. If the instructions direct a
student to “write a paragraph of up to 15 sentences”, we can examine how closely the student is able to do this, and how clear the parts of a paragraph are (e.g. topic, support, and concluding sentences). There was a definite lack of clear direction in the previous test. The instructions supplied to all students are identical, and although there are two separate sections for the normal and easier prompts, the instructions are outlined with no clear indication of what students should write. This form of instructions creates a risk of discouraging a low-proficiency student from writing anything. If the student has a clear directive, such as “write eight sentences,” they could feel more empowered, because they can write sentences. If they see this paper with lines covering front and back, they may think the several sentences they can write are insufficient. Setting aside the theoretical affective component, it is advantageous for the ALI to direct students to write towards criteria we use for advancement in writing. By explaining what our top expectation is for their reading range, they would demonstrate their skill more appropriately for the levels we are considering for them.

3.2 Level ranges and curricular expectations.

The level ranges for each prompt set will be low, mid, and high-level sets. The rationale for this division is based on the similarities between goals per level for example, an effective, multi-paragraph essay is one of the criteria for advancement between Advanced 1 and Advanced 2. So for a student whose reading score places them into Advanced 1 or 2, they will receive a prompt from the high level prompt set, and be instructed to write a multi-paragraph essay from the prompt. If their attempt is deemed successful, then they would be placed in Advanced 2, if it is deemed unsuccessful, then
Advanced 1. This same process can be applied with the expectations at the other levels also.

Although Kroll and Reid (1994) outline six variables to consider when designing prompts, for this project, we will focus on only four: the content of responses required by the prompts, linguistic accessibility of the prompt text for each proficiency level, task relevance and difficulty for each proficiency level, and evaluation criteria outlined explicitly in each level’s instructions. We exclude context because the students will be expected to produce timed writings occasionally in their ALI courses, so this is their contextual reality after placement. The rhetorical variable is left out because writing with regard to rhetorical modes does not occur until the terminal reading and writing course, Advanced 2; therefore this variable does not directly cause a change in placement the way composition of a short essay does.

3.3 Prompt and instruction modification.

The composition prompts that will be used for this are in Figure B.1 and B.2 (see appendix B). We will start with the low-level prompt set because it requires the least amount of modification. Due to the simplicity and similarity between the ALI’s Foundation, Basic 1, and Basic 2 levels, these levels are combined as one. The low prompt set, and the “Easier Topic (for students with less English)” can work well for these levels with minimal to no modification to the prompt itself. The instructions must be made clearer for this level, however, and placed at the top of the page rather than following the mid-high instructions. The range that the test had to cover is the cause of the flaw in the instructions; each of these papers had to reach from students in the
Foundation level, virtually no English, to the top of ALI’s curriculum, university level English. You can see the changes to these prompts in Table 3.2.

<table>
<thead>
<tr>
<th>Original Prompt</th>
<th>Modified Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Tell as much as you can about a good friend of yours.”</td>
<td>“Write as much as you can about a friend of yours now, or from your childhood.”</td>
</tr>
<tr>
<td>“Tell about some of the happiest memories in your life.”</td>
<td>“Write about your average day now, or from your past.”</td>
</tr>
</tbody>
</table>

Table 3.2. Low-Level Prompt Set Examples. This table indicates the original prompt used by the ALI and a suggested modified prompt.

The first prompt was changed only slightly so it offered the opportunity to write in the past tense. The second required a complete change to offer this, as memories cannot occur in the present tense, and it demonstrates an issue of equity between two prompts that were intended to be for the same proficiency range. These two prompts do not call for temporally similar answers. With the modified version of prompt 2 in Table 3.2, they both now include the opportunity for the student to demonstrate their command of the present tense, as well as the past tense, but past tense is not required for completion. This again creates a ceiling to reach for those students who are able, but leaves the opportunity open for students who don’t have the knowledge of past tense. The modifications to this prompt set’s instructions are in Table 3.3.
<table>
<thead>
<tr>
<th>Original Instructions</th>
<th>Modified Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>“This is a writing test. We want you to write well. We are more concerned with how well you write than how much you write. But you should write as much as you can carefully and thoughtfully write in the time given. You may not use a dictionary or computer. You will have 30 minutes to choose a topic, organize your ideas, then write. When you have finished, take a few minutes to check your work.”</td>
<td>“This is a writing test. We want you to write well. You have 30 minutes to write at least five (5) sentences. (≥5 sentences) Write the best paragraph you can in 30 minutes. Check your sentences and your paragraph.”</td>
</tr>
</tbody>
</table>

Table 3.3. Low-level Instructions. This table shows the original low-level instructions and suggested modified instructions for writing assessment.

These instructions have been greatly simplified. The linguistic variable is a primary concern in the low-level prompt set, but it is in the instructions that are unnecessarily complex. By simplifying them and giving them a composition sheet that only has this text on it, it should provoke less anxiety from the students around this level, as they can understand very little of the initial instructions, which are, of course, directed at the more proficient students.

For the mid-level prompt set, we are now looking to draw out effective paragraphs of more than 8 sentences. This level will also examine introductory and concluding sentences, and although it is possible to leave this expectation out of the instructions, it is reasonable that it be explicit. For if a student is able to produce adequate introduction and concluding sentences on command in a longer paragraph, that would serve as the demonstration of the skill, because they still must: 1) have the knowledge of
what these parts of a paragraph are, 2) be able to demonstrate them, and 3) produce a longer essay, of up to 14 sentences for placement in the Intermediate 2 level. The instructions will reflect these expectations while also offering more clarity without risking undue anxiety for the student. You can see the original and modified prompts at the mid-level in Table 3.4.

<table>
<thead>
<tr>
<th>Original Prompt</th>
<th>Modified Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>“In most situations, would you personally prefer to be a member of a group or to be the leader of the group? Use specific reasons and examples to support your answer.”</td>
<td>“In a group situation, do you prefer to be a leader of the group, or just a member? Explain why you prefer a leader or member position. Use 1-2 examples to illustrate your preference.”</td>
</tr>
<tr>
<td>“New technologies have advantages and disadvantages. They can make life easier in a certain way, but at the same time cause unexpected problems, social change, etc. Write about the positive and negative effects of one or two recent new technologies.”</td>
<td>“There are many new technologies in the world today, for example, personal computers, smartphones, or tablets like the iPad. Choose a technology and explain the ways (good or bad) it has affected your life. Use 1-2 examples to illustrate this.”</td>
</tr>
</tbody>
</table>

Table 3.4. Mid-Level Prompt Set Examples. This table describes the original mid-level writing prompts and suggested modified prompts.

For the mid-level prompt set, you will notice there were more changes to these prompts than to the low-level prompts. The rationale for the changes to the first prompt in this set involves, again, the variable of linguistic accessibility, but also the task variable. If a student’s top expectation for this range is simply a long, cohesive paragraph, the original prompts requested too much from the writer. The jump from the easier topic to the, what I will call, normal topic is too great. These two levels are essentially the same, with the only difference being paragraph length and two paragraphs’ interconnection. The prompt now calls for one example, fitting in a single paragraph for Intermediate 1, or two
examples, which would give the Intermediate 2 students enough instruction to write two paragraphs. Intermediate 1 should be capable of a short paragraph, and Intermediate 2 should be able to write two internally cohesive paragraphs that also transition from the first to the second well. When one considers that the students beginning the Intermediate 2 reading and writing class have just mastered a cohesive paragraph with 11-14 sentences, the prompts should reflect that extent of writing.

Referring to the second prompt in the mid-level set, it suffered from the 13-task prompt problem discussed earlier, in section 2.2. Itemizing each task, the student must write about both positive and negative effects (multiple) of one or possibly two new technologies. Limiting the number of tasks in this prompt allows the student to answer all the questions fully without needing more than the instructed target.

Another change to the second prompt is, rather than discussing society at large, the scope is minimized to only involve the student’s experience. Narrowing the content keeps the prompt in the private category. As Hamp-Lyons and Mathias (1994) found, the private category is easier for low- and mid-level students. For the longer writing for the upper levels, public writing might be a more reasonable expectation, but at this level, it is still too high. We must keep in mind the length expectations from these levels in the curriculum, and that helps to make our prompts more precise for the target.

The offering of a couple possible technologies is supported as well, by on Smith et al. (1985). The brevity of these suggestions (i.e. “personal computers, smartphones, or tablets…”) minimizes any chances that there is an advantage or disadvantage caused by reading ability, but it offers a few common technologies for the student to write about. The specific mention of the iPad is strategic because the ALI faculty deliver all course
content to students using the iPad, and many students have stated that their applications to ALI were based exclusively on that fact.

The instructions can be made more complex for this level, too, but will explain precisely what is expected of the student, and will set the expectations higher. See these changes in Table 3.5.

<table>
<thead>
<tr>
<th>Original Instructions</th>
<th>Modified Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>“This is a writing test. We want you to write well. We are more concerned with how well you write than how much you write. But you should write as much as you can carefully and thoughtfully write in the time given. You may not use a dictionary or computer. You will have 30 minutes to choose a topic, organize your ideas, then write. When you have finished, take a few minutes to check your work.”</td>
<td>“This is a writing test and we want you to write well. It is more important that you write well, than how much you write. Write as much as you can carefully. You may not use a dictionary or computer. You will have 30 minutes to organize your ideas and complete your writing. Write 2 paragraphs in the 30 minutes. Remember to use the parts of effective paragraphs for this test.”</td>
</tr>
</tbody>
</table>

Table 3.5. Mid-level Instructions. This table explains the original mid-level writing instructions and potential modified instructions.

The instructions are much closer to the original instructions with this prompt set, but there are still differences, such as the inclusion of expectations. It adds transparency makes the response personal, decreasing the task difficulty. The instruction to write at least two paragraphs also challenges the Intermediate 1 students, which creates a minimum for them to aim for. At this level and higher, the goal is to place them into one of the two levels, so setting higher expectations is acceptable, because we are attempting to place students more accurately and this action of “weeding out” the students from the
upper level in the prompt set will work best. A similar effort will be used with the high-
level prompt set, too.

Using the remaining two prompts in Figure B.1 (see appendix B), we will now
modify the prompts and adjust the instructions to target our goals for the Advanced 1 and
Advanced 2 levels. The two example prompts can be seen in Table 3.6.

<table>
<thead>
<tr>
<th>Original Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Some people argue that the most important years of a person’s life are the childhood years (from birth to 12 years of age). What do you think? Do you agree or disagree? Use specific reasons and examples in your answer.”</td>
</tr>
<tr>
<td>“Many cultures interpret the term friendship differently. How would you define friendship? What are some of the qualities you would expect to find in your close friends? How would you expect them to behave? Do you see much difference between your expectations and expectations of close friends in any other cultures?”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modified Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Some people argue that the most important years of a person’s life are the childhood years (from birth to 12 years of age). Do you agree? Why? OR Do you disagree? When do you think the most important years of life are? Use specific reasons and examples to support your opinion and persuade your reader.”</td>
</tr>
<tr>
<td>“Many cultures interpret the term friendship differently. How would you define friendship? What are some of the qualities you would expect to find in your close friends? What differences do you expect of Americans regarding friendship, based on what you know?”</td>
</tr>
</tbody>
</table>

Table 3.6. High-Level Prompt Set Examples. This table describes the original high-level prompt examples with suggested modified prompts.

As you can see from Table 3.6, the high level prompts are quite similar to the original, however the modified prompts a more simplified and far more readable appearance for students. It is justified by awareness of the task variable. Unnecessary
questions were removed, such as “What do you think?” and “How would you expect them to behave?” Though a student might be able to provide an answer to these questions, they are redundant. “What do you think?” would be answered in their agreement or disagreement with supporting examples, and “How would you expect them to behave?” would be answered in the description of a friend’s qualities. They tie into other questions too closely, and could be difficult for the student to separate from each other. By using the questions in this fashion, it creates an implicit outline that they can follow with their response. If the student is able, this outline will lead them to give a paragraph for each question, and, in effect, yield an adequate writing sample.

Although this is only a demonstration of what can be done using six example prompts at the ALI, it will be easy for a committee to come together using this thesis as a guide for more prompts. There is still more work to be done, but with the modifications made to both the prompts themselves and the instructions for each of the ranges, the ALI will have much more precise placement of new students. This new composition placement system will work alongside the CALT to bring in students and place them more quickly as well. The final part is the scoring the writing samples, but this, too, is made much easier with the new system.

3.4 Scoring of writing placement samples.

The simplification of the placement test, and the creation of the ranges of prompts should cause very high inter-rater reliability. Inter-rater reliability is when a writing sample or essay is rated by two individual scorers and their scores are in agreement. For many composition tests, for example, the Test of Written English by Educational Testing Service uses a scale of 1-6, ranging from very little control of written English at 1, up to
near-native control, eloquence even, at 6. For tests like the Test of Written English and others with a broader scale, there is often a reasonable concern for inter-rater reliability, because the more increments in any rating scale, the higher the probability of disagreement. The system in this composition test brings the scale down to a two-point scale (or three-point for the low-level prompt set). The TWE uses a range that includes nearly every English learner. Its designers have deemed a 6-point scale sufficient discrimination for its market and purpose (e.g. admission, matriculation, etc.). The composition test in this thesis benefits from its institute-specific criteria and function. It has a clear curriculum on which to be measured and this facilitates the level-specific prompts, thus it has a much narrower range per prompt set.

The narrower range of two (or three) proficiency levels based directly on the ALI curriculum. When a student finishes their level-specific composition, the scorer must only decide between a two level split, for example Intermediate 1 or Intermediate 2. If a student demonstrates mastery of the criteria that is learned in the lower of the two levels, they can be confidently placed in the upper level of the range. As in the example earlier in this chapter, if a high-level student writes a composition, and it demonstrates a good understanding of a short essay, then this student can be placed in Advanced 2, however if it does not demonstrate this understanding, then they will be placed in Advanced 1.

3.5 Looking into the future.

Research shows that a scorer who is an expert in a particular level is more accurate at placement around that level (Hamp-Lyons & Mathias, 1994). This makes intuitive sense, because someone who is familiar with the writing expectations could make the fine distinctions between a Basic 2 writer and a Basic 1 writer. For this reason,
the faculty at the ALI will be trained during the pilot of this placement test to recognize writing that is demonstrably around their level of expertise. They will also have calibration practice before they begin rating samples, too, which minimizes any disagreements that might occur, though again, with a scale that is as narrow as this, inter-rater reliability should be near 100%. Using the 5-year research plan designed for the Test of Written English after it was launched (Stansfield & Ross, 1988), the ALI will have a limited rollout of prompts to the different levels, and collect these writing samples, but they will not be used for placement until they are statistically reviewed.

From these validated samples we can develop prototypical models that are indicative of particular levels, and these models will be chosen by instructors who teach at the level(s) in question to make final judgments of the level of the sample. Over the summer of 2014, the ALI will be able to collect many writing samples from the classes, and find the best-fitting prototypes to use for calibrations before the beginning of fall term, when we have an intake of new students. Questions of validity and reliability will be part of a follow-up study of this new system, as we will need to see if it is truly effective, and increases the precision of our placement after the necessary pilot study research is complete.
References


### Appendix A

**Table A.1. ALI Writing Levels with Curricular Goals**

<table>
<thead>
<tr>
<th>Level</th>
<th>Prompt set</th>
<th>Exit goals per level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced 2</td>
<td></td>
<td>- Academic reading with paraphrasing and/or quotation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 2-3 page paper</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Introduction (4-5 sentences)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Hook, background information, and thesis statement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Conclusion (4-5 sentences)</td>
</tr>
<tr>
<td>Advanced 1</td>
<td>High</td>
<td>- Academic reading with paraphrasing and/or quotation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Short essay</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Introduction (4-5 sentences)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Hook, background information, and thesis statement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Conclusion (4-5 sentences)</td>
</tr>
<tr>
<td>Intermediate 2</td>
<td>Mid</td>
<td>- Short essay</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Introduction (&gt;= 2 sentences)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Hook, background information, and thesis statement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Conclusion (&gt;= 2 sentences)</td>
</tr>
<tr>
<td>Intermediate 1</td>
<td>Low</td>
<td>- Longer paragraph (11-14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Clear expression of ideas</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Effective topic and concluding sentences</td>
</tr>
<tr>
<td>Basic 2</td>
<td>Low</td>
<td>- Full paragraph (&gt;= 8 sentences)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- With topic and concluding sentences</td>
</tr>
<tr>
<td>Basic 1</td>
<td>Low</td>
<td>- Paragraph (5-8 sentences)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- General to specific organization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Present and past of <em>be</em></td>
</tr>
<tr>
<td>Foundations</td>
<td></td>
<td>- Simple sentences with subjects and verbs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Few/if any, adjectives</td>
</tr>
</tbody>
</table>
Appendix B

Figure B.1. Prompt Handout for Previous Composition Test

[Table]

<table>
<thead>
<tr>
<th>New-Student TOEFL Composition</th>
<th>NAME __________________________, __________________________</th>
</tr>
</thead>
<tbody>
<tr>
<td>(family)</td>
<td>(given)</td>
</tr>
<tr>
<td>Student ID Number</td>
<td></td>
</tr>
<tr>
<td>DATE</td>
<td></td>
</tr>
</tbody>
</table>

This is a writing test. We want you to write well. We are more concerned with how well you write than how much you write. But you should write as much as you can carefully and thoughtfully write in the time given. You may not use a dictionary or computer.

You will have 30 minutes to choose a topic, organize your ideas, then write. When you have finished, take a few minutes to check your work.

Choose one topic.

**Topic 1:** New technologies have advantages and disadvantages. They can make life easier in a certain way but at the same time cause unexpected problems, social change, etc. Write about the positive and negative effects of one or two recent new technologies.

**Topic 2:** Many cultures interpret the term *friendship* differently. How would you define friendship? What are some of the qualities you would expect to find in your close friends? How would you expect them to behave? Do you see much difference between your expectations and expectations of close friends in any other cultures?

If the two topics above are too difficult for you to write about, you should be in a lower level and you may write about an easier topic so that we can get an example of your writing and find the best level for you.

**Easier Topic (for students with less English):** Tell as much as you can about a good friend of yours.
Appendix B

Figure B.2. Prompt Handout for Previous Composition Test

<table>
<thead>
<tr>
<th>New-Student TOEFL Composition</th>
<th>NAME ____________________________, ____________________________</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(family)</td>
</tr>
<tr>
<td></td>
<td>(given)</td>
</tr>
<tr>
<td>Student ID Number</td>
<td></td>
</tr>
<tr>
<td>DATE _________________________</td>
<td></td>
</tr>
</tbody>
</table>

This is a writing test. We want you to write well. We are more concerned with how well you write than how much you write. But you should write as much as you can carefully and thoughtfully write in the time given. You may not use a dictionary or computer.

You will have 30 minutes to choose a topic, organize your ideas, then write. When you have finished, take a few minutes to check your work.

Choose one topic.

**Topic 1:** In most situations, would you personally prefer to be a member of a group or to be the leader of the group? Use specific reasons and examples to support your answer.

**Topic 2:** Some people argue that the most important years of a person’s life are the childhood years (from birth to 12 years of age). What do you think? Do you agree or disagree? Use specific reasons and examples in your answer.

If the two topics above are too difficult for you to write about, you should be in a lower level and you may write about an easier topic so that we can get an example of your writing and find the best level for you.

**Easier Topic (for students with less English):** Tell about some of the happiest memories in your life.

______________________________________________
______________________________________________