



2013-06-16

# Assisting Novice Raters in Addressing the In-Between Scores When Rating Writing

Brittney Greer

*Brigham Young University - Provo*

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Linguistics Commons](#)

---

## BYU ScholarsArchive Citation

Greer, Brittney, "Assisting Novice Raters in Addressing the In-Between Scores When Rating Writing" (2013). *All Theses and Dissertations*. 4066.

<https://scholarsarchive.byu.edu/etd/4066>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu).

Assisting Novice Raters in Addressing the In-Between Scores  
When Rating Writing

Brittney Greer

A thesis submitted to the faculty of  
Brigham Young University  
in partial fulfillment of the requirements for the degree of

Master of Arts

Norman W. Evans, Chair  
Wendy Baker-Smemoe  
Troy L. Cox

Department of Linguistics and English Language  
Brigham Young University

June 2013

Copyright © 2013 Brittney Greer

All Rights Reserved

## ABSTRACT

### Assisting Novice Raters in Addressing the In-Between Scores When Rating Writing

Brittney Greer

Department of Linguistics and English Language, BYU  
Master of Arts

In the research regarding rating ESL writing assessments, borderline writing samples are mentioned, but a solution has yet to be addressed. Borderline samples are writing samples that do not perfectly fit a set level within the rubric, but instead have characteristics from multiple levels.

The aim of this thesis is to provide an improved training module in the setting of an Intensive English Program by exposing new raters to borderline samples and rating rationale from experienced raters. The purpose of this training is to increase the confidence, consistency, and accuracy of novice raters when rating borderline samples of writing. The training consists of a workbook with a rubric and instructions for use, benchmark examples of writing, borderline examples of writing with comments from experienced raters defending the established scores, then a variety of writing samples for practice.

The selection of the benchmark and the borderline examples of writing was informed by the fit statistic from existing datasets that had been analyzed with many-facet Rasch measurement. Eight experienced raters participated in providing rationale based upon the rubric explaining why each borderline sample was rated with its established score, and describing why the sample could be considered at a different level.

In order to assess the effectiveness of the training workbook, it was piloted by 10 novice raters who rated a series of essays and responded to a survey. Results of the survey demonstrated that rater confidence increased following the training, but that they needed more time with the training materials to use them properly. The statistical analyses showed insignificant changes, which could be due to the limitations of the data collection. Further research regarding the effectiveness of this training workbook is necessary, as well as an increased discussion in the field regarding the prevalent issue of rating borderline samples of writing.

Keywords: borderline, benchmark, writing assessment, rater error, rater training, language assessment, experienced raters, novice raters, English as a second language, intensive English program, L2.

## ACKNOWLEDGEMENTS

First and foremost I would like to thank my parents, Mathew and Terri Greer, and my brothers and sisters for their constant support and love, faith, countless prayers, and gift of finding humor throughout this process. I would not have done it without you. I am immensely grateful to my committee chair, Dr. Norman W. Evans, for his insights, patience, and sacrifice of time that made this project possible. I would also like to thank my thesis committee members: Dr. Troy Cox for his countless hours of tutelage, guidance, and friendship, and Dr. Wendy Baker-Smemoe for her unfailing positivity and flexibility. Additionally I would like to express gratitude for the prayers, poetry and proofreading, distractions, encouragement, and support of family, friends, roommates, professors, and peers. I am overwhelmed at the outpouring of kindness and goodness from each of you. You know who you are, and I thank you.

## Table of Contents

ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	iii
Table of Contents .....	iv
Index of Tables .....	viii
Table of Figures .....	ix
Chapter 1: Introduction .....	1
Context and Rationale .....	3
Aims and Questions .....	4
Definitions .....	5
Organization of Thesis .....	6
Chapter 2: Review of Literature .....	8
Training .....	9
Rubrics .....	9
Holistic .....	10
Analytic .....	10
Comparison of rubric types .....	11
Anchor Samples .....	13
Borderline Samples .....	14
Experienced Rater Feedback .....	16

Practice Writing Samples .....	17
Rater Error .....	18
Fair Averages .....	19
Summary .....	20
Chapter 3: Methodology .....	22
Instrument Development .....	23
Rubric .....	24
Benchmark examples .....	25
Borderline examples .....	28
Practice writing samples .....	31
Content experts .....	31
Selecting experienced rater rationale .....	32
Research Design .....	33
Subjects .....	33
Crossover design .....	33
Pre-treatment .....	34
Treatment .....	35
Post-treatment .....	35
Structure of Analysis .....	35
Qualitative data collection .....	35

Quantitative data collection.....	36
Chapter 4: Results.....	38
Question 1 and Survey Responses.....	38
Confidence.....	39
Difficulty.....	39
Usefulness.....	40
Reliance upon the workbook segments.....	41
Structural preferences for experienced rater rationale.....	42
TESOL Presentation Feedback.....	43
Question 2.....	43
Pearson product-moment correlation.....	43
Cohen's Kappa.....	44
Means (fixed effects ANOVA).....	44
Question 3.....	45
Summary.....	46
Chapter 5: Discussion.....	47
Question 1.....	47
Confidence.....	47
Difficulty.....	48
Usefulness and reliance upon the workbook segments.....	49

Question 2 .....	49
Question 3 .....	50
Limitations .....	51
Recommendations for Future Research .....	53
General Conclusion.....	55
References.....	56
Appendix A.....	62
Appendix B.....	119
Appendix C .....	124
Appendix D.....	128

## Index of Tables

Table 1: Benchmark Selection Data .....	27
Table 2: Borderline Selection Data.....	30
Table 3: Survey Responses of Usefulness/Helpfulness of Training Workbook Segments .....	41
Table 4: Ranking of Rater Reliance upon Training Workbook Segments and Other Skills .....	42
Table 5: Rater Measurement Report.....	46

Table of Figures

Figure 1: Basic Exposure Model..... 9

Figure 2: A Model to Train Novice Raters of Writing ..... 21

## Chapter 1: Introduction

Every year, worldwide, there are new teachers entering the field of English as a second language (ESL) teaching. Often these new teachers are called upon to rate the performance of students through writing assessments designed to measure English language skills. The outcome of these assessments can ultimately shape the future of the students. Without proper training, teachers new to the field risk inaccuracy and inconsistency in their ratings. This lack of experience and training is then compounded by the challenge of rating papers whose quality is debatable. Indeed, one of the difficulties for novice raters and experienced raters alike is the rating of borderline samples of writing that do not easily fit into the levels of a rubric. Some may say it is acceptable to allow novice teachers time to gain experience, or that some rating inconsistency is inevitable for new raters. While improvement may indeed come with time, the negative impact on examinees can be great.

Consider the following scenario. Two teachers at an Intensive English Program (IEP) have been asked to blindly rate the placement exam of a student named Ricardo who intends to work to pay for school. They have been provided with a holistic scoring rubric and the basic training which has been created by the administration of the program. The essay does not fit neatly into the rubric provided because Ricardo has good command of rhetorical aspects but his grammatical accuracy is weak. One teacher is in her first semester of teaching, and has no experience rating exams. As she rates the essay that Ricardo wrote for his placement exam, she tries to follow the rubric, but looks mainly at his grammar. Based on the rubric and his weakness in grammar, she gives him a rating of

3 out of 7 on his essay. She may see his strength in rhetoric, but does not know if that compensates for his lack of grammatical accuracy. The other teacher is in her second semester of teaching and rating exams. She is not very experienced, but also not entirely new to the process. Her background has led her to believe that fluency and structure are the most important aspects in writing. She too strives to follow the rubric and training provided by the administration, and after reading Ricardo's essay, she gives him a rating of 4 because the essay appears very fluent and is well structured. She sees the grammatical inaccuracies, but is unsure of whether they nullify his good work in rhetoric. Because of school policy, Ricardo needs a rating of 4 in order to qualify for employment.

The administration looks at the scores given by the raters and is unable to determine whether Ricardo should remain in lower level courses due to the score of 3 he received, or move up to the higher level classes because of the 4 he also received. The raters did not know how to rate an essay that did not conform to the rubric definitions, and as a consequence the administration cannot decipher the reasoning behind the scoring discrepancy. While the outcome of this test is high stakes for Ricardo, at this point the administrators do not have enough information to make an educated decision. Lack of confidence in rating such borderline samples can be a grievous problem. For the sake of consistency, and for students like Ricardo, there must be guidance for new raters regarding borderline writing samples, and reliability between raters on high stakes testing. (Brown & Hudson, 2002).

The overarching goal of this thesis is to create a model by which new writing raters can be trained to be more consistent and accurate in their ability to rate borderline writing samples. This goal will be achieved through the development of materials and

practices that will systematically expose novice raters to borderline samples of writing that have been assessed and annotated by experienced raters. The product takes form as a training workbook to be used in Brigham Young University's English Language Center (ELC) to train new raters of writing (See Appendices A & B). While improving the writing rater training at the ELC drove the initial focus of this project, it is defined and designed in such a way as to be applicable elsewhere. The principles defining the model can serve as a guide to administrators of other Intensive English Programs (IEPs) to create training workbooks for their own inexperienced raters. Student placement is a critical and ongoing challenge in any program and writing assessment can be subjective. The goal of this thesis is to minimize the subjectivity through proper training and exposure to writing samples.

### **Context and Rationale**

For many programs, there is a long list of matters requiring the attention of administrators, not the least of which is assessment. As a result of the potential impact assessment can have on students' futures, many programs include some form of consistency training for their raters. Returning to the example of Ricardo, in the absence of minimal training, many more students would face Ricardo's unfortunate situation. However, according to an informal survey distributed to IEP administrators, to be discussed more in depth later, many programs still appear to lack consistency and direction in providing effective training (Greer, 2012). The purpose of this training model is to provide clear guidance for improving training programs designed for new raters of writing assessments.

Previously, the English Language Center at Brigham Young University had a carefully designed training module for new raters. This training was in the form of a workbook containing a rubric and its instructions for use, benchmark samples of writing representing each scoring increment in the rubric, and un-scored samples of actual student writing with a key to allow for rating practice.

The benchmark samples, rubric, practice samples, and the attached key provided new raters with a concrete baseline for understanding how an essay was rated and allowed new raters to improve their rating skills. However, feedback from new raters who went through the existing training suggested a gap in the training to be filled. New raters lack confidence in their own judgment and a solid understanding of how to assess more challenging writing samples. While they may become more confident in rating samples that conform to the definitions of the rubric, they are less confident in rating a borderline sample, which has characteristics from different levels of the rubric. These new raters need more than they currently receive. They require better communication with past raters and more confidence and experience in rating “fence-sitter” or borderline samples. Regardless of the program, there will always be new raters who require additional guidance and training.

### **Aims and Questions**

The aim of this thesis is to provide an improved training module in the setting of an Intensive English Program by exposing new raters to borderline samples and more rating rationale from experienced raters. To assess its effectiveness, such materials were created and given to a group of novice raters to examine if they were more confident in their rating ability. The questions for this research are:

1. To what extent does borderline training with experienced rater rationales increase the confidence of the novice raters?
2. To what extent does borderline training with experienced rater rationales result in more consistent agreement between novice raters and the standards?
3. Following exposure to borderline training with experienced rater rationales, do novice raters have sufficient consistency, or intra-rater reliability, to rate high-stakes tests?

### **Definitions**

The following is a list of words and their definitions for the purposes of this research and study:

*Benchmarks*, also known as *anchors*: Samples of writing which clearly represent each score level on a rubric. For example, in a holistic rubric with ratings one through seven, the benchmarks would be writing samples which clearly represent a one, a two, a three, a four, a five, a six and a seven.

*Borderline* examples, also known as “fence-sitters:” Samples of writing which do not clearly represent any clear-cut score on the rubric. For example, when rated, many raters may defend the sample with a rating of two based on certain characteristics, while many other raters may defend that the sample should be rated as a one based on different characteristics. Personal bias and lack of training make these samples the most difficult to rate.

*Established score* also referred to as *fair average*: An average which takes rater error into account. This average is found by many-facet Rasch measurement (MFRM) - a statistical method for compensating for consistent human error (Eckes, 2011). If a teacher is known as a strict or generous rater, their scores will be adjusted to compensate for their bias. The adjusted score is then accepted as the actual score of the writing sample, and is used as a baseline against which new raters are compared.

*Experienced raters*: Although the term “experienced” implies someone with a great deal of experience, the use of this term at the ELC and for the purposes of this paper generally applies to any rater who is not in his or her first semester of teaching or rating.

*New/novice raters*: Raters who have no past experience with rating in the given program or field.

*Observed Average*: The average which most plainly appears without further analysis of the data. If one teacher rates an essay as a 3 and another teacher rates the essay as a 2, the observed average is a 2.5.

### **Organization of Thesis**

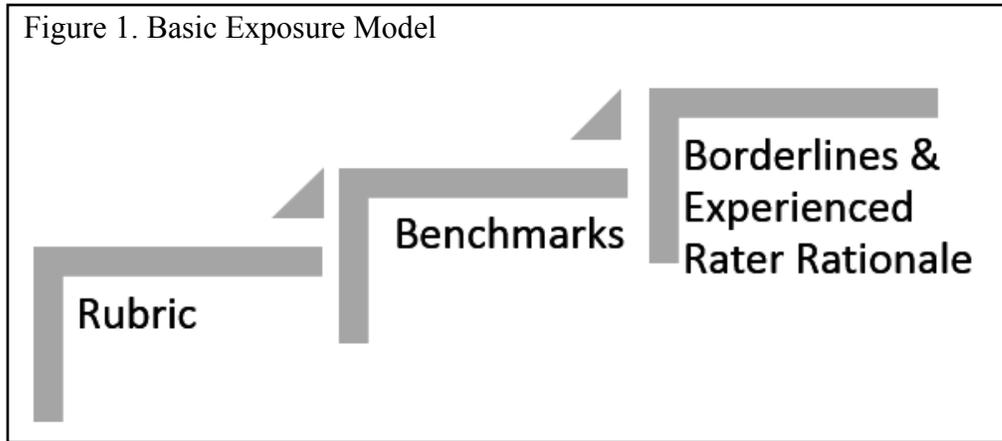
In order to clearly define the theoretical background of the principles utilized in the study, Chapter 2 reviews existing literature. Chapter 3 will clearly define how the training workbook was created, then discuss how the study was organized in order to answer the research questions. It also describes the procedures by which the data was collected and analyzed. The perspectives of the participants regarding the training workbook and other data collected will be shown in Chapter 4, after which the results and recommendations for implementation and further research will be delineated in Chapter

5. The Appendices include the updated training workbook created and resources used in the evaluation of the product.

## Chapter 2: Review of Literature

Since the purpose of this research is to create an improved training model for new raters of writing, the literature review will explore both why and how new raters need to gain experience and confidence, improve consistency, and become more reliable. The various elements which make up a proper rater training will be reviewed in order to: (1) define each element, (2) defend the positive portions of the existing model: such as exposure to a rubric and the use of anchor writing examples and practice writing samples and (3) endorse the use of other elements such as further exposure to borderline samples of writing and experienced rater feedback, which the research will show as beneficial.

The basic model upon which this model expands is the progress from exposure to the rubric to exposure to benchmarks to exposure to borderlines with experienced rater rationale, as shown in Figure 1. However, even after using the best training systems, novice and experienced raters will continue to make errors and will not be perfect raters. In order to inform the reader of the difficulties faced even after implementing the training, some of these common errors will be discussed, as well as strategies to combat post-training rater error. The literature will stand as a foundation to inform the improvement of an existing training model, including necessary post-training treatments. This chapter will begin by detailing what is already known of existing training programs, describe two main types of rubrics and detail their necessity, discuss the need for exposure to a variety of writing samples and the importance of feedback from experienced raters, continue with a discussion of common errors made by raters even after exposure to these resources, then discuss some of the mathematics which can assist in engineering rater reliability.



### **Training**

Because there is little to no empirical research regarding the training of new raters of writing using borderline examples, Greer (2012) conducted an informal survey to surmise what methods of training are being utilized in IEPs. This survey was distributed by means of an official TESOL organization list-serve designed for people involved in IEPs. While only six administrators responded, those who gave specific information regarding their training programs all mentioned the use of a rubric, the use of multiple raters, and some type of calibration in which they assess if the raters are commonly in agreement. Of the six who responded, three emailed in responses without viewing the survey, and the other three completed the survey. Due to this, only three saw and responded to a question regarding the effectiveness of their training programs. All three who were questioned, however, shared feelings that the training was not based in research, and expressed interest in further research providing information regarding a global model of training.

### **Rubrics**

“Rubrics are tools that can help multiple instructors come to similar conclusions about construction of higher-level conceptual knowledge, performance skills, and

attitudes” (Bargainnier, 2004, p. 1). Rubrics are a valuable tool in minimizing the effects of rater errors. There are a variety of options to consider when choosing a rubric. In order to select or create a valid and functional rubric, it must first be consistent across levels (Tierney & Simon, 2004). Additionally, it must fit the needs of the program. The first major consideration is whether to choose a holistic rubric or an analytic rubric.

**Holistic.** A holistic rubric is one that considers all criteria simultaneously and assigns a single score (Brookhart, 1999). These rubrics allow for a broader judgment of the overall product being rated (Moskal, 2000). Many agree that the holistic scoring process takes less time for the rater because they are only required to give one score for the overall product (Brookhart, 1999; Bargainnier, 2004; Schreyer Institute, 2007; Weigle, 2002). This can be a good method of scoring when the characteristics of the product cannot easily be broken into distinct parts, or when the “big picture” of the students’ ability is being sought. A rater must be careful, however, that the student is not docked an exorbitant amount of points for an error that would not carry the same weight in an analytically scored product (Moskal, 2000). The rater must find balance between all the characteristics to create a whole or holistic view of the product.

**Analytic.** According to Moskal, (2000) an analytic rubric is much like a checklist. It allows for the evaluation of separate factors. Each factor receives a separate score based on whether the qualities of each criteria are present (Brookhart, 1999). Because it is like a checklist, there are not always numeric values assigned to each quality or criteria sought. A simple way to understand an analytic rubric is as several scales applied to the same work (Brookhart, 2004). Some argue that analytic scoring is more helpful to students because they can see their strengths and weaknesses more clearly (Brookhart,

1999). But others suggest that analytic scoring is more complex, takes more time, and is therefore more costly (White, 1984).

**Comparison of rubric types.** According to Weigle (2002), some say that the holistic style rubric achieves high inter rater reliability, but in the process loses some validity. She continues by explaining that others state holistic scoring is considered even more valid than the analytic style of scoring because it can more clearly show the real, all-encompassing reaction of a rater to the text. Also argued is that through analytic scoring, the big picture of the overall score is obscured by the attention to its parts, rather than attention to the whole work (Weigle, 2002). Thus the attention to the whole allowed by the holistic score is often a more accurate depiction of the overall ability of the author or examinee.

However, another disadvantage of holistic scoring is that not all raters use the same criteria or standards by which to judge the same writing sample. It is not always simple to interpret a holistic score because one rater may give it a score of 4 based on its rhetorical features while another rater rates it as a 4 for its linguistic features (Weigle, 2002). However, White (1984) suggests that holistic scoring has been shown to have statistically significant reliability. Because raters sometimes allow themselves to be influenced by their background and understanding of different areas of language, training on how to use the rubric can help lead to an increase in inter rater reliability and in rater confidence (Lovorn & Rezaei, 2011).

The ELC and other programs like the Cambridge ESOL Exam and the TOEFL-ibt still use a holistic rubric when rating the writing portions of their tests (Becker, 2010)

although it appears that holistic rubrics are no longer the norm, with testing services and programs like IELTS and the Michigan test turning to analytic scoring (Becker, 2010; Veerappan & Sulaiman, 2012). Although analytic is recognized as more typical and has many advantages, as previously mentioned, the ELC continues to use a holistic rubric. This training workbook is based upon a holistic rubric in part due to the programmatic restraints, and also because holistic rubrics have been proven a faster, more valid way to access the overall reaction of the rater to the writing (Becker, 2010).

Many teachers feel that rubrics increase uniformity, dependability and therefore, agreement and consistency (Spandel, 2006). Additionally, some comment that an untrained rater is less consistent and effective in his or her rating than one who has been trained in rubric use (Rezaei & Lovorn, 2010). Any rubric, no matter how well written, will have ambiguous terms or aspects. Before a rater can consider rating writing, he or she must be familiar with the rubric and its intended meaning, and must therefore have solid exposure and explanation thereof. This exposure will increase the rater's confidence and consistency. Studies have shown that reliability of rated assessments, especially those assessments rated by new teachers, significantly improves with thorough, quality rubric trainings (Lovorn & Rezaei, 2011). However, the rubric training alone is not enough. Harsch and Martin's research (2012) of adapting rubric descriptors also argues the necessity of additional training beyond the rubric for new raters, expressing that trained raters must have the ability to interpret the wording of the rubric in the way test developers intended, which requires more than simply exposure to the rubric. Based on a need for more, the researcher attempts to expose novice raters to the types of writing they will encounter as they rate in order to increase confidence, consistency and agreement.

## **Anchor Samples**

According to some public education departments, anchor papers are a necessary and vital part to a successful training in writing assessment (Arizona Department of Education, 2005; Driscoll, 1996; Nevada Department of Education, 2000; Washington Office of the State Superintendent of Public Instruction, 2001). Anchors, sometimes also referred to as benchmarks, are clear-cut examples of each scoring increment in the rubric. For example, there may be a sample paper that clearly depicts the characteristics of a level 1 in the rubric, another sample paper whose characteristics demonstrate the requirements of a level 2 in the rubric, and so forth. These benchmark samples are important because they “ensure assessment quality and high inter-rater reliability through clearly illustrating the intent of the rubric to the raters” (Osborn Popp, Ryan, & Thompson, 2009, p. 257). White (1984) supports this by discussing the helpfulness of allowing raters to internalize the scale by looking at both the descriptions of the rubric and the examples of the benchmarks. Consequently, anchor papers should be chosen to make the concepts of the rubric more concrete and visible to raters. These samples set the standard and serve as a reference point for the raters, which in turn will increase their agreement as they rate.

Often benchmark scores are chosen by a method called range finding. In this process, the rubric is carefully analyzed and writing samples are chosen which accurately represent the points of the rubric. In some methods, the rubric is created by using a group of sample papers to write concrete reasons for each score in the rubric. However in others, multiple sample papers are measured against the preexisting rubric until papers are chosen which exemplify the established standards of the rubric (Osborn Popp et al.,

2009). One last method is to select benchmarks based on the feedback of experienced raters. If a high percentage of raters agree on the score, it signifies that the writing sample clearly exemplifies the score according to the rubric (White, 1984). Although rated by a variety of raters, the scores assigned are fairly consistent; the various raters tend to rate it at the same level. This signifies that the sample is a clear example of the scoring level in the rubric.

### **Borderline Samples**

In addition to exposure to the anchor samples, new raters need exposure to samples of writing that simply do not fit in the uni-dimensional view of writing created by the use of a holistic rubric. These writing samples are referred to as borderlines because they fall between the score categories, or represent the borderline between two scores (Parke, 2001; Kegley, 1986). A borderline is when an essay does not perfectly conform to the rubric.

Borderline essays present a problem when rating with any rubric and many approaches have been tried to deal with the issue, some less effective than others. Haswell (1998) describes his desire to remedy one process used in a community college because of its over-simplicity. In the program, borderline placement essays were decided by a group vote or decision if they were not obvious to the initial raters. On the other end of the spectrum, Johnson and VanBrackle (2012) describe how the Georgia University Regents' exam rater training gives two borderline samples first to their raters (for the holistic three score rubric), followed by a short description of why it is borderline and then asks them to rate actual essays based on how they compare to the borderline samples. In assessing writing, other universities have created a third category of

borderline to hold students who fit in neither class but can be placed according to the easiest enrollment (Willard-Traub, Decker, Reed, & Johnston, 1999). All of these methods have been used in placing students and show that there is indecision and confusion concerning how to handle borderline writing.

In addition to experienced rater and program confusion on what to do with borderline writing is the confusion that new raters feel when trying to fit borderline essays to a holistic rubric. Remember, for example, the anecdote of Ricardo whose rhetorical aspects to appear to deserve a 4 out of 7 on the given rubric. However, he was weak in grammatical accuracy to the point where he appeared to deserve a 3 out of 7. New raters especially have difficulty deciding which of the two levels is more appropriate for the writing sample's holistic score, especially when they know that placement will result from their rating. A common complaint among new raters is that they cannot give the sample a score of 3.5 or 3-plus, etc. because they feel that their sample reflects parts of both levels in the rubric.

While much research references the problem of borderline writing samples, none could be found to address a solution (i.e. Cho, 2003; Janopoulos, 1986; Kegley, 1986; Penny, Johnson, & Gordon, 2000; Smith, 2000). Based on experiences while still a novice rater and discussions with more experienced raters, the researcher hypothesized that recognizing and understanding how best to score writing samples that seem to be on a borderline comes from experience and exposure to writing. While the benchmark samples alone provide a more concrete example to help new raters, the researcher theorized that new raters also need communication with experienced raters and exposure to borderline writing samples. Exposure to more writing that is on the border of two

rubric scores will allow the new raters more practice. This exposure will improve the confidence and consistency of new raters.

### **Experienced Rater Feedback**

Another tool for assisting new raters in preventing or overcoming personal bias and inconsistency is to train them using the knowledge of more experienced raters. Harsch and Martin (2012) describe one of the purposes of new rater training as “facilitating novice raters to approximate expert raters’ behaviour” (p. 233). Through guidance and feedback, new raters can better understand the human logic and rationale behind the true score, or expected observable score (Jang, 2010). An increase in input from experienced to new raters will help new raters improve their own skills in making logic-based choices. In formal assessments, this input is can be given to raters through written feedback on how a writing sample fits the rubric or does not, as seen in Johnson and VanBrackle’s (2012) description of the Regent’s exam in Georgia.

As mentioned previously, the decision to utilize a holistic rubric can result in difficulty interpreting the scores because not all raters use the same criteria or rationale to select the appropriate score (Weigle, 2002). Therefore, feedback from experienced raters will provide a written rationale assisting new raters in comprehending which characteristics of the writing are most important. When experienced raters refer to the rubric to detail why they selected a particular score, it will provide one more useful resource for new raters to rely upon as they grow accustomed to rating. Similarly, in Harsch and Martin’s study (2012), they found that by conducting training with experienced raters and new raters together they were able to best interpret the rubric descriptors, rewrite them and increase rater reliability.

## Practice Writing Samples

According to Bloom's Taxonomy, there is a difference between knowing information and being capable of applying that information (Anderson & Krathwohl, 2001). In order to assist new raters in the application of their knowledge, the raters must have the opportunity to practice and gain experience. New raters should have the opportunity to blindly rate a variety of writing samples to allow them to practice rating with realistic samples in a low-stakes setting, but also to receive proper feedback following their ratings (White, 1984).

As Knoch, Read and von Randow (2007) clearly state:

In writing assessment programmes it is routine practice to moderate the influence of rater effects by double- or triple-rating of the scripts and, in larger-scale programmes, by using many-faceted Rasch measurement to make statistical adjustments to the ratings on a post-hoc basis. However, the most important way to address these effects is by means of rater training whereby— both initially and as an ongoing activity – raters are (re-)introduced to the assessment criteria and then required to demonstrate the extent to which they have arrived at a common interpretation of what the criteria mean by rating a number of writing samples. (p. 27)

After the new raters have tried to apply their knowledge and use the rubric to score the samples, it would be helpful to them to see a key of the established scores, or what are considered the actual sample score to which novice raters' scores are compared, and then to discuss or receive training on why the samples were rated accordingly. This

allows the new rater to assess their own progress and the administrator to discover what type of rater effect the new rater tends to use.

Despite all the training, discussion, and calibration, there will still be rater error and bias which have an effect on the outcome of the scores, regardless of the raters' level of experience. Based on a study by Vaughan (1991), he suggests even after raters receive training, they will depend on their own individual rating approaches and ideas, especially when rating borderline samples.

### **Rater Error**

No two raters are created equal. All raters approach their tasks with individual bias and varied experiences. The result of these biases manifests itself in what are called rater effects (Eckes, 2011). Many researchers have devoted time to defining rater effects. Scullen, Mount, and Goff (2000) define rater effects as a "broad category of effects [referring] to the systematic variance in performance ratings that is associated in some way with the rater and not with the actual performance of the ratee" (p. 957). Myford and Wolfe (2003) suggest that rater effect occurs when a rater introduces error into rating. The most commonly discussed rater effects are those of leniency/severity, halo, central tendency and restriction of range. For the purposes of this research, a short definition of each type of rater effect seems appropriate.

Leniency and severity are descriptions for raters who have the inclination to rate far above or below the midpoint (Myford & Wolfe, 2003). As suggested by their names, a lenient rater would consistently rate higher than the midpoint, and a severe rater would consistently rate below. Another type, halo effect, describes a rater who has an idea of

what type of person the ratee is: inherently good or inherently inferior. This could be due to the rater recognizing the writing style as coming from a particular student or a particular culture of which the rater has an opinion or bias. Raters who struggle with the halo effect then allow that idea of the person to affect their judgment of the separate qualities of the sample being rated (Myford & Wolfe, 2003). Other raters struggle with central tendency effect. This describes a rater who “overuses the middle category of a rating scale while avoiding the extreme categories” (Myford & Wolfe, 2003, p. 396). Lastly, when raters cluster their scores around any given point on the rating continuum, but not the midpoint, it is labeled as restriction of range effect (Myford & Wolfe, 2003). All these effects are errors typical of raters. As administrators and test proctors, the goal should be to minimize these effects in order to ensure the highest accuracy rating possible, understanding that even experienced raters still make errors.

### **Fair Averages**

In order to minimize the rater effects, various methods can be used. Immediately following rating, there is what is called an observed score. As with the previous example of Ricardo, if two raters rated one essay with respective scores of three and four, the observed average or score would be 3.5. However, a consideration in assigning a fair average, or an average that takes rater error into account, is if the rater who awarded a four was known to be a generous rater. Some methods employed in creating a fair average are Multi-dimensional Item Response Theory (MIRT), Generalizability Theory, and many-facet Rasch Measurement (MFRM). MIRT is defended by some as the best method to utilize because writing is multi-dimensional (Reckase, 2009). The Educational Testing Company (ETS) defends their use of Generalizability Theory because within that

framework, when different assessment designs change the reliability of the scores, they can still be examined. For example, if there are different combinations of number of tasks and raters, the scores can still be assessed reliably (“Using Generalizability Theory,” 2004). MFRM is the method employed at the ELC in keeping with the common trend of language testing investigations which utilize uni-dimensional models like MRFM (McNamara & Knoch, 2012). With MFRM, a series of mathematical equations computes what an examinee would have been rated by an average of all the raters. MFRM scores are averaged and then adjusted for rater error. The common error trends of each rater are factored into the fair average score through MFRM (Eckes, 2011).

Osborn Popp, Ryan and Thompson (2009) defend their use of MFRM as it allows for multiple facets when analyzing the data. This means that student ability is being examined while allowing for rater error and item difficulty in the exam. By factoring in these rater effects, Rasch modeling allows researchers to use the fair average rather than the observed average, and to take into account more factors that could skew the data of the observed score found. Utilizing MFRM makes it possible to overcome some unreliability between raters, thus allowing for more accuracy and consistency in scoring.

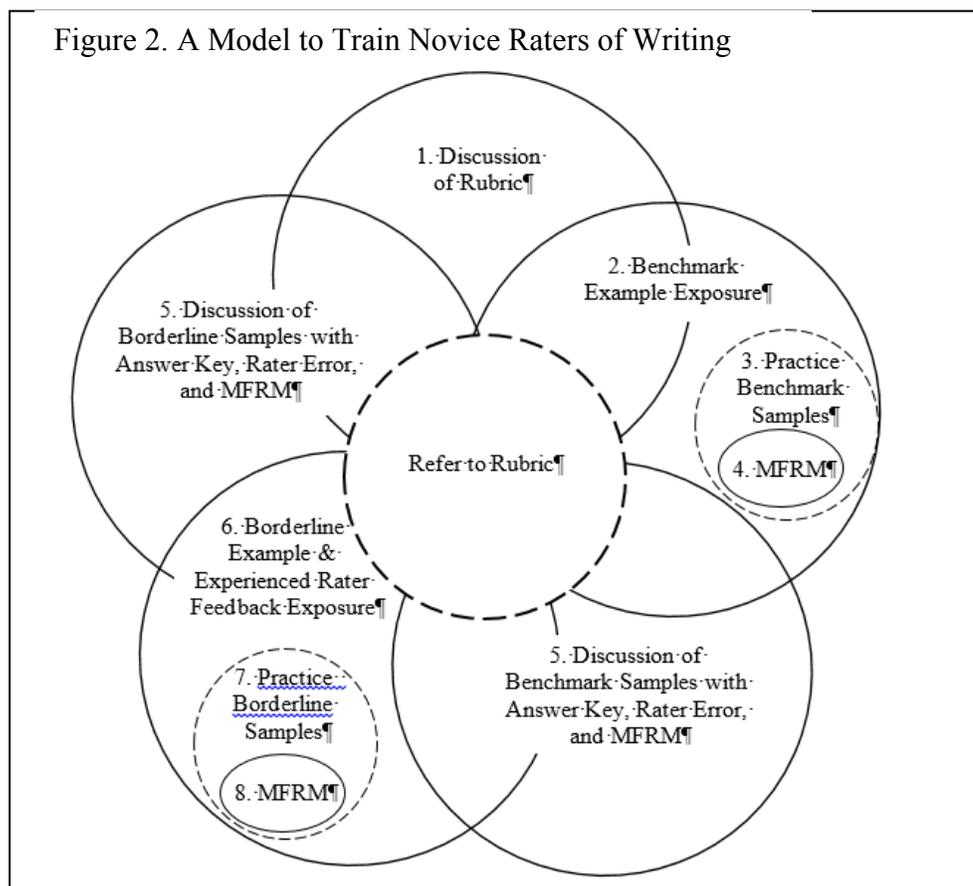
### **Summary**

Given the research presented here, new raters need much more than access to a rubric. Until they have gained adequate experience, new raters need reliable training that will help them rate writing more confidently and consistently. Through application of the principles presented here, a model, described more in the following chapter and shown in Figure 2, can be created to assist in the training of novice raters. With the proper application of the model, new raters of writing will become more familiar with the rubric,

will have input from experienced raters, and a better understanding of how to rate borderline samples of writing. This will improve the self-consistency and agreement of the new raters with established scores, combat the issues of rater error, and will ease the anxiety and improve the confidence of new raters, as the aims and questions of this thesis suggest. Additionally, the model can be useful on a larger scale. Any program that administers assessments requiring multiple raters will be able to look to this model as a foundation and guide to train their own new raters, assisting in the improvement of assessment and consistency on a more global level.

### Chapter 3: Methodology

In order to expose novice raters to the principles and strategies outlined in Chapter 2, an instrument for training new raters was created. This instrument comes in the form of a training workbook (see Appendices A & B), designed to present borderline samples of writing with experienced rater rational. The workbook follows a model, as shown in Figure 2, which allows for access to the materials in a sequence with the rubric presented first. Additionally, Figure 2 shows how the raters will have continual access to the rubric, benchmark examples, borderline examples, and experienced rater feedback once they have been exposed to each, as each section of the training model is overlapping. This model is an outline to be followed in the creation of a training workbook for novice raters of writing.



The graphic in Figure 2 represents the model by which the training workbook should be utilized. First, this chapter will outline the process by which the instrument was created, data will be presented to determine the extent to which the newly created materials improve consistency, agreement and confidence in novice raters.

### **Instrument Development**

The work for this thesis was carried out at the English Language Center, the Intensive English Program at Brigham Young University in Provo, Utah. The program consisted of students divided into two tracks: a Foundations track for beginning level students, and an Academic track for more advanced level students. Within each track there were three levels plus two preparatory levels.

At the start of their first semester, students took a placement test to determine their level of proficiency. Likewise, at the close of each semester, all students were required to take Level Achievement Tests (LATs) as part of their qualifications to advance to a higher proficiency level. As part of their contract, teachers in the program were required to participate in the rating of the LATs each semester. Teachers were asked to assist in rating LATs in one of three productive areas: 30-minute Timed Writing, Portfolio Writing, or Speaking. The training for this thesis focuses on the 30-minute Timed Writing LAT. The writing examples, provided to raters with the established score to act as a guide in their rating, and samples, provided to raters without the established score to allow for practice, are actual student responses to the 30-minute Timed Writing LAT over the course of several semesters.

An existing training model was already in place at the ELC. The following describes the various parts of the existing training workbook which were retained, as well as the additions carried out to improve the model.

**Rubric.** The existing training workbook at the English Language Center contains instructions for use of the rubric, followed by a copy of the rubric (Appendix A). The program uses a holistic rubric. It is based on an eight-point scale, and each numeric score corresponds to one of eight class levels at the ELC (i.e. A score of 0 implies that the student is ready to begin the class level at the ELC titled “Foundations Preparation,” a score of 1 implies that a student is ready to begin “Foundations A,” etc.).

Each numeric level is segregated by varying levels of “Text Type,” “Content,” and “Accuracy,” which lends an analytic aspect to the rubric. In the rubric, “Text Type” is described as the length and organization, with more specific guidance provided in each of the rubrics’ eight levels. “Content” refers to the writer’s functional ability with the language, and to their vocabulary. Lastly, the “Accuracy” section refers to the meaning provided, and grammatical complexity (see Appendix A).

The instructions for proper use of the rubric specify that the rater should begin judging the sample by “Text Type,” then moving on to “Content” and then “Accuracy.” If the rater is still unsure, he or she should “consider semantics, pragmatics, punctuation, spelling or concision to inform [his or her] choice” (*Timed Writing Rater Training*, 2011). This instruction helps to combat the weakness of holistic scoring when each rater puts varied levels of weight on the different characteristics of the whole.

However, in the existing instructions, they also suggest that if a student's skills are spread over different levels the rater should use their "best judgment, bearing in mind that sometimes a student's demonstrated strength in one area can compensate for weakness in another" (*Timed Writing Rater Training*, 2011). This instruction embodies the very issue being addressed through this research, trusting the judgment of various raters to be consistent when rating samples which fall between the levels of the rubric.

**Benchmark examples.** Benchmark examples in the existing training workbook were chosen by an expert rater. He viewed a Facet Output, the results provided by software used to apply many-facet Rasch measurement, and selected writing samples with a solid score, for example a solid 3 or a solid 4 as opposed to a 3.5. These essays he selected as benchmarks. He did not, however, analyze the writing of the essays to see what characteristics they had which could explain why they received the established score, relying strictly upon the statistics (T. Cox, personal communication, 2012).

In selecting writing examples for the updated training, fit statistics were an important variable. However, if a program does not have access to the Facets program, benchmarks and borderlines can be selected based on feedback from multiple raters, as Haswell (1998) described. In his study, he mentions gathering experienced raters to discuss samples of writing and decide if they are clear examples of the rubric (benchmarks) or not (borderlines). While a more primitive method, feedback from multiple experienced raters is another option for selecting borderline and benchmark examples. If this method were pursued, specific care would need to be taken in selecting raters with extensive experience in rating, in utilizing the given rubric, and in

understanding the levels under scrutiny. Because of this and other research presented in the previous chapter, it is recommended to utilize the fit statistics whenever possible.

An infit statistic shows how well the writer did in responding to an item at his or her level (Linacre, 2007). If the student did not perform as expected on a test item at his or her level, either scoring much higher or much lower than the expected score, the infit statistic would be high. Having an infit above zero shows that the raters are human. If raters consistently had an infit of zero, they would be acting as rating machines. Instead, an infit slightly above zero shows that their reactions to essays are not always perfectly consistent. In selecting the benchmark samples, the researcher typically chose infit statistics between 0.5 and 1.5 as shown in Table 1. If there were no samples with an infit statistic which followed these guidelines, the researcher selected an example where the infit statistic was lower, not higher. Although the writer over fits the model with a low infit statistic (below 0.5), it shows a better example of level-appropriate writing than when infit statistic is above 1.5 (Linacre, 2007).

Table 1

Benchmark Selection Data						
ID Number	Number of Raters	Observed Average	Fair Average	MSE	Infit Statistic	Outfit Statistic
12792	2	7.0	6.66	1.98		
13254	2	7.0	6.97	2.00		
13554	14	3.3	3.26	0.53	0.77	0.70
17398	8	3.1	3.05	0.73	0.56	0.48
20367	2	1.0	1.30	2.10		
22304	2	2.0	2.01	1.35	0.79	0.59
25722	4	4.0	3.93	0.82	0.50	0.49
27576	2	6.0	6.37	1.19	0.58	0.58
29541	2	7.0	6.88	2.07		
36339	4	4.0	3.81	0.82	0.76	0.75
39493	5	5.2	5.31	0.76	0.55	0.56
39890	2	1.0	1.00	4.12	0.03	0.03
47935	2	6.0	6.37	1.19	0.58	0.58
49343	7	4.1	4.10	0.79	0.84	0.76
50685	7	5.1	5.08	0.68	0.88	0.89
50875	4	5.0	5.03	0.75	0.50	0.50
54232	4	4.0	3.93	0.82	0.96	1.05
64941	2	1.0	0.99	9.40	0.00	0.00
72406	5	5.0	5.04	0.81	0.45	0.47
72815	4	2.0	1.97	1.27	0.08	0.07
73415	3	6.0	5.70	1.02	0.59	0.59
73871	3	3.0	3.00	1.56	0.21	0.17
76495	5	3.0	3.01	1.26	0.55	0.30
99846	4	2.3	2.42	1.12	0.95	0.80

The researcher selected 24 benchmark essays. Two were selected to act as benchmark examples for each of the seven levels of the rubric, and the remaining 10 served as practice samples. When selecting benchmarks, the researcher referred to the number of raters and the observed and fair average scores. Updated benchmarks were selected if 3 or more raters had rated the essay and concurred on its score, so the observed average was a whole number (2.0, 4.0, 5.0, etc.) as opposed to an average with a decimal point (2.4, 4.7, 5.8, etc.) as shown in Table 1. The fair average was also considered, especially if an essay was selected which had an observed average with a decimal point.

In the workbook, the benchmark examples are labeled with their actual score and ordered from lowest to highest. For example, two benchmark examples for Level One are first, then two benchmark examples for Level Two, etc. This allows raters to view the benchmarks while looking at the rubric and to see a concrete example of the characteristics described in the rubric.

**Borderline examples.** Recognizing and understanding how best to score writing samples that seem to be on a borderline comes from experience and exposure to writing. While the existing ELC training workbook provided benchmark writing samples, new raters need a combination of borderline writing samples and the written rationale provided by previous raters. Exposure to more writing that is on the border of two rubric scores will allow the new raters more practice. This exposure will improve the confidence and consistency of new raters.

As mentioned, the fit statistics are important when selecting writing samples. When selecting borderline examples, the infit and outfit statistics should be higher than those of the benchmark examples. If the infit statistic is high, it implies that the writer did not perform as expected on the item. This means that the writing sample is not a clear illustration of one score on the rubric. The raters were unexpected to classify the essay if the infit and outfit statistics are high. The outfit statistic shows how well the writer responded to an item that was not at his or her difficulty level. The item could have been far too easy for his or her level or much above the level. If the outfit is high, it implies that the writer either did very poorly on an easy item or performed above level on a difficult item (Linacre, 2007). Extreme fit statistics imply that the writing sample was not a clear example of the level at which it was rated.

More heavily considered in selecting the borderline examples, however, was the Model Standard Error (MSE) statistic. A higher MSE signifies a larger gap between the scores assigned by the raters (Linacre, 2007). A larger gap between raters implies disagreement when rating. Disagreement shows that the sample does not clearly follow the patterns outlined in the rubric, which means that it is not a clear example of a given level in the rubric and is therefore more likely a good example of a borderline score.

In selecting the 18 borderline essays, the researcher looked for writing samples which had been rated by 2 or more raters, had a decimal point in the observed average (3.5 instead of 3.0), had high fit statistics (over 1.0), and a high Model Standard Error statistic (typically over 1.0) as shown in Table 2. However, insufficient writing samples following these characteristics were available. Therefore, the researcher selected one borderline sample from the ELC's existing training booklet. Statistics were not available for any writing samples in the existing training booklet, so this sample was selected based on feedback from raters that it was indeed a borderline sample. Therefore, for one of the 18 borderline examples, the researcher selected the sample based on other guidelines.

Table 2  
Borderline Selection Data

ID Number	Number of Raters	Observed Average	Fair Average	MSE	Infit Statistic	Oufit Statistic
10726	4	3.5	3.62	0.92	0.54	0.51
17865	5	3.6	3.60	1.15	1.11	0.82
20487	2	6.5	6.14	1.32	1.52	1.55
29066	2	1.5	1.47	1.42	1.29	1.29
29615	2	3.5	3.58	3.16	0.01	0.01
31220	5	4.8	4.76	1.04	0.43	0.33
36092	6	1.7	1.82	0.85	0.96	0.92
41393	3	2.3	2.64	1.34	0.38	0.33
48419	3	5.7	5.93	2.18	0.01	0.01
57252	4	2.3	2.28	1.12	1.13	0.97
57753	4	4.3	4.45	1.02	1.83	1.82
61718	7	4.4				
62695	2	3.5	3.67	1.68	3.78	3.88
68121	3	4.7	4.73	1.14	1.52	2.25
71076	5	4.4	4.28	0.99	0.55	0.48
92112	3	5.3	5.52	1.07	1.54	1.56
95177	3	3.3	3.17	1.48	0.27	0.24
99840	5	3.4	3.31	1.09	1.06	0.83

Of the 18 borderline essays, eight served as borderline examples. These examples were provided to the raters with the established scores indicated to act as models. The eight included one example per scoring level of the rubric (an unclear 1, an unclear 2, etc.) and a second example for the borderline scores which have proven most difficult for raters: those in the 3-5 range. The remaining 10 borderline essays served as borderline practice samples, and therefore did not have the established score listed.

In the workbook, the borderline examples are labeled with their established scores, and are ordered from lowest to highest. This allows new raters to view examples of writing that do not perfectly fit the model of the rubric and to see how “fence-sitter” examples of writing are rated. In the adjacent page, feedback and comments from

experienced raters are included to describe why the borderline could be rated with one of two scores.

**Practice writing samples.** In the ELC’s original training packet, the writing samples provided to the novice raters for practice were selected from essays which had been rated many times and had been run in a facets analysis to ensure their established score.

In the updated training workbook, the samples provided for practice are a randomized mix of both benchmark and borderline samples. They were selected by the same criteria as the benchmark and borderline example essays used at the beginning of the training workbook, and are made up of the 10 extra benchmark essays and 10 extra borderline essays mentioned in the previous sections. The only difference between the practice writing samples and the benchmark and borderline examples is that the final/established score and the experienced rater rationale is not included in the practice samples. Instead, the new raters are to rate them without guidance. The purpose of these writing samples is for the new raters to practice rating on their own, without being influenced by an established score or by experienced rater feedback. From this point forward, the words “samples” and “examples” will not be used interchangeably, for sample refers to un-marked practice essays and example refers to reference essays with the established score listed.

**Content experts.** The teachers at the ELC could be divided into three groups. There were graduate students who taught part-time in addition to their studies, non-student community teachers who were hired part-time, and full-time ELC staff. Teacher

levels of experience and exposure to different skill areas and levels vary. Eight teachers were selected to assist in the creation of the training workbook by supplying experienced rater feedback. All were experienced teachers at the English Language Center, having no less than three semesters experience teaching at the ELC and rating LATs, and an average of six semesters of experience. The researcher attempted to utilize teachers of all three teacher populations by selecting three part-time graduate students, two part-time community teachers, and three full-time ELC teachers.

**Selecting experienced rater rationale.** Following the selection of the writing examples and samples, comments from experienced raters were collected. The content experts were provided with the writing samples, asked to read them, then to describe in their own words, as well as words, phrases, and examples from the rubric to support their answers in three sentences or less why the essay could have been rated at either its established score or the score which borders the established score in the rubric. For example, if the established score of a borderline essay was a level three, they were asked to defend why it was a three, but they were also asked to describe why it could have been mistaken for a level four.

After the experienced rater feedback was collected, the researcher compiled the comments of the experienced raters, edited them for clarity purposes by eliminating any abbreviations and formatting them as complete sentences when they were fragments or bullet points. Then the researcher positioned them adjacent to the borderline writing example as a resource to new raters. Not every content expert chose to respond to every essay. For some essays there are as few as three experienced rater responses, while others have comments from all eight raters. When the novice raters read the borderline

examples, they can refer to both the rubric and to the experienced rater feedback as a guide to understand what experienced raters saw as most important when rating (Appendices A & B).

### **Research Design**

**Subjects.** Due to teacher turn-over at the ELC, there was a high volume of new raters each semester, necessitating recurrent training. One group of new teachers each year came from the TESOL graduate program at BYU. These first year graduate students spent a year preparing to teach at the ELC, then for the remainder of their graduate studies were usually hired as part-time teachers. During their first year of courses, they took the course, Linguistics 611: “Methods and Materials Application” in which they learned about rubrics and practiced rating. For many, this was their first exposure to rating writing. Therefore, they were an ideal group in which to pilot a training for new raters of writing.

From January to April of 2013, there were 11 first-year graduate students enrolled in Linguistics 611. The course was divided into two sections, with six students in one section, pilot Group A, and five students in the second section, pilot Group B. In Group A, one student opted not to participate in the pre-treatment rating, which invalidated his post-treatment ratings. His results and data were not included in the analysis. Therefore, there were 5 participants in each pilot group. Due to curriculum scheduling, Group A received the pilot, then three weeks later, Group B went through the same test.

**Crossover design.** A crossover design can yield a more efficient comparison than a parallel design because fewer subjects are required in the crossover design to attain the

same level of statistical evidence as in a parallel design (Vonesh & Chinchilli, 1997; Chow & Liu, 2004). In striving to validate the data with so few participants, a crossover design was utilized. Group A rated the first ten essays, listed on Handout A, then received the treatment, after which they rated the second ten essays, listed on Handout B (see Appendix C). Three weeks later, Group B went through the pilot. They rated the second group of ten essays first, listed on Handout B, then they received the treatment, after which they rated the first ten essays: listed on Handout A. This crossover design was used in case one group of essays (either Handout A or Handout B) was more difficult or there were errors or issues within one of the groups of essays (Chow & Liu, 2004).

**Pre-treatment.** The purpose of the pilot was to learn if exposure to borderline samples of writing and experienced rater feedback as developed in the training packet would assist new raters in more accurately rating borderline samples of writing. In order to do so, the pre-treatment consisted of ten practice essays: five benchmark essays and five borderline essays, which were not labeled with their established score. The participants were provided the rubric, instructions for the use of the rubric, and benchmark examples (labeled with their established score) of each level of writing described within the rubric (See Appendix A). They were instructed to read through the benchmark examples while looking at the rubric to familiarize themselves with the levels of the rubric, after which they were to rate the 10 writing samples provided on either Handout A or Handout B, dependent upon which group they were in (See Appendix C). They were given this assignment in class on a Monday, and had two days to accomplish this assignment.

**Treatment.** After completing the pre-treatment, students came to class the following Wednesday for the treatment. In the class the researcher led a discussion of common rater error, many-facet Rasch measurement, the established scores of the essays they initially rated, and an explanation of borderline writing as shown in a filled copy of either Handout A or Handout B (See Appendix C). As a part of the treatment, students were then given the remainder of the training workbook: examples of borderline writing (labeled with established scores) with corresponding comments from experienced raters (discussed below; See Appendix A).

**Post-treatment.** Students were given instructions to read the borderline writing examples and experienced rater comments while reviewing the rubric, then to rate ten more writing samples included in either Handout A or Handout B (See Appendix C). This was done to analyze change in the new raters rating abilities.

### **Structure of Analysis**

**Qualitative data collection.** Because the purpose of this thesis is to create the updated training workbook by including exposure to borderline samples of writing, there was insufficient time to carry out a full empirical analysis to collect quantitative data. In order to gather more data, the researcher relied on qualitative responses from those involved in the pilot of the updated rater training workbook. The responses from the survey will answer Question 1: To what extent does borderline training with experienced rater rationales increase the confidence of the novice raters? The responses will also address other issues of usefulness in relation to the training workbook.

In order to gather qualitative data, the researcher created a survey to collect the opinions and reactions of the new raters to the training (see Appendix D). The researcher used online data collection software, Qualtrics. The researcher asked the participants questions to assess their perceived progress throughout the training and rating process, and to gauge the usefulness and structural clarity of the training workbook.

**Quantitative data collection.** After piloting, the data collected was analyzed using Pearson product-moment correlation, Cohen's Kappa, the fixed effects analysis of variance (ANOVA) and lastly, MFRM. These analyses will answer Question 2: To what extent does borderline training with experienced rater rationales result in more consistent agreement between novice raters and the standards? As well as Question 3: How consistent are the raters after exposure to the rating process?

The Pearson will disclose if there is correlation between the ordering of the established scores and the novice raters' scores, and the strength of the relationship between those two variables (Brown & Hudson, 2002; Eckes, 2011). The researcher hypothesized that post-treatment, the novice raters' Pearson correlation would be higher. Cohen's Kappa is a statistical measure of inter-rater reliability which shows agreement between the ratings given by new raters and the established ratings ("Interrater Reliability," 2013). The researcher postulated that the agreement of the novice raters' scores and the established scores would increase. The fixed effects ANOVA shows the mean of the difference between the established scores and the novice raters' scores. This statistic is used to analyze data in a crossover design. If the novice raters give the same ratings as the established score, the difference would be zero. The null hypothesis was that there is no difference between pre-treatment and post-treatment in rating the

borderlines. It is hypothesized that the mean difference will be closer to zero after treatment as that would indicate the novice raters are improving. Lastly this chapter will briefly look at a Facet analysis using many-facet Rasch measurement to answer Question 3, which the researcher expected to show high intra-rater reliability, or self-consistency in rating.

## Chapter 4: Results

The main purpose of this thesis is to create a training workbook which can be used to improve the confidence of novice raters of writing when rating borderline samples of writing (see Appendices A & B). In addition to the creation of the workbook, ten novice raters were observed using the training workbook to assess the workbook's effectiveness in developing more accurate novice raters. In addition to the results of the data collected through the pilot study, a brief description of a public presentation of this model will be outlined. This chapter presents the results of the data collected in an effort to answer the following questions:

1. To what extent does borderline training with experienced rater rationales increase the confidence of the novice raters?
2. To what extent does borderline training with experienced rater rationales result in more consistent agreement between novice raters and the standards?
3. Following exposure to borderline training with experienced rater rationales, do novice raters have sufficient consistency, or intra-rater reliability, to rate high-stakes tests?

### **Question 1 and Survey Responses**

This first question examining the confidence levels of novice raters was answered through the collection of qualitative data in a survey (see Appendix D). The survey also collected data regarding the difficulty level of rating and the usefulness of various aspects of the training workbook. This additional information is important in discovering if the creation of this training workbook is effective. This section will answer the first research

question, but will also outline other responses to the survey in order to gauge the usefulness of the training workbook.

**Confidence.** Of the nine raters who responded to the survey, six stated that they felt more confident in their abilities to rate timed writing after the practice samples and the treatment. Two stated they did not feel a change in confidence, and one stated she felt less confident. Of the three who did not feel more confident, two cited their lack of agreement with the established score as the reason for lowered/unchanged confidence. The third rater felt uncomfortable reporting a change in confidence without seeing the ratings of her peers. The six who reported feeling more confident referenced increased familiarity through practice and exposure to the rubric.

**Difficulty.** Raters were then asked which writing samples were more difficult to rate; those provided pre-treatment or those provided post-treatment. The researcher hypothesized that pre-treatment samples would be more difficult due to lack of exposure to the rating process, and therefore lack of confidence in rating. However, seven of nine raters reported the post-treatment writing samples as more difficult to rate. Their explanations of why touched on lack of agreement with established scores, more elements to be aware of, and lack of time to complete the ratings. The two raters who felt the pre-treatment samples were more difficult cited the reasons expected by the researcher in the hypothesis: lack of previous exposure, lack of confidence, and they were unsure how to categorize “in-between scores.”

In spite of reporting higher difficulty when rating post-treatment samples, seven of eight raters reported feeling more prepared to rate more essays following the treatment.

As rationale for increased feelings of preparedness, the seven raters cited comparisons of their ratings to the established scores and to the ratings of their peers, more confidence in use of the rubric and four of the seven referenced increased confidence stemming from the understanding of compensation for their errors through MFRM. The eighth rater who responded to the question reported feeling unchanged in preparedness.

**Usefulness.** The raters were then asked to answer and provide rationale for whether the various segments of the training workbook were useful and/or helpful in rating. The survey asked about the three main parts of the training workbook: the benchmark examples, the borderline examples, and experienced rater feedback. Table 3 displays the results of these questions. As can be seen, a majority of the raters who responded stated that each of the segments was helpful and/or useful all or some of the time: benchmark examples, nine of nine, or 100%, borderline examples, five of seven, or 71%, and experienced rater feedback five of six, or 83%.

While the majority of the novice raters were positive about the usefulness of all three aforementioned segments of the training workbook, the borderline examples were not as well-received as the other segments, with two of the seven responding raters (29%) stating that the borderline examples were not helpful or useful.

Table 3  
Survey Responses of Usefulness/Helpfulness of Training Workbook Segments

Segment Type	Useful/Helpful (No. of Responses)	Summarized Responses of Why/Why Not (number of similar responses)
Benchmark Examples	Yes (7)	Helpful in gauging what each score looks like in the rubric and in guiding how to rate the samples (4).
	Sometimes (2)	Sometimes they were useful, other times they created more confusion if I disagreed with the score (2).
Borderline Examples	Yes (4)	Helpful in overcoming the difficulty of deciding between two particular scores (1). Helpful in determining how the writing matched elements of the rubric (1).
	Sometimes (1)	Sometimes they were useful, other times they made it more confusing (1).
	No (2)	Caused more confusion, perhaps caused by lack of familiarity with the borderline examples (1). Can be interpreted in different ways (1).
Experienced Rater Feedback	Yes (4)	Provided rationale for established ratings (3). Improved my judgment (1) Provided clarity to the rubric (1) They were a little helpful. I tend to hold my same opinions as I rate (1).
	Sometimes (1)	It was useful at times, but sometimes I still did not understand (1)
	No (1)	I do better comparing two essays than comparing comments to an essay, so the benchmark samples were the most useful. The comments did not help me (1).

**Reliance upon the workbook segments.** Next in the survey, raters were asked to rank how much they relied on each segment or skill during the rating process. In other words, they were to choose the workbook segment they relied upon most, and rank it first. Then whichever segment they relied upon second most, they ranked as second. They were to rank the rubric, benchmark examples, borderline examples, experienced rater feedback, as well as their own intuition. The results are shown in Table 4. Notably, seven students stated that they relied most heavily upon the rubric when rating, and all nine

ranked it as what they relied on either first or second. This implies that they understood the importance of reliance upon the rubric, and that they were willing to adhere to that instruction. It appears that of the segments and skills, borderline examples and experienced rater feedback were relied on least, with eight students ranking borderlines as either fourth or fifth, and seven students ranking experienced rater feedback as either fourth or fifth. All nine ranked experienced rater feedback as one of the last three things upon which they relied.

Table 4  
Ranking of Rater Reliance upon Training Workbook Segments and Other Skills

Ranking	First	Second	Third	Fourth	Fifth
Rubric	7	2			
Benchmark Examples	1	3	3	2	
Borderline Examples	1			5	3
Experienced Rater Feedback			2	2	5
Your Own Intuition		4	4		1
Total Rater Responses	9	9	9	9	9

**Structural preferences for experienced rater rationale.** Lastly, raters were asked about the structure of the experienced rater feedback. In the Handouts, the experienced rater feedback was condensed and summarized by the researcher (see Appendix C). In the training workbook, experienced rater feedback was individually labeled, and had not been summarized or condensed with other raters' comments based on similarities (see Appendices A & B). The new raters were asked in the survey which structure they preferred. Only four responded. One revealed no preference, one preferred the separated comments in order to view different ideas, and two stated their preference

of the condensed comments because it was easier to follow. This is not a large enough sample size to determine which structure is more effective and helpful to novice raters.

### **TESOL Presentation Feedback**

In addition to the qualitative data collected from the survey, some feedback was provided when the model by which the training workbook was created was presented in the Graduate Student Forum (GSF) in Dallas, Texas, at the International TESOL Convention in March, 2013. Approximately 18 members of the TESOL GSF were in attendance, and the title of the presentation was, “Rescuing New Writing Assessment Raters from Inexperience.” Not much time was allotted for questions following the presentation of the outline, but one attendee asked how long it would take for new raters to become “experienced” when utilizing this training method. A longitudinal study would be necessary to assess this information. Two attendees who addressed the researcher/presenter following the presentation expressed the frustration of dealing with borderline samples of writing as novice raters, and showed interest in future research on the topic.

### **Question 2**

The second question refers to agreement between novice raters and established scores. In order to analyze agreement, both correlation and agreement and the difference of the means were reviewed through Pearson product-moment correlation ( $r$ ), Cohen’s Kappa, and a Fixed Effects ANOVA.

**Pearson product-moment correlation.** The first step in data analysis was to cross-tabulate the pre- and post-treatment novice ratings with the established ratings of

borderline writing samples. After cross-tabulation, Pearson's product-moment correlation ( $r$ ) was 0.817 pre-treatment. This suggests a very strong correlation in the ordering between the novice raters' scores and the established scores. In other words, regardless of whether the numeric score assigned actually matched the numeric score established, the novice raters were ranking or ordering the essays similarly to the established rating. For example, they assigned a lower score to a lower-level essay and a higher score to a higher-level essay. Correlation remained statistically unchanged post-treatment, where ( $r = 0.809$ ).

**Cohen's Kappa.** After cross-tabulating the pre-treatment novice ratings of borderline writing samples with the established ratings, the Cohen's Kappa had a value of 0.330 agreement. This is a fair agreement ("Interrater Reliability," 2013). The hypothesis posited that the Cohen's Kappa would increase post treatment, which would imply that raters were better able to match the established score when rating, but post-treatment agreement was 0.235. While this is a decrease, it is still categorized as fair agreement, which shows that the decrease is not statistically significant.

**Means (fixed effects ANOVA).** The dependent variable was the difference between the established rating and the novice ratings. The researcher hypothesized that the difference would be larger pre-treatment and diminish post-treatment. However, none of the means were significant. As shown in Table 5, the pre-treatment scoring of the benchmarks resulted in a mean of the differences between the established ratings and the novice ratings of 0.14, and post treatment the mean was 0.06. The mean of the difference did decrease in rating the benchmarks, as hypothesized, showing that the novice raters got closer to the established scores post-treatment, but the change was not statistically

significant. Additionally, in rating the borderline samples, the pre-treatment difference was 0.10, and the post-treatment mean was 0.18, showing an increase in the mean, however statistically insignificant. Therefore, an analysis of the means of the difference between the established scores and the novices' scores does not suggest that the treatment caused an improvement in the novice raters' proximity to the established scores.

Table 5

Means of Difference Between Established and Novice Rating				
Treatment	Essay Type	Mean	SD	N
Pre-Treatment	Benchmark	-0.14	0.75620	50
	Borderline	0.10	0.93131	50
	Total	-0.02	0.85257	100
Post-Treatment	Benchmark	0.06	0.91272	50
	Borderline	0.18	0.96235	50
	Total	0.12	0.93506	100
Total	Benchmark	-0.04	0.83991	100
	Borderline	0.14	0.94302	100
	Total	0.05	0.89527	200

### Question 3

In actual practice, a software called Facets, used to apply Rasch modeling, can be used to model out rater effects such as severity and leniency (Linacre, 2007). However in answering the third question, the main goal was to observe the level of intra-rater reliability following exposure to the training workbook. Analysis of the rater measurement was collected in Facets to show intra-rater reliability of the novice raters.

As shown in Table 6, the infit scores for all raters but one, Rater 11, are between the appropriate 0.5-1.5 range. This suggests that following the training, all raters were

sufficiently consistent to be allowed to rate high-stakes tests. Rater 11 was slightly above what was established as an appropriate level for consistency. Rater 11 had a 1.82 infit and 1.62 outfit, which implies that more intervention is necessary with her before she be allowed to rate high-stakes tests. Otherwise, the novice raters show appropriate levels of consistency following exposure to the training workbook.

Table 6  
Rater Measurement Report

Total Score	Total Count	Measure	Model S.E	Infit MnSq	ZStd	Outfit MnSq	ZStd	Rater Number
71	20	-0.12	0.37	0.6	-1.3	0.59	-1.3	1
82	20	1.34	0.36	0.77	-0.7	0.74	-0.8	3
68	20	-0.55	0.38	1.21	0.7	1.3	0.9	4
78	20	0.82	0.36	0.76	-0.7	0.71	-0.9	5
75	20	0.42	0.36	0.67	-1	0.62	-1.2	7
69	20	-0.39	0.37	1.1	0.4	0.97	0	8
72	20	0.02	0.37	0.51	-1.7	0.61	-1.2	9
68	20	-0.53	0.38	1.36	1.1	1.41	1.1	10
74	20	0.29	0.37	1.82	2.1	1.62	1.7	11
63	20	-1.3	0.4	0.71	-0.8	0.56	-1.3	12
72	20	0	0.37	0.95	-0.2	0.91	-0.3	Mean (Count: 10)
5.2	0	0.72	0.01	0.39	1.2	0.37	1.1	S.D. (Population)
5.5	0	0.76	0.01	0.41	1.3	0.39	1.2	S.D. (Sample)

## Summary

The data collected to answer the research questions of the thesis included qualitative survey data, responses from those in attendance in a formal presentation of the model, and quantitative pre- and post-treatment intra-rater reliability data. The next chapter will discuss the implications of these results and provide recommendations for future program improvements and research.

## Chapter 5: Discussion

The purpose of this thesis is to create a training workbook which could be used to train new raters of writing by including rating rationale from experienced raters and exposure to borderline samples of writing. Infit and MSE statistics were used in the selection of the essay examples in the training. Experienced raters were carefully selected and questioned regarding their rationale for each score they assigned while training. The updated training workbook now serves as the training for calibration meetings each semester when raters are preparing to rate Timed Writing LATs at the ELC. Now that it has been created using the established method, it can be further tested for effectiveness, replicated, changed for application to training raters in other skills, or utilized in other programs.

Following creation of the workbook, data was collected to understand whether or not the training workbook was useful or effective in increasing novice rater confidence and consistency as well as improving agreement between novice raters and their more experienced counterparts.

### **Question 1**

The first question examined the effect of borderline writing samples and experienced rater rationales on the confidence of novice raters. Results collected from a survey provide answers to this question, but the survey also provided other valuable data to be discussed here (see Appendix D).

**Confidence.** Based on the results, the majority of the novice raters reported feeling more confident in their abilities to rate timed writing after exposure to the practice samples and treatment. Those raters who reported feeling less confident post-treatment

may have done so for several reasons. The less-confident raters mainly cited disagreement with the established score as the source of their unchanged or decreased confidence. It is possible that these raters had been overloaded with too much information in a short period of time. During a period of only three days, the raters learned about rating for the first time, learned about rubrics and benchmark samples of writing, were asked to rate essays, learned about rater errors and MFRM, then reviewed borderline samples of writing and rater rationales, and were finally asked to rate even more essays. In all this, it is possible that their exposure to the rating process overwhelmed them and lowered their agreement due to overload.

**Difficulty.** When asked which essays were more difficult to rate, those provided pre-treatment or those provided post-treatment, the new raters largely and surprisingly reported the post-treatment essays to be more difficult. Similar to the evaluation of the data regarding their confidence, it is possible that the new raters found the second group of essays to be more difficult because they had received more knowledge and understanding.

Following this line of thinking, it might be supposed that the raters were initially naïve to their own lack of expertise when rating. Then, at treatment, they were exposed to the established scores of those first essays. They learned about rater effects in general, and were then able to surmise their own tendencies for error. It is possible that exposure to the treatment heightened their awareness of the many aspects of rating, which in turn made them more aware of all the ways in which they could err. This greater awareness could be a factor in causing the raters to sense a higher level of difficulty when rating the post-treatment essays.

**Usefulness and reliance upon the workbook segments.** The novice raters were asked to describe the level of helpfulness and usefulness of the benchmark and the borderline examples of writing and the experienced rater rationale. They were also asked to rank the level of their reliance upon various segments of the training workbook. In doing so, all agreed that the benchmark examples were useful and helpful, and almost all ranked them higher than borderline examples in terms of what they relied upon. The majority of raters described the borderline examples and experienced rater feedback as helpful, but overall these aspects of the workbook were less well received than the benchmark examples. While this could indeed imply that they were less helpful than the benchmark samples of writing, this could also be explained by the lack of time novice raters had the borderline examples and experienced rater feedback, or that while important, the borderline examples have a lower priority than the benchmark examples of writing. One new rater stated that she was more comfortable using the benchmark examples because she had them longer, and therefore found herself relying on them more than the newer aspect of borderline examples. With that perspective in mind, other raters may have felt rushed to complete the post-treatment ratings and therefore delved less seriously into the borderline examples than they had the benchmark examples pre-treatment.

## **Question 2**

In answering the second research question of consistent agreement between novice raters and the established scores, high correlation was shown, but agreement was statistically unchanged. Cohen's Kappa and the fixed effect ANOVA displayed that the novice raters' ability to rate a practice sample and match the rating of the established

score remained unaffected by the exposure they received to borderline samples and experienced rater rationale. This lack of agreement could be due to weakness of the experimentation. Due to time constraints and the limitation of how many students were enrolled in Linguistics 611, a low sample size of novice raters as well as an insufficient number of practice examples made any statistics insubstantial in order to make any lasting judgments on the effectiveness of the training workbook.

However, the high correlation shown by Pearson product-moment correlation implies that the novice raters were developing the key skill of recognizing the relative strength of each piece of writing compared to another, even if they were struggling to situate their scores on the rubric with the accuracy of an experienced rater. In other words, the novice raters were rating higher-level essays with a higher score and lower-level essays with a lower score. While their agreement was not shown to be improved by the training, this correlation shows that they were consistent in their ability to rank the essays properly.

### **Question 3**

The third question addresses novice raters' intra-rater reliability following exposure to borderline training with experienced rater rationales. The question inquires if their level of consistency is adequate for the raters to safely begin rating high-stakes tests. In viewing the infit scores provided through Rasch modeling, novice rater consistency was generally satisfactory. The data collected implies that while each rater does not have perfect agreement between their ratings and the established scores (as described in answering the second question of this thesis), they are consistent. Therefore, novice rater errors can be modeled out using MFRM. If the raters were inconsistent: sometimes rating

harshly, sometimes rating with more leniency, and showing no pattern to their rating, then the raters would require more intervention, training, and practice before they could safely rate high-stakes tests. This outcome is apparent in one rater, Rater 11, whose infit scores show that she is not consistent in her rating, and therefore requires more training before she should be allowed to rate high-stakes tests. Overall, however, the majority of the novice raters showed that following exposure to the training workbook, they were effectively prepared to rate due to their consistency. Further training is suggested regarding the rating of borderline samples, but overall the raters have proven themselves consistent enough that their errors can be modeled out.

### **Limitations**

Because the nature of the thesis was to create a training workbook inclusive of borderline samples of writing and experienced rater rationale, and the goal of evaluation was secondary, the actual data collection and analysis had several limitations. Many of these limitations have already been mentioned in the discussion of data presented in this chapter, but will also be summarized and evaluated here. In the future, efforts should be made to more fully assess the validity and influence of the implementation of the training workbook created. It appears that had further emphasis been placed on the design phase of this training workbook, the evaluation phase may have gone more smoothly and produced stronger statistical results and more solid evidence in favor of the positive impact of its use.

First, the sample size of only 10 novice raters was too small to truly assess the effects of the training. Future studies would be strengthened with a minimum of 30 novice raters in the sample to gather data that could be more statistically sound

(Campbell, 2011). Larger sample sizes are often difficult, if not impossible to provide since there are rarely 30 new raters in any given cohort of raters. This challenge became evident as the data collection was being established for the evaluation of the training workbook, and out of necessity, it was decided that the ten raters available would be sufficient.

Secondly, in a future study of the effectiveness of the training workbook, the number of practice samples which each novice rater was expected to rate could be increased to a minimum of 30. This could increase the statistical significance of the study. This was not considered in the planning stages of data collection, and each novice rater rated 20 essays over the course of the training. This use of fewer essays did not render the results invalid, but future studies could be improved by the use of this minimum (Campbell, 2011).

Next, it is possible that the meeting with novice raters to provide the treatment and calibration was not sufficiently thorough. High levels of time and attention should be given to the discussion of the rubric and the samples rated. The meeting for the pilot was organized and planned, with careful consideration given to the material presented and the topics discussed, but it may have been more effective with more time spent rating a higher quantity of samples together and discussing their characteristics. Further research regarding the best practice for training and calibration meetings should be included in a later study.

Additionally, raters mentioned a greater comfort level with the benchmark examples of writing because they had been exposed to them for a longer period of time.

The pilot of this study was affected by the time constraints set up in the class curriculum. In the future, researchers could allot more time to rating strictly benchmark samples of writing to become completely comfortable with writing that clearly follows the rubric. Then researchers could slowly expose the novice raters to borderline samples. After being required to rate a mix of benchmark samples and borderline samples, the treatment of borderline examples and experienced rater feedback should be introduced, after which the novice raters would be asked to rate more borderline samples of writing. This staggering of the training which was minimal in the pilot due to time constraints may provide a more complete view of the effect of the training workbook upon novice raters' ability to rate borderline samples of writing. This will also combat the concern mentioned in the discussion of the data that the novice raters could have been exposed to too much information in a short period of time, causing overload.

One final limitation to be discussed involves the line of questioning in the survey (see Appendix D). The goal of the survey was to assess how the novice raters felt post-treatment. One main question from the survey asked raters upon which segments of the training workbook they relied while rating. However, the goal of the training workbook is to improve rater consistency when rating borderline samples of writing. The survey does not directly address when novice raters relied upon each of the segments. It may be beneficial to do more in-depth analysis of when raters use each segment of the training workbook as a resource.

### **Recommendations for Future Research**

The results of this study suggest several possibilities for future research in regards to borderline samples of writing. The first few recommendations, mentioned in the

context of the limitations, are to add to the study of the effectiveness of the training workbook created here. This thesis, its aims, questions, and the discussion of limitations can be used as a guide to properly assess the validity of the training workbook.

Another possible topic for future study regards assessing the tenets of the creation of the training workbook. One question to be researched is if infit statistics are the most effective method for selecting benchmark and borderline examples of writing. Researchers could select a large quantity of examples based on infit statistics, then ask a sample of experienced raters to rate the writing. Raters could then be asked if they feel each sample is a clear example of a level of the rubric, or if they feel it is a borderline sample which contains elements of two levels of the rubric. The creation of this training workbook relied upon such raters, and a question to be addressed in the future is what truly qualifies a rater to be considered experienced.

Finally, the data seems to suggest that access to the borderline samples and experienced rater feedback during initial training does not significantly improve intra-rater reliability. However, a longitudinal study might provide insight into the development of rater expertise over time, including when to expose developing raters to borderline samples and experienced rater feedback. As such, a longitudinal study is recommended where researchers determine what it means for novice raters to become experienced raters, and how long it takes when utilizing the training workbook provided here.

## **General Conclusion**

Specific training for novice raters regarding borderline samples of writing has not been discussed in the ESL field. It is the researcher's hope that this thesis will start a conversation regarding the need to address borderline samples of writing. It is a pervasive problem for all raters, but most especially novice raters, and the creation of this workbook only begins to address that pervasive need. Although there were limitations in the evaluation of this training workbook, the qualitative data suggests that future evaluation may prove this training more effective. It is hoped that future research will be pursued and will validate the use of the product created. Future researchers interested in the topic of borderline writing samples can use the training workbook and accompanying project assessment as a starting point to ask further questions regarding borderline samples, and as guide to begin training new raters concerning these samples.

## References

- Anderson, L.W., & Krathwohl, D.R. (eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman
- Arizona Department of Education. (2005). Six trait rubric—Official scoring guide for AIMS. Retrieved from <http://ade.state.az.us/sbtl/6traits>.
- Bargainnier, S. (2004). Fundamentals of Rubrics. In D. Apple (Ed.), *Faculty Guidebook*. Lisle, IL: Pacific Crest Inc.
- Becker, A. (2010). Examining Rubrics Used to Measure Writing Performance in U.S. Intensive English Programs. *The CATESOL Journal*, 22(1), 113-130.
- Brookhart, S.M. (1999). *The Art and Science of Classroom Assessment: The Missing Part of Pedagogy*. ASHE-ERIC Higher Education Report (Vol. 27, No. 1). Washington, DC: The George Washington University, Graduate School of Education and Human Development.
- Brookhart, S. M. (2004), Assessment theory for college classrooms. *New Directions for Teaching and Learning*, 2004: 5–14. doi: 10.1002/tl.165
- Brown, J.D. & Hudson, T. (2002). *Criterion Referenced Language Testing*. Cambridge UK/New York: Cambridge University Press.
- Campbell, J. (2011). Why is 30 the “Magic Number” for Sample Size? *.problem.solve.create.share*. Retrieved May 8, 2013, from <http://www.jedcampbell.com/?p=262>

- Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, 8, 165-191. doi:10.1016/S1075-2935(03)00018-7
- Chow, S., & Liu, J. (2004). *Design and analysis of clinical trials: concepts and methodologies* (2nd ed.). Hoboken, N.J.: Wiley-Interscience.
- Driscoll, L. A. (1996). *10 steps to district performance assessment*. Memphis, TN: Memphis City Schools, Department of Research, Standards, and Accountability. (ERIC Document Reproduction Service No. ED406396).
- Eckes, T. (2011). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments*. Frankfurt: Language Testing and Evaluation.
- English Language Center. (2011). *Timed Writing Rater Training*. Provo, UT.
- Greer (2012). [Survey of IEP Administrators]. Unpublished raw data.
- Harsch, C., & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing*, 17, 228-250. Retrieved from the Elsevier database.
- Haswell, R.H. (1998). Rubrics, Prototypes, and Exemplars: Categorization and Systems of Writing Placement," *Journal of Assessing Writing* 5.2 (1998): 231-268.
- Interrater reliability (Kappa) using SPSS. (n.d.). *Statistics Tutorials for SAS, SPSS, WINKS, Excel*. Retrieved May 7, 2013, from <http://stattutorials.com/SPSS/TUTORIAL-SPSS-Interrater-Reliability-Kappa.htm>

- Jang, S.Y. (2010). *The development and evaluation of a systematic training program for increasing both rater reliability and rating accuracy*. Retrieved from <http://hdl.handle.net/2142/15599>
- Janopoulos, M. (1986). The Relationship of Pleasure Reading and Second Language Writing Proficiency. *TESOL Quarterly*, 20(4), 763-768. doi: 10.2307/3586526
- Johnson, D., and VanBrackle, L. (2012). Linguistic discrimination in writing assessment: How raters react to African American “errors,” ESL errors, and standard English errors on a state-mandated writing exam. *Assessing Writing*, 17(1), 35-54. Retrieved from ScienceDirect database.
- Kegley, P. H. (1986). The Effect of mode discourse on student writing performance: Implications for policy. *Educational Evaluation and Policy Analysis*, 8(2), 147. Retrieved May 9, 2013, from the Sage Journals database.
- Knoch, U., Read, J., and von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12, 26-43. Retrieved from Elsevier database.
- Linacre, J. M. (2007). *A user's guide to FACETS*. Chicago: Winsteps.com.
- Lovorn, M.G., and Rezaei, A.R. (2011). Assessing the assessment: Rubrics training for pre service and new in-service Teachers. *Practical Assessment, Research & Evaluation* 16(16). Retrieved from [www.pareonline.net/pdf/v16n16.pdf](http://www.pareonline.net/pdf/v16n16.pdf)

- McNamara, T., & Knoch, U. (2012). The Rasch wars: The emergence of Rasch measurement in language testing. *Language Testing*, 29(4), 555-576. Retrieved from the EBSCO Host database.
- Moskal, B. M. (2000). Scoring rubrics: what, when and how? *Practical Assessment, Research & Evaluation*, 7(3). Retrieved from <http://PAREonline.net/getvn.asp?v=7&n=3>
- Myford, C.M. & Wolfe, E.W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422.
- Nevada Department of Education. (2000). *Writing proficiency examination guide 2000–2001*. Carson City, NV. (ERIC Document Reproduction Service No. ED447485).
- Osborn Popp, S.E., Ryan, J.M., & Thompson, M.S. (2009). The critical role of anchor paper selection in writing assessment. *Applied Measurement in Education*, 22, 255-271, doi: 10.1080/08957340902984026
- Parke, C. S. (2001). An approach that examines sources of misfit to improve performance assessment items and rubrics. *Educational Assessment*, 7(3), 201-255. doi: 10.1207/S15326977EA0703\_02
- Penny, J., Johnson, R. L., & Gordon, B. (2000). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing*, 7(2), 143-164. Retrieved May 9, 2013, from the Science Direct database.
- Reckase, M. D. (2009). Historical background for multidimensional item response theory (MIRT). *Multidimensional Item Response Theory*, 57-77.

- Rezaei, A. R., & Lovorn, M. G. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing, 15*(1), 18-39. doi:10.1016/j.asw.2010.01.003
- Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*(6), 956-970. doi:10.1037/0021-9010.85.6.956
- Schreyer Institute (2007). The Basics of Rubrics. Retrieved from <http://www.schreyerinstitute.psu.edu/Tools/Rubric/>
- Smith, D. (2000). Rater judgments in the direct assessment of competency-based second language writing ability. In G. Brindley (Ed.), *Studies in immigrant English language assessment* (pp. 159-189). North Ryde, N.S.W.: National Centre for English Language Teaching and Research.
- Spandel, V. (2006). In defense of rubrics. *English Journal, 96*(1), 19-22. Retrieved from <http://www.jstor.org/stable/30046656>.
- Tierney, R. & Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation, 9*(2). Retrieved from <http://PAREonline.net/getvn.asp?v=9&n=2>.
- Using Generalizability Theory in Test Design (2004). *The Researcher, Spring*, 1-2. Retrieved from <http://www.ets.org/Media/Research/pdf/researcher04.pdf>

- Vaughan, C. (1991). Holistic assessment: What goes on in the raters' minds? In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex Publishing Corporation
- Veerappan, V. a/l., & Sulaiman, T. (2012). A Review on IELTS Writing Test, Its Test Results and Inter Rater Reliability. *Theory and Practice in Language Studies*, 2(1), 138-143.
- Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. New York: M. Dekker.
- Washington Office of the State Superintendent of Public Instruction. (2001). *Grade 10 anchor set annotations from the spring 2001 Washington Assessment of Student Learning in Writing (with) presentation guide for principals*. Olympia, WA. (ERIC Document Reproduction Service No. ED466812).
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- White, E. M. (1984). Holisticism. *College Composition and Communication*, 35(4), 400-409. Retrieved May 7, 2013, from [www.jstor.org/stable/357792](http://www.jstor.org/stable/357792)
- Willard-Traub, M., Decker, E., Reed, R., & Johnston, J. (1999). The development of large-scale portfolio placement assessment at the University of Michigan: 1992–1998. *Assessing Writing*, 6(1), 41-84. Retrieved June 10, 2013, from the ScienceDirect database.

# English Language Center

# Timed Writing Rater Training

Winter 2013

Packet Instructions

Rubric and instructions for use

Benchmarks Examples

Borderline Examples and Experienced Rater Feedback

Practice Samples

**Please return to Dr. Troy Cox**

Dear Novice Rater,

This packet is designed to assist you in gaining confidence and experience in rating Timed Writing. The following instructions will guide you as a new rater through the use of this packet.

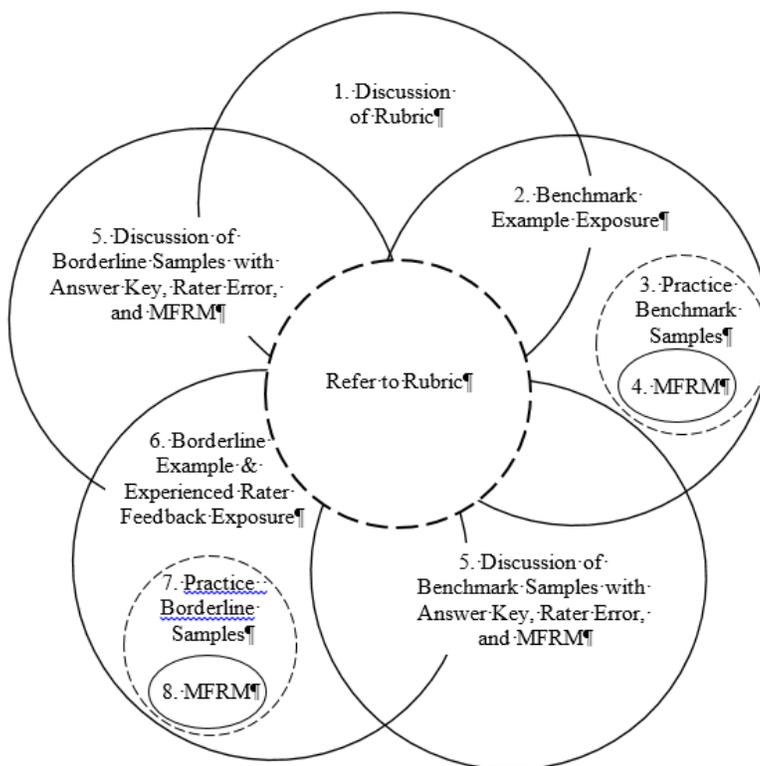
Instructions; Section 1:

1. Read through and become familiar with the Rubric and its instructions for use.
2. Read the Benchmark examples from Part 1 of the binder while comparing them to the rubric. Try to find sentences in the rubric that prove why the benchmark is a clear example of its given score.  
*Example: Why did the Benchmark 3 receive a score of 3?*  
\*You don't need to write anything here, just become familiar with the rubric and the benchmarks.
3. Your administrator will provide you with Handout 1.gap
4. Read each practice essay whose ID number is listed on Handout 1.gap
  - a. Rate each essay according to the rubric.
  - b. Record your scores on Handout 1 in pen.
  - c. Avoid using .5 or plus or minus scoring. Give the essays a solid score of a 3 or a 4, etc. Not a 3+.
  - d. Avoid discussing the rating with others. This is not a collaborative process.
  - e. Write any comments you have as you rate in the comment section.  
*Comment examples:*  
*This was difficult to rate because...*  
*It was clear this should receive a score of... because...*  
*I wanted to give this a 3- or a 2+ because... but I gave it a 2 because...*  
*In the rubric, it states... so I gave this essay a score of...*
5. Submit Handout 1.gap to your administrator.
6. Your administrator will arrange for a training meeting. Bring your binder and be prepared to discuss the rating process and experience. At this meeting you will be provided with the actual scores of the essays you have rated (Handout 1), and will participate in a discussion of benchmark vs. borderline samples of writing.

## Instructions; Section 2:

1. Read the borderline examples from Part 2 of the binder. Compare each example to the rubric. Read the Experienced Rater Comments to understand why the borderline was eventually rated with its given score.  
\*You don't need to write anything here, just become familiar with the rubric, borderlines, and experienced rater feedback.
2. Your administrator will provide you with Handout 2.gap
3. Read each practice essay whose ID number is listed on Handout 2.gap
  - a. Rate each essay according to the rubric.
  - b. Record your scores on Handout 2 in pen.
  - c. Avoid using .5 or plus or minus scoring. Give the essays a solid score of a 3 or a 4, etc. Not a 3+.
  - d. Avoid discussing the rating with others. This is not a collaborative process.
  - e. Write any comments you have as you rate in the comment section.

*Comment examples:*  
*This was difficult to rate because...*  
*It was clear this should receive a score of... because...*  
*I wanted to give this a 3- or a 2+ because... but I gave it a 2 because...*  
*In the rubric, it states... so I gave this essay a score of...*
4. Submit Handout 2.gap to your administrator.
5. Your administrator will arrange for another meeting where you will discuss the established scores of the essays you have rated (listed on Handout 2). It will be determined if you are prepared to rate high-stakes tests.



**Instructions for using the *ELC Timed Writing* rubric:**

1. Use the “Text Type” box on the left to get a general feel for the **length** and **organization** of the response.
2. If you need additional clarification about a student’s level, check the “Content” box. Here you can get a sense of the writer’s **functional ability with language** and **vocabulary**.
3. If you are still not sure which level to choose, look in the “Accuracy” box to get a feel for appropriate **grammatical complexity** and the conveyance of **meaning**.
4. Still feeling a little unsure? Consider semantics, pragmatics, punctuation, spelling, or concision to inform your choice.
5. What if a student’s skills are spread over different levels (ie: highly complex and cohesive writing but with exceptionally poor grammar)? Use your best judgment, bearing in mind that sometimes a student’s demonstrated strength in one area can compensate for weakness in another. On the other hand, sometimes a student’s weakness is so overpowering that it negates some of their strengths.

<b>LATs Timed Writing Rubric Foundations Program</b>			
Level	Text Type	Content	Accuracy
	<ul style="list-style-type: none"> <li>• <b>Length</b></li> <li>• <b>Organization</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Functional Ability with the Language</b></li> <li>• <b>Vocabulary</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Grammatical Complexity</b></li> <li>• <b>Meaning</b></li> </ul>
4—ready for Academic A	Multiple paragraphs are present with solid organization on the paragraph level (topic sentence, supporting detail, etc.)—and basic organization on the essay level.	Writing is usually in the context of personal interests and experiences, daily routines, common events, and immediate surroundings. Concrete topics are discussed.	Some mastery of past narration (past progressive, simple past, etc.) with both regular and irregular verbs. Inconsistencies occur in other time frames. The majority of sentences will be shorter. Complex sentences are common and generally accurate. Problems in accuracy may occur, but the overall meaning is not obscured.
3—ready for Foundations C	At least one paragraph (for 30-minute writing portion). Organization is weak with multiple paragraphs.	Able to meet some limited practical writing needs—writing about personal interests and experiences, daily routines, common events and immediate surroundings. Structure and meaning are highly predictable. Usually relating to personal information or immediate surroundings. Writing exhibits a small range of vocabulary.	Solid writing of short and simple conversational style sentences with basic subject-verb-object word order. Exhibits some consistent success with compound and complex sentences. Basic errors in grammar, word choice, punctuation and spelling. Most writing framed in the present. Some mastery of past narration in the simple past with regular verbs. Other time frames may be attempted with some success. However, natives used to the writing of non-natives can usually understand the meaning.
2—ready for Foundations B	Simple sentences; some compound and complex sentences with repetitive structure. Lacks clear paragraph organization.	Close, personal explanations with very limited vocabulary. Writers can express themselves within a very limited context.	Able to write clear simple and compound sentences with limited vocabulary and conjunctions. Attempt to create some compound sentences using connectors like “because.” Writing is successful in present tense, occasional and often incorrect use of past or future tenses. Text is writer-centered.
1—ready for Foundations A	Some simple sentences.	Reliance on formulaic/memorized language.	Exhibit accuracy when writing on well-practiced familiar topics using limited formulaic language. Sentence-level constructions. The volume of writing may be so small that it undermines the reader’s ability to evaluate accuracy, or errors occur so frequently that the purpose of the writing task may not be completely clear.
0—ready for Foundations prep	Able to supply limited information on forms and documents—i.e. names, numbers, nationality, etc.	With adequate time and cues may be able to produce a limited number of isolated words.	Inability to use sentence forms. Volume of writing is insufficient to assess accuracy.

## Timed Writing Rubric Academic Program

Level	Text Type	Content	Accuracy
	<ul style="list-style-type: none"> <li>• Length</li> <li>• Organization</li> </ul>	<ul style="list-style-type: none"> <li>• Functional Ability with the Language</li> <li>• Vocabulary</li> </ul>	<ul style="list-style-type: none"> <li>• Grammatical Complexity</li> <li>• Meaning</li> </ul>
7—ready for university courses	Essays may be a full page or more. Organization and transitions make writing very easy to read and understand.	Able to write more complex elaborations (i.e. summaries and paraphrases dependent on task). Uses a range of general and academic vocabulary. Writing uses a variety of cohesive devices. Provides sufficient background information as evidence that the writer is generally aware of the readers' needs. Readily understood by native readers.	Excellent control of a full range of grammatical structures. Small errors in grammar, syntax, spelling or punctuation may occasionally distract a native reader, but there is no evidence of a pattern of errors. Writing is easy to read, but the writer may fail to convey the subtlety and nuances of the language.
6—ready for Academic C	Multiple paragraph essays with clear organization.	Appropriately uses abstract and concrete language to convey meaning. Message is pragmatically accurate for easy reading. Attempts to use cohesive devices but they may be redundant. Wide and varied general and academic vocabulary and topics.	Able to use language in detail in all time frames. Control of syntax in word order, coordination, and subordination while not perfect, does not distract greatly from meaning. No or very few spelling problems. Evident use of a wide range of structures. May be a few errors with complex and infrequent grammatical structures.
5—ready for Academic B	Multiple paragraphs with evidence of organizational markers on the essay level.	Able to meet all practical writing needs. Favors concrete ideas and some more abstract topics may be discussed but meaning is perhaps unclear. Vocabulary is quite varied, but not to the extent of level 6.	Able to use language in major time frames. There is apparent subordination, but it is more like oral discourse. Mastery of grammar with simple sentences. More complex sentences are attempted, but contain errors and may not be clear.
4—ready for Academic A	Multiple paragraphs are present with organization on the paragraph level (topic sentence, supporting detail, etc.)—but perhaps not on the essay level.	Writing is usually in the context of personal interests and experiences, daily routines, common events, and immediate surroundings. Concrete topics are discussed. Some examples and explanations may not be clear. Some points may not be well supported or explained.	Some mastery of past narration (past progressive, simple past, etc.) with both regular and irregular verbs. Inconsistencies occur in other time frames. The majority of sentences will be shorter. Complex sentences are common and generally accurate. Problems in accuracy may occur, and the overall meaning may occasionally be obscured.

Packet Part 1

LAT Essays  
Benchmark Example Score: 1

ID: 20367

30 Minute Essay

Word Count (excluding prompt) 29

905520367

*Prompt: You are the freshman counselor at a university. You are writing an essay to all incoming freshman that answers the following question: What are the 3 most important subjects that students should study in college to prepare them for the future? Why?*

You have studing for better future.

You nide study hard for take better college.

Study if your more nice.

Your future is better whem you have one good college.

Packet Part 1

LAT Essays  
Benchmark Example Score: 1

ID: 39890

30 Minute Essay

Word Count (excluding prompt) 14

086539890

*Prompt: Do you agree or disagree with the following statement? All drugs should be legal for people over 18. Use details and examples to support your opinion.*

In my opinion I agreee. Because All drugs should be legal for people over.

Packet Part 1

LAT Essays  
Benchmark Example Score: 2

ID: 22304

30 Minute Essay

Word Count (excluding prompt) 186

184122304

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

There were a lot of animals. First I went to see monkey. They was eating some bananas, and they was playing with visitor. They had red hip. I wanted to touch them. When I saw them, They looked at me interestly. Second I went to see lions. When I saw them, I felt scery, because they have dangerous teeth. Then I went to see elefant. They was so big. And they have long nose and big ears. I wanted to ride them. So I paid some money. It is very interesting. Then I went to see tutle. They were very slow. So while I saw them, they moved only 1metar. Then I went to see birds. They were in small cage. I thought that they want to fly to sky. So I leaved ealy. It was 3:00 PM. I was hungry. So I walked to find restaurant. But I couldn't find any restaurant. So I walked and walked for 1 hour. When I found restaurant, the sun set. I was very hungry. So I couldn't any thought. There was only one manu, hamberger. I like the

Packet Part 1

LAT Essays  
Benchmark Example Score: 2

ID: 72815

30 Minute Essay

Word Count (excluding prompt) 130

723972815

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

I think is good idea other corporation sponsor the countries because people don't have money and they give help for represent their countries. One example is about nike help united states why the company think is good idea help the people why is sports and is not bad and is good for the healthy another example more than one the countries and specially the government don't have money for help the women and men for the olympics . I agree with differents companies help the sports , they have money for help the sports it is good and they need to reconognize for the help . It is good idea that all companies help all sports.The gobeernants want help but can not help for don't have money. my opinion inclus

Packet Part 1

LAT Essays  
Benchmark Example Score: 3

ID: 76495

30 Minute Essay

Word Count (excluding prompt) 236

795876495  
 Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.

I disagree with this statement because I have two reasons why I disagree for experience for kids and knowledge.

First, a zoo has useful purpose for experience for kids. If the zoo doesn't have in the world, people can't show any animals. Of course people can see animals on the TV but they can't touch them. For example, when my kids really want to meet a tiger, I should go to the zoo and I can explain my kids about tigers. However, we don't have a zoo, I can't express anymore if they want to meet any animals. When my kids meet the tiger at the zoo, they should memorise easily and it will give my kids for good experience.

Second, They should build their knowledge. When people learn something, there are many different ways. However, the best way is that people directly watch them. For example, When my kids study at home, they should memorise everything with books. Also I push their to study very hard. But it is not necessary because many memory disappear from their brain. However, they directly watch and touch some animals, their memory may be forever. It will be related their knowledge.

In conclusion, a zoo has useful purpose for experience and knowledge for kids. If people know about animals through books, it may disappear quickly. So it is necessary to exist the zoo in each countries.

Packet Part 1

LAT Essays  
Benchmark Example Score: 3

ID: 73871

30 Minute Essay

Word Count (excluding prompt) 184

167273871

*Prompt: Do you agree or disagree with the following statement? All drugs should be legal for people over 18. Use details and examples to support your opinion.*

I agree for all drugs should be legal for people over 18 for many reasons such as more more mature, could be to do decisions, understand better .

First, one teenager don't understand many things, because is very young , and can not decided about your decisions and acts. Because don't have mature for nothing, many teenager depends of the parents. Also, when the person is over 18 understand about the consequences.

Second, the people over 18 understand when to make his decisions , all drugs are very dangerous for health, many teenagers in this time was a dependent. They understand one happening after that.

Finally, drugs is very prejudicial for all ages , but one person over 18 understand better than other person that less 18. We can choose what to do.

In conclusion, in my opinion drugs never can be legal. Also, because can be damage for health , family , and life. But, if we need to choose what age is better, I think for people over 18 because can be mature, could be to do decision , and understand better.

Packet Part 1

## LAT Essays

### Benchmark Example Score: 4

ID: 36339

30 Minute Essay

Word Count (excluding prompt) 437

747036339

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

why Countries do not ffnacially their Athletes

Did you know that many countries do not financially their athletes who compete in the olympics? Mos of the time many excellents athletes have to get some sponsor if they want to go some olympics. I think that they have to do that for thre reasons. First, their countries are poor countries. Second, their country prefer to help in others things like wars, and finally their countries think that they are not a good athletes.

Many contries on the world do not have enough money for financial their athletes for many reasons. First of all, these countries are like in the third world and they do not have many products for export to others countries. In addition, they do not recive any help from rich countries, so they have to survivor just with the things thta their land can give them. Due these although they have a good athletes they can not help them, so their athletes have to get money for their competition from others.

In the similar manner, many countries prefer to give mostly money for finacial wars for several reason. First, these countries think that is better give money to army for bild the best guns in the world. they think that in this way they can show thier power to all the world. Second, they think that give money to sports are not a good idea because for them that mean waste their money in unnecesary thing. For this reason some countries do not help their ahtletes when they need to go to represent their countrie in some olimpycs.

Finally, some countries think that their athetes are not good in their sports. These countries think that is no good give money to some athletes because for them their athletes are not good, so they decide that is not a good idea send their atheltes to olimpycs because they can go down the imagen of their countries. Also, these countries sometimes think that if their athletes do not show that they are the best they prefer that them stay in their countries.

In conclusion, many countries in this world have like last option give thier money to finacial their athletes in the olimpycs. the reasons for this decision are many. These countries are too much poor and they do not have maney, they prefer to financial wars, and they think that their athletes are not good in sports that they practice. I think that is not good because in this way they do not give the change to these people for they can progress in their lives.

Packet Part 1

LAT Essays  
Benchmark Example Score: 4

ID: 54232

30 Minute Essay

Word Count (excluding prompt) 316

039454232

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

The Olympics is held by England this year. Everyone likes to watch those sports events. All of the athletes want to get a good score in this competition. However, the government plays an important role in the Olympics. Those athletes need supports from the government to focus on their competition.

My county gives financially support to the athlets who compete in the Olympics. The government wants those athletes to concentrate on training. I know there are some countries which not support their athletes. I think that will be hard for those athletes because they need to worry about money and try to find sponsors. In addition, it is not easy to find a sponsor to help them. Even they can find a sponsor, it takes time to find and persuate them to help athletes.

I think that the government should think about one issue: winning the prize for thier countries from the Olympics is a good thing or not? If it is a good thing, the government should support them and give them financially support. Therefore, those athletes can focus on their training and try their best to achieve thier goals.

I know there are some athletes who not get any support from their government. They do not any support and audiences to encourage them. I think that they must feel lonely and hopeless. They need some strength from thier country. However, I know there some countries which have their own problems, such as financial crisis and political issue. Therefore, the governments can not support them.

Generally speaking, the best way to support athletes is giving money by the government. The government has a responsibility to support their athletes and help them to win the prizes and achieve their goals. Regardless of the government is rich or poor, the governemt should support them with money or encouragement. There must have a way to help their athletes.

Packet Part 1

## LAT Essays

### Benchmark Example Score: 5

ID: 72406

30 Minute Essay

Word Count (excluding prompt) 667

748572406

*Prompt: Do you agree or disagree with the following statement? It is better to be a member of a group than to be the leader of a group. Use details and examples to support your opinion.*

Being a leader give us the advantage to lead people based on our right thoughts, ideas and feelings. Being a leader also brings responsibility for others and make us more careful at the moment of make a choice. the opportunities to lead a group make us sensitive for others necessities and provide us a sense that every decision that we make can affect not just to the person who make the choice but also to those members who are part of the group. Even though be part of a group is a great opportunity to bring support to the leader and help him in many ways, however, being a leader will cause a lot of positive impact in any group that need to make right choices just when the leader act with righteousness. Being a leader of a group with a right common sense and making the last choice will have the most positive impact.

Being leader of a group with a right common sense has great benefits. when I mention being a leader of a group with a right common sense I mention the opportunity that the leader has to lead others under his thoughts, ideas and opinions mixed with the ideas and opinions of the other members of the group. Leading group with your ideas and with other ideas allow us to achieve goals in a more effective way that just if the leader lead people with their own thoughts and ideas. For example, United States government is based on the ideas that the congress and others are to support and bring companionship to the the President. As we understand the President don't take every decision alone rather than he listen to the congress and other people and after he make a choice. Having the opportunity to lead a group with a right common sense and listening others will bring an important success in the future because 2 heads think more than one.

Being a leader you make the last choice. Making the last choice means choosing the best thing for a group of members. When a leader make a choice is because he already has thought about it, he already has studied any point and came up with what is the best thing for the rest of the group. A leader always provides help others in special the members of the group, so if he realized that something is going to affect them they can notice with clarity and do the best make a final choice. For example, a couple years ago my family have started a business. after some years the economy of the country was going down, so my father who was the leader of the business, made a gathering with the all members of that business group in order to make a choice about how to keep the business active in the future. All of the member thought that the crisis would last for a while and they said that the business should keep going in the same way, but my father who had a good vision about how to make business succeed made the choice to save the 30 percent of all revenues in order to avoid a loss of the business. After years the business reach succeed because of that money that was save. Being a leader give us the opportunity to make the last choice even though he is part of the minority.

In conclusion, having the advantage to lead people with our good judge and making the last decision for them will have more impact in that group in the present and in the future more than being part of the group and support the leader. If a leader leads his group with careful and caring others opinion he will be successful achieving the group goals, but being part of the group our life will be affected for every choice that the leader make. Being a leader will carry out great things for us and for others when we lead them wisely.

Packet Part 1

## LAT Essays

### Benchmark Example Score: 5

ID: 50875

30 Minute Essay

Word Count (excluding prompt) 340

752250875

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

Olympics is the most famous competition and biggest world festival among the athletes. People are also really interested in Olympic games and they are crazy about it when their country's athletes get medals. For this Olympic, some countries financially support their athletes who compete in the Olympics and other countries do not support so they need to find sponsor. I think countries need to financially support for their athletes because athletes can focus on only their position and it will help to introduce their country through their good score in Olympic games.

First of all, if athletes do not have sponsor who can help them financially, they cannot continue their position. They need to find other job that they can afford by themselves. If they have families, it is way harder. However, when athletes can get support financially from the countries, they do not need to worry about purchasing the training and finding sponsors. Therefore they can more concentrate on their job. They can practice without concern or anxious and it also really helpful to get great result in the games.

In addition, Olympics is the biggest world festival and people love sports. All the countries of the world are focus on Olympics and they wondered that which country will get golden medals. Therefore, in naturally, countries can introduce their nation by making good result in the Olympic games. I probably help their economics and cultures according to be well-known. It is really good way to show and get proud of themselves. Thus, it is way better to support their athletes to get good score than use another way to introduce their countries.

In conclusion, there are some benefits not only for athletes but also for countries. Sports can make people to get interests. Hence, it is really good opportunities to be well-known in the world festival. Countries can show that how they are good at sports and what country we are.

Moreover, the athletes can make good record according to get support from countries. It is good for them.

Packet Part 1

## LAT Essays Benchmark Example Score: 6

ID: 73415

30 Minute Essay

Word Count (excluding prompt) 612

814873415

*Prompt: Do you agree or disagree with the following statement? It is better to be a member of a group than to be the leader of a group. Use details and examples to support your opinion.*

Being a leader of a group requires more responsibility than being a member of a group; therefore, being a member of a group will be better. There are some people think that being a leader can be better, but they do not know their abilities. Hence, they fail in being a leader. Leader is the one who lead their members to the best thing and they have to make the greatest decision for their group. It requires so many skills like leadership, dealing, communication, and some others. Nonetheless, being a member requires responsibility too, but less than the leader. A member just needs to take care of his or her responsibility and combine with the others. By being a member of a group, people can avoid some steps or requirements for a leader such as making decision, dealing with other resources, connecting between members.

First, making decision is one of the important requirement of a leader, so members do not need to do that. Members of a group just need to focus on what are their responsibility and finish what they have to. Leader will be the person who make the decision after all. By being a member of a group, people do not need to make a decision, but they have responsibility to assist the leader to make the right decision for the group. Furthermore, members should finish their works on time in order to assist the leader to make the right decision in the right time. Members should focus on what they have to do to turn in their works on time. Finishing their works will assist the leader so much in making decision.

Second, dealing with other resources is important too, but members sometimes do not need to do that. The leaders should deal with other resources in order to help their members acquire the best result. For example, a study group of journalist need to find some sources from the old newspaper and present their works. The leader should contact the library to find that they may have the resources for them or contact someone who has the evidence for their presentation. After that, the leader can assign his or her members to do that specific case. Moreover, members do not need to deal with so many things because they have their specific works, so they just need to focus on that and finish that. Hence, being a member do not require dealing skills.

Finally, the connection between members in a group is the responsibility of the leader; therefore, members should assist the leader but do not take the main responsibility about that. Being a group required the connection between members. Every members in a group cannot avoid their responsibility about that. However, the leader should be the one who takes the main responsibility. If the leader has great skill in communication and leadership, he or she will lead their group perfectly. The connection between members becomes great or worse depends on the leaders. Hence, the members could assist the leader but this is not the main point that the members should focus on.

In conclusion, making decision, dealing with other resources, and connecting members together should not be responsible of the members. The members should focus on their works and their time. Members can have their specific jobs to do, they should turn it in on time. Furthermore, the members can have more benefits because members do not need to spend more time while the leader has to. Being a leader requires so many responsibility and skills. Hence, being a member of a group will be better if someone do not want to take so many responsibility.

Packet Part 1

## LAT Essays Benchmark Example Score: 6

ID: 47935

30 Minute Essay

Word Count (excluding prompt) 440

021947935

*Prompt: You are the freshman counselor at a university. You are writing an essay to all incoming freshman that answers the following question: What are the 3 most important subjects that students should study in college to prepare them for the future? Why?*

Have you ever thought about what are the most important subjects you should learn in college? During the college years you will have opportunities to learn many important subjects and develop skills in a sort of different areas. All of the subjects will be important for your future and all of them will help you to develop different skills that will be useful in your life. However, there are three subjects that are the most important because they will be essential and affect many different areas and tasks in your lives.

The first most important subject is the language. To know how to use the language properly will be essential for succeeding in many different phases in your life. For instance, to apply for a good educational program you will usually be required to take tests that involve the language and require different language skills. In addition, essays are often required to get into a good educational program and sometimes they will be decisive for your acceptance. Moreover, language skills will be very important in your career. In order to be a good leader in a company, for instance, it is very important to know how to communicate well.

The second subject that is essential is math. Not only in the school math will be important for you, but also in your future job. In other words, math will be very important in the university because your math skills will be frequently tested and required to succeed and to be accepted in educational programs. In addition, if you decide to work in the finance area for example, your math skills will be essential and help you to achieve a good position in the area.

The last of the three most important subjects is leadership. Leadership skills and abilities will be very significant and essential in your future. The ability to lead, manage and influence people is one of the more important and most appreciated today. During the college years you will have many opportunities to develop leadership skills. You must seek for these opportunities and use them wisely. Many times you will have assignments to work in groups or to participate in projects with other students. Use these opportunities to learn how to communicate well and interact with others. Also try to help other students, have ideas for the projects and to solve problems. It is also important to know how to listen to others' ideas.

After all, although there are many important subjects in college, the language, math and leadership skills you acquire during this time will be essential to bring you success and satisfaction during your whole life.

Packet Part 1

## LAT Essays

### Benchmark Example Score: 7

ID: 12792

30 Minute Essay

Word Count (excluding prompt) 488

695012792

*Prompt: Do you agree or disagree with the following statement? It is better to be a member of a group than to be the leader of a group. Use details and examples to support your opinion.*

It is better to be a leader of a group so you can reach your potential. Being the leader of a group can be compared to the statement of "don't give a fish to a man but teach him how to fish it." Although you can learn from a leader as part of a group, when you are a leader you tend to learn more than the members of the group because the leader has to do an extra effort to lead or to guide a group, and he also has the responsibility to look after others. Being a leader of a group is better because it helps you to reach your potential as human being, it helps you to become selfless and it helps you to know how to solve challenges.

Being a leader can help you to reach your personal potential because you will be able to develop skills you didn't know you had because of the responsibility to lead others. When you are a leader, you will have to find ways to help people to trust you so they can follow you. In order to get people's trust, you need to be accurate in the way you live. People will be looking at you as an example; therefore, you will have to do an extra effort to do the right thing for the right reason even under tough conditions. When you develop the a strong character of righteousness in making desitions, people will see the result in the team, and as a consequence they will trust you enough to follow your instructions. Thus, being a leader will help you to reach your potential as you help others to find theirs by guiding and supporting them.

Along with reaching your personal potential, being a leader will help you to be selfless.

Nowadays, people tend to be focused on their own interests and that may be the case with the members of a group; however, the leader has to forget or to leave aside his own interests to help those under his stewardship. Being selfless is an attribute that can only be learned when you put others' interests before yours. A leader cannot be a good leader if he is going after his own interests. Being a leader will provide you the opportunity to serve others and to make other's life better.

Finally, when a leader becomes selfless, his attitude toward solving problems can also be enhanced. A good leader has to be good at solving problems in less time than the usual. Thus, being a leader will help you to develop skills that will help you to focus more on the solution than on the problem.

In conclusion, it is better to be a leader because you have been sent to reach your personal potential through the opportunity of rightly guiding others, you will be able to be less selfish by watching after others, and you

Packet Part 1

LAT Essays  
Benchmark Example Score: 7

ID: 13254

30 Minute Essay

Word Count (excluding prompt) 441

634013254

*Prompt: Do you agree or disagree with the following statement? All students should be required to study art and music in secondary school. Use specific reasons to support your answer.*

Art is an essential component of human life. Every single culture has developed a certain kind of art that identifies them. The reason for what art is very important is that people express themselves when playing music or painting, sculpting or building. Because it has a huge importance to express emotions and feelings, all students in secondary school should be required to study art and music.

Teenagers should work on their creativity in art and music classes. Most people know that the brain has a logic component and a creative one. At school, in subjects like linguistic or mathematics, the logic component is developed, and in music and art subjects, the focus is creativity. Creativity is an important part of our lives; it helps make connections in our brains to use the theory in the practical life. Through different experiences, people react in different ways. For instance, some people analyze their lives and create great ideas while listening to music. In art classes adolescents would have the opportunity to think in a different way.

Another benefit of studying music and art is that people discover talents and abilities they did not know they have. When people do not try something, they do not know if they like it or not, or if they are good at that or not. If students have these classes at school, they will have new opportunities to learn about their own capacities. Furthermore, they will feel satisfaction when they achieve the goal of expressing themselves in a certain way they enjoy. They will feel they are valuable because they can express what they feel and nobody could say that they are not right, like at maths. In these classes, students will discover talents that would help them to feel their value.

The last reason to study music and art at school is that it is part of general culture. People need to know more about their own nation through the different ways that people used to express their emotions. For example, paintings tell us with too much detail how people feel in war times, or when they enjoyed a stable economy. When people know about art and music, they have more issues to talk about and discuss with different people and they can learn more about other's opinions.

In summary, students should be required to learn art and music in secondary school firstly because they will develop their creativity to think in a different way; secondly, they will discover their abilities and that will give them the opportunity to feel satisfaction; and, finally, they will gain new information and skills to continue increasing their general culture.



Packet Part 2

LAT Essays  
Borderline Example    Score: 1

ID: 36092

30 Minute Essay

Word Count (excluding prompt) 140

331236092

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

Countrys and companys should support their athletes in the Olimpics because the olympics is very important to our country. Almost country support this sports and the big companys in the world support Olympics sport because it is so important. United state of America spend alot of money to support this sports.

Almost countrys spend millions dollars to support it, for example China, Syria, and France. There are many companys support Olampic sports also because there are country can not pay cause no alot of moneny.

Packet Part 2

## Experienced Rater Feedback

ID: 36092

---

### **Why could this have been rated a 2?**

Rater 1: *The complex sentences and good use of vocabulary. Almost paragraph structure.*

Rater 2: *Simple Sentences some complex; vocabulary seems to be quite limited; restricted to words in the prompt; successful use of because connectors; almost exclusively present tense but is relatively accurate in present tense.*

Rater 3: *The student attempted to write two paragraphs. They had some compound/complex structures being used.*

### **Why was this rated a 1?**

Rater 1: *It is sentence level; there is a limited quantity of text.*

Rater 2: *The most successful sentences were the simpler sentences. Language seems quite formulaic because of the repetition of words and structures. Volume of writing. Errors reduce the volume of understandable communication even further.*

Rater 3: *The sentences were fairly accurate but formulaic as they used the same sentence patterns and often repeated the same information.*

Rater 4: *Sentences are simple or incomplete. Language seems more formulaic. Frequent errors.*

Packet Part 2

LAT Essays  
Borderline Example      Score: 2

ID: 41393

30 Minute Essay

Word Count (excluding prompt) 422

441941393

*Prompt: Do you agree or disagree with the following statement? All students should be required to study art and music in secondary school. Use specific reasons to support your answer.*

current people are learning other required to study like play piano, and art and also a lot of instruments. i think this is good for our ability. but parents wanted to their children learn when they are young. but it's not good way to learn. because when children learn this things when they are young they don't know much about this thing so they don't care bout it. but when they turn teenager it is good time to learn. i agree with learning those things. baccuse nowadays when we want to go to college or university have they good exception person who is rearlly good at piano they can go to good univerisy for example person who really good at piano they can go to hardverd for music. but if we didn't learn those kind of things we have to learn only math, science or history. then what can we do with this?. some people are not interesting about learning other instrument, art and music. but other people are really good at those things. in my country these things are not requirement. but i want to learn these things. unfortunailly we don't have time to learn after school. so a lot of people want those things in school. we can learn during school time. and also we can have more ability and it is good when we have good talent we can show to other people. and also whe n they become a adult they can use their abiliy for get a job. person who good at art they can get a good for artist.nowdays a lot of people are really good at music instrument like piano and viloin. i think that is because people are learning when they were young and they also enjoyed learning thosre things. i heard that some school is mandatory that learning art and music. when i was in my country i learned choir we learned how to sing a song and we prepared performance it was really good experince. when we learn those things we can also spend good time with these things. as a result i think learning other things like music and art these are good for our ability also students can enjoy their school life.it has a lot of reason to learn art and music those things are good for our life.

## Experienced Rater Feedback

ID: 41393

### Why could this have been rated a 3?

Rater 1: *Solid writing of conversational style sentences, good use of subordinating and coordinative conjunctions, length of text*

Rater 2: *Very strong and appropriate use of vocab, highly accurate in basic subject-verb word order; some success with complex/compound sentences*

Rater 3: *The student uses a larger range of vocab than a level 2. The student uses the present tense successfully and at times uses the past tense correctly.*

Rater 4: *Fluency- quantity of expression was impressive. Complexity: compound complex sentences. Present and past tense.*

Rater 5: *Length but don't be deceived! Perhaps a few cognates.*

Rater 6: *The writing portion is long, but I think the writer lacks some consistency in sentence structure and compound complex sentences that would qualify him for a 3. No it could not be 3.*

Rater 7: *It has at least 1 paragraph.*

Rater 8: *Could be called a paragraph- compound and complex sentences mostly successful.*

### Why was this rated a 2?

Rater 1: *Lacks clear paragraph organization. Text is writer focused*

Rater 2: *Lacks clear paragraph organization*

Rater 3: *The student wrote a lot, but there is no clear organization. The student tries to use connectors in complex/compound sentences, but can't quite do it successfully.*

Rater 4: *Lacks basic writing conventions—capitalization excessive punctuation. Sentences, even simple sentences, lack clarity.*

Rater 5: *In order to write at paragraphs level they have to write good sentences. Most of these are not good sentences. It also lacks organization on the paragraph level and is simple sentences. They attempt to create compound sentences*

Rater 6: *Lacks clear paragraph organization. The writer attempts compound sentences using conjunctions like "but, so and," but the ideas aren't complete.*

Rater 7: *Punctuation errors, lack of clear paragraph structure.*

Rater 8: *Personal explanations. Lacks clear paragraph organization. Uses conjunctions.*

Packet Part 2

LAT Essays  
Borderline Example      Score: 3

ID: 99840

30 Minute Essay

Word Count (excluding prompt) 219

672999840

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

I agree that A zoo is usful for three reasons. First, a study of animals is very important and it needs a special enviroment such as a zoo. Human life is related to not only farm animals but also wild animals. To study wild animails efficiently, a zoo is very useful. For example, if scientists want to study tigers, it is difficult to go the wild areas everytime. It is more efficient to set up the environment like a wild. Second, a zoo is good for students to educate. When students need to be educated about animals , it can make them study easily. Third, a zoo is a good place where people have a good time with their family and friends. If a zoo is set up for animals well, there is no problem for people to enjoy it there. Especially, it is difficult for people in a big city to find a good place where they have a time with family. for A family with children , a zoo is the one of the best place to enjoy. In conclusion, I agree thant a zoo has an useful purpose because a zoo can use for a study of animal , education for children. Moreover, it also is the place where people relax and enjoy with their family.

## Experienced Rater Feedback

ID: 99840

### **Why could this have been rated a 4?**

Rater 1: *Basic essay organization- though not divided into multiple paragraphs, compound sentences, good vocab use.*

Rater 2: *I would not give this a 4. Some development into fuller essay. Attempts to use narration. Attempts to use academic vocab*

Rater 3: *The content is not personal, it is more abstract. The student has a larger range of vocabulary than level 3.*

Rater 4: *I don't know that it could have. Insufficient quantity. There are multiple paragraphs.*

Rater 5: *Even though it is not formatted into paragraphs, there are multiple paragraphs with good paragraph organization (meaning details and topic sentences)*

Rater 6: *There is a solid organization on the paragraph level. Topic sentence, supporting details and concluding sentences.*

### **Why was this rated a 3?**

Rater 1: *One good paragraph, solid writing of short sentences*

Rater 2: *Clear paragraph structure, paragraphs are not developed enough to receive a higher score; examples and support are too personal; grammar lacks refinement.*

Rater 3: *The student wrote one paragraph to answer the prompt. The student stayed in the present tense and did not attempt other time frames.*

Rater 4: *Text type meets a 3. At least one paragraph. Successful simple sentences.*

Rater 5: *Not all ideas are developed, only one paragraph with formatting. Lots of basic punctuation, capitalization and grammar errors. Small range of vocabulary.*

Rater 6: *There is clear paragraph organization. The writer writes in the present. The writer attempts to use compound sentences.*

Packet Part 2

**LAT Essays**  
**Borderline Example      Score: 3**

ID: 29615

30 Minute Essay

Word Count (excluding prompt) 170

002529615

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

Would you like to live in a jail? the answer is simply, anybody is grateful for your freedom so why animals have to live in a zoo. In this sense, some people believe that animals must live in their original habitat because they need happiness, freedom, and make a life like humans.

Some investigation shows that animals which live in the jungle, woods, etc. are more happy than the animals that live in a zoo. For example their expression of their face is more natural than the zoo animals. In addition some attitudes or behaviors show that they can sing, dance or speak more than the animals that live in a zoo, so the animals which live in a zoo have a boring life.

The animals appreciate their freedom because they can run, play, and eat anywhere, so they can be more happy than the others.

Finally when animals live free they learn how to get food, how to build a home, and how to make a family. This

Packet Part 2

## Experienced Rater Feedback

ID: 29615

### Why could this have been rated a 4?

Rater 1: *Multiple paragraphs. Basic org on an essay level- good thesis statement, complex sentences*

Rater 2: *Some development into multiple paragraphs; thesis statement and topic sentences exist but need refinement*

Rater 3: *The student has a larger range of vocab than a level 3. The student attempts to have multiple paragraphs with limited organization.*

Rater 4: *This shouldn't have received a four. Nowhere near enough indication that they are ready for academic writing in terms of quantity and quality.*

Rater 5: *I don't think it should have received a four. I think someone probably rated it a 4 because they attempt to use high vocabulary and abstract tone.*

Rater 6: *I guess because the writer showed some mastery over complex sentences.*

Rater 7: *Clear structure, at least one paragraph*

Rater 8: *Complex sentences, vocabulary?*

### Why was this rated a 3?

Rater 1: *basic errors in grammar word choice and spelling, two good paragraphs, two weak "paragraphs" (sentences)*

Rater 2: *Vocab and examples are simple and personal; poor word choice confounds meaning. Structure is highly predictable.*

Rater 3: *Some of the words are a mix of Spanish and English. The last two paragraphs are not really paragraphs.*

Rater 4: *There are some sloppy paragraph structures but more in content less in execution. Lots of problems with word choice and spelling. Less conversational in tone. Attempts at complex expression more than expected from a 3, but not enough to warrant a 4.*

Rater 5: *Formatting-wise it looks like multiple paragraphs but 2 of them are not real paragraphs. There is no solid organization within the paragraph a little bit. There are too many basic errors to be a 4.*

Rater 6: *I think the organization is weak on the paragraph level and essay level.*

Rater 7: *Lots of mechanical problems. Basic errors. Small vocabulary*

Packet Part 2

## LAT Essays Borderline Example      Score: 4

ID: 71076

30 Minute Essay

Word Count (excluding prompt) 514

805071076

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

According to the topic, I disagree because they have many reasons but I want to choose only three reasons. A zoo has no useful purpose. My three reasons are because, to preserve animals who are going to disappear. Second, to give people a big occasion to go to see animals that they never see in their life. Third, to protect their race or their groups.

First, nowadays they have a lot of animals in the world which started to disappear around the globe because of many things. The main things I learned long time ago people from one generation to another generation they start to kill them and to eat it. According to my own opinion, God says in the Bible what animal you need to eat you know I believe in religious rules. But people from the twenty to the twenty-one century they start to kill animals that we do not suppose to eat like birds, dogs, bats, etc. to much to mention. A zoo is really great because they create this place in a lot of countries to save them to make them multiply without getting hunted by people who are crazy. I really thank to have a zoo to preserve them because animals are really important they are in life and it's really important they can feel in security, this is the main purpose of a zoo to put them in a great reserve and make them feel home in nature.

Second, a zoo was also created about the 1800s to give people a chance to visit animals without getting aggression from them. Before the 1800s, people need to go to the savanna or in a forest to observe animals. Most of that time people get aggressive by wild animals and they thought insecure at that time period of their life. People try to invent a zoo so tourism and visitors can visit the main animals who are going to extinct for money and business. At the time people feel more protected about that. People from different cultures really don't know all the animals on the earth but since they create the zoo people are more interested to know about and great moment for them.

Third, animals nowadays are rarely rare because of the global area and natural disasters such as fire in a forest. But the main things which is a global area is air pollution. I want to give an example in the Indian Ocean where I'm from they had about one point five million turtles before (2008) with air pollution and every time it happens they start to disappear, they have plastic bags all around the beach and they disappear because they eat that thinking it's food. That's why they try to put these animals in this kind of place whose name is zoo. People were so smart to create a place like that.

In conclusion, a zoo has a purpose in the life of the animals and people. Because our planet is going down. A zoo is here to give people a chance to see animals and for the animals themselves they can multiply their race and feel secure.

Packet Part 2

## Experienced Rater Feedback

ID: 71076

### **Why could this have been rated a 5?**

*Rater 1: Evidence of organizational markers at essay level (first, sec, third), good length (514), complex grammar and vocabulary attempted (in security, generation)*

*Rater 2: Extensive writing; highly developed paragraphs; able to use language in major time frames*

*Rater 3: There is some evidence of organization.*

*Rater 4: Great quantity. Nice variety and academic expression.*

*Rater 5: The other paragraphs have some more abstract examples but still a personal context feel. Good essay organization and fluency. Complex sentences are attempted.*

*Rater 6: This could have received a 5 because there is strong organization on the essay level.*

*Rater 7: Has organizational markers, varied vocabulary.*

### **Why was this rated a 4?**

*Rater 1: Multiple paragraphs-good organization at paragraph and essay level, context of personal interests/experiences, problems in accuracy-at times writing is unclear but overall meaning is not obscured*

*Rater 2: Errors in basic subject/verb agreement and singular/plural nouns; errors in complete sentences; ideas within paragraphs are not cohesive*

*Rater 3: The vocabulary is varied. The topics are concrete. The student doesn't have a mastery of the past tense.*

*Rater 4: Sufficient length- 5 paragraphs all functioning according to essay structure. Complex sentences generally accurate. Patterned errors in subject-verb agreement. Lacking determiners, tense problems.*

*Rater 5: Solid paragraph organization and basic essay organization but writing in the second paragraphs is in the personal context- has a colloquial tone. There are a lot of basic errors that impede comprehension.*

*Rater 6: The writer speaks about personal experiences and anecdotes. There are also errors on the sentence level.*

*Rater 7: Multiple paragraphs, solid organization*

Packet Part 2

**LAT Essays**  
**Borderline Example      Score: 4**

ID: 61718

30 Minute Essay

Word Count (excluding prompt) 305

407823779

*Prompt: Do you agree or disagree with the following statement? All students should be required to study art and music in secondary school. Use specific reasons to support your answer.*

I agree with the statement, because study music and art during secondary school can be a benefit for the rest of your life, for many reasons like develop different skills, create ideas or own projects to the future, increase and found new talents and habilities.

In the first hand you can develop more skills maybe your hear, voice, creativity, etc. You can be a integrity student with many talents and abilities to share with others, and this is very usefully during secondary school, improving all of your capacities.

Also you can create ideas for your future, like a musical company, be a teacher with your talent, create your own business in the world of the arts. Is very important to recognize this habilities in this time of your life. I you can improve your talents in this age of your life is an incredible chance to starting a professional carrer in this area, you can be an artist with distinction arround the world, also you can be a famous singer with your own cd's and stuff, fans club, videos, movies, tv shows, everything. So is very important create this spaces for the students during secondary school.

Finally you can help other, the music and the art can be an beatiful way to inspire people, because is very powerful channel, with your talent other can be feel better, maybe you can create a smile, emotions, litteraly touch hearts through the art. And this can be by a song, just music, notes, your voice, a picture, paint, etc. By many ways you can inspire the world by your talents and habilities.

There are all of this reasons to support this stament, is an a very good idea start to help the student with this chances during our first years in school. I completly agree with the statement.

Packet Part 2

## Experienced Rater Feedback

ID: 61718

---

### **Why could this have been rated a 5?**

Rater 1: *They used different time frames, but there were inconsistencies between present and simple past. Organizational markers, used more concrete ideas, some grammatical errors in structure*

Rater 2: *Good paragraphs. Essay level organizations grammar errors.*

Rater 3: *Multiple paragraphs with organizational markers, favors concrete ideas, vocabulary is good, but not as good as a 5.*

### **Why was this rated a 4?**

Rater 1: *Several grammar errors, but they don't obscure meaning. Multiple paragraph with topic sentence and examples some essay level organization.*

Rater 2: *Organization at paragraph and some essay level, not all ideas are well explained, vocabulary and errors.*

Rater 3: *Errors/typos*

Packet Part 2

## LAT Essays

### Borderline Example      Score: 5

ID: 31220

30 Minute Essay

Word Count (excluding prompt) 436

133131220

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

#### The Purpose of a Zoo

Many people consider that a zoo has no useful purpose, that is a waste of resources. Other people consider the opposite that a zoo is good, and useful for the people. Having a zoo in a city can be really useful if the people consider the advantages of it. In many developed countries zoo are a great attraction to tourist, and helps to improve the economy of a city. A zoo have a useful purpose because it helps people to learn more about animals and nature, It can help the economy of a city, and it help the people to have a more green attitude.

First, many people that likes nature and animals will visit a zoo. In many contries around the world zoo's help people to learn more about nature and the animals. If someone have the opportunity to go to a zoo this person will learn more about the animals that live in his/her country. Also will learn the correct way to take care of those animals, and how to protect them. All the zoo's around the world have vegetation, so people can learn about plants, vegetation, and the benefits of some of this plants to our life. People that have had the opportunity to went to a zoo says that it's an amazing experience and that you will always learn something new and intersting.

Second, having a zoo in in a city helps to improve the economy of it (the city). For example in Australia, there are many zoo's. Each of those zoo's helps to improve the economy of the city. Many people travel across the world just to see the rare and exotic animals in Australia. as a result of this the city will have good profits in various sector such as the zoo, hotels, restaurants, etc. The tourism rate will increase. This will have an important effect on the economy of the city. Probably because of this more people will have a job, a better salary, and in general a better quality of life.

Finally, people that has visited a zoo have learn how to protect the enviroment, so this people will have a green attitud. A green attitud is what the world needs. People need to protect the enviroment, and they can learn it through visiting a zoo.

A zoo have various useful purposes such as helping people to learn more about animals and nature, helping the economy of a city, and helping the people to have a more green attitude. I personal believe that a zoo can have more positive effects rather than negative.

## Experienced Rater Feedback

ID: 31220

### **Why could this have been rated a 4?**

Rater 1: *basic grammar errors- subject/verb agreement*

Rater 2: *Some examples and explanations are not clear*

Rater 3: *The vocabulary could be more varied. There are more errors in accuracy than a level 5.*

Rater 4: *Incomplete towards the end some problematic errors in verb agreement.*

Rater 5: *Some grammar and spelling errors, capitalization errors, some are in basic sentences.*

Rater 6: *Favors simple vocab, some subject-verb agreement problems.*

### **Why was this rated a 5?**

Rater 1: *Evidence of organizational markers at the essay level-strong paragraph/essay organization, complex grammar and vocabulary, some abstract topics*

Rater 2: *Examples in paragraphs are general rather than personal; multiple paragraphs with essay level organization; some paragraphs have extended development*

Rater 3: *The student uses concrete and some abstract topics, but the abstract is not as clear. The student has multiple paragraphs with evidence of organization in paragraphs and the essay.*

Rater 4: *Nice variety of sentences structures. Solid thesis than mirrors topic sentences of body paragraphs. Body paragraphs well supported with specific details.*

Rater 5: *The tone is abstract and impersonal. Great fluency and organization. Paragraphs are solid and almost all ideas/details are supported. Complex sentences, title.*

Rater 6: *Strong organization on the essay level with appropriate organizational markers. Writer able to write about abstract topics.*

Rater 7: *Paragraphs used good organization markers, academic vocabulary higher.*

LAT Essays  
Borderline Example    Score: 6

ID: 48419

30 Minute Essay

Word Count (excluding prompt) 518

217348419

Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.

When I was a child I loved going to the zoo; it was always an amazing experience to see all of the animals. As when I was a child, I still enjoy going to the zoo. But is there any purpose on having children going to the zoo? Zoos have useful puposes. Zoos allow children to know the different species of animals around the world, it creates a love for the animals that can be used as a factor to protect them, and it also helps children to see how much we need to safe the environment so that animals can live.

First of all, knowledge can be found in the zoo. Going to the zoo gives children the opportunity to learn while they are having fun. It allows them to have an interactive learning. This way to learn, makes a more profound impact on children's minds. Therefore, they learn more about the animals while they are at a zoo that just in a classroom. Also, a zoo gives children he experience to get to know many animals that come from all over the world. It is not always possible to travel around the world and see the animals each place has. But in a zoo, since there are many species of animals from all over the world, children can learn about the animals that in other situation they wouldn't see or even hear about them.

Second, when children get to see and learn of the animals that are endangered, they develop a love for those animals. That love, can be later on used as a way to engage children into protecting the animals that are endangered. A filmmaker once said: "we protect what we love". That is a true statement when it comes to protectim the species that are endangered. The love that is created when children are at a zoo can be used to protect animals. Children are a great source of power. They influence the family they live in. Therefore, if children want to protect the endangered species they are more likely to influence their families into doing the same.

Third, in the zoo, children can learn how dependent are the animals from humans since humans are destroying their environment. Many times it is hard to communicate people in general why animals are endangered. Zoos usually have endangered species so that they can protect them; thus, children can learn about the animal that are in there and how they can help to save the environment they live in.

In conclusion, zoos are not only a nice and amazing experience for children, but a way to educate them. Education is everywhere in a zoo, everywhere you look there is a new animal you can learn about. Therefore, children learn about animals from all over the world that in other situations they wouldn't know. Also, they learn to love the animals and how to protect the environment they live in. Zoos are useful, they are a great tool for educators and environment protectors and we as a society should encourage the creation of more and better zoos.

## Experienced Rater Feedback

ID: 48419

### Why could this have been rated a 7?

Rater 1: *At times more like oral discourse, grammatical errors*

Rater 2: *Vocabulary is very appropriate but a little informal; sometimes writing is more personal than it should be; repetition in sentence structure and word choice*

Rater 3: *There are some errors in accuracy with the more difficult structures.*

Rater 4: *Errors are very contained. All ideas seem full and well developed.*

Rater 5: *There are sometimes too many simple sentences, especially next to each other. A few small basic errors.*

Rater 6: *Some punctuation/ grammar errors.*

Rater 7: *Low use of academic vocabulary, complexity low.*

### Why was this rated a 6?

Rater 1: *multiple paragraphs- clear organization, vocabulary, cohesive devices, grammatical complexity*

Rater 2: *Able to use language in detail in various time frames; multiple, highly developed paragraphs with clear essay organization; uses abstract and concrete examples.*

Rater 3: *There are multiple paragraphs that have clear organization. The essay is also well organized. The student uses abstract and concrete ideas.*

Rater 4: *Complete from top to bottom, cohesive devices. Academic in tone. Clean in execution.*

Rater 5: *Vocabulary, language and content is abstract and impersonal tone. Strong paragraphs and essay organization. Successful use of complex and compound sentences.*

Rater 6: *Multiple paragraph essay organization. Message is pragmatically accurate. Control over word order although there are some errors.*

Rater 7: *Multiple paragraphs- clear organization, pragmatic accuracy variety of some tenses.*

Practice

## LAT Essays

ID: 02112

30 Minute Essay

Word Count (excluding prompt) 351

189702112

*Prompt: Do you agree or disagree with the following statement? All students should be required to study art and music in secondary school. Use specific reasons to support your answer.*

I agree that all students should be required to study art and music in secondary school. I can understand that some people do not agree this statement, but I believe it will give many experiences to students. It is not just wasting time because of several reasons, variety of their lives, good refreshment, and consideration to others.

First, to learn art or music is a good opportunity to earn a good sense or a perception towards things or people. It will encourage students to enhance their lives. For example, I have been to many museums in the world. When I saw painting, I was impressed because I could have seen the paintings which might have been written long time ago. So I can learn things from our ancestors. To learn art or music is not just a art or music. it can also learn history of the world.

In addition, it also a good chance to know that there are so many ways of thinking depend on the person. For instance, we can see difference between two people even if they paint the same thing. We can learn that we should respect each other through learning art or music because there are no correct answer especially the way of feelings. Even if a work of some students look like weird, no one can blame or decide that it is not correct.

Finally, I believe that no one can just keep studying whole a day. To learn art and music can be a good refreshment for students, and they can also learn about consideration or history. For me, art class was important to change my feelings between the classes when I was a high school student. I did not think other things while I was working to paint my work.

Studying art and music is not just like a usual study. It can be a good opportunity for students to know the world. I'm sure that if students stop to study art and music, there will be many selfish people more than now because they could not take art and music class.

Practice

## LAT Essays

ID: 05177

30 Minute Essay

Word Count (excluding prompt) 292

195405177

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

Could you like to live in a jail? So, why do you think that animals could have a good life there? I have a lot of reasons, but I will tell you only three to explain why I disagree with zoos and they are it spend a lot of money, the animals were not created to live in jails and they cannot be health living far to they natural habitat.

The first reason is a zoo spend a lot of money with workers, to get a big and good place and to take care with the animals, etc. This money could be spend with different things that are more useful, like parks, schools, and everything.

Another reason is that the animals were not created to live in jails, they need to live in they natural habitat without noise and a lot of people close to bother them. Could you imagine the sensation that an animal that can fly falls when is locked in a jail? And an animal that could run across the forest and be independent and after is living in the limit of his jail?

Finally, animals that live in a zoo cannot be health at all. They cannot to express their natural extint because they live in a limited space. Usually, zoos are in big cities, this places have a lot of noise and it is totally different that place that wild animals live.

Lock animals in a zoo is not a good option to get fun, people could find another way to get it. If you would not like to live in a jail, why do you need to agree that animals can do it? Everyone have your natural place to live and never would get out there.

Practice

## LAT Essays

ID: 10726

30 Minute Essay

Word Count (excluding prompt) 167

744010726

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

I disagree that a zoo has no useful purpose. For my opinion, zoo is a good place for people to learn about many kinds of animals which we can't pet them.

It is also a nice place for families to go to visit, especially for a family has kids. Moreover, it is a place for children's schools to go for visit and picnic.

First, Animals are very interesting. There are variety of animals in the world. But, we don't have opportunity to a lot of types of animals around us. Such as monkeys, elephants, dolphins and so on. Therefore, we should have a zoo to see different animals and we can learn about different animals from a zoo.

Second, most of kids like to go to the zoo to see animals with their parents. Parents can show and teach their kids about those animals they see from pictures.

Third, most elementary school students study about different kinds of animals, and animals' habits, features. It is better to have opportunity to see real animals a

Practice

## LAT Essays

ID: 13554

30 Minute Essay

Word Count (excluding prompt) 169

218713554

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

We all crazy about the Olympics, and it is starting at London. Some countries financially support their athletes, while some countries don't give money to their athletes. I think that countries should financially support thir athletes who compete in the Olympics for two significant reasons.

The government should support our athletes because they fight for our counrty. When A good athlet is going to attend a international activety, we should do everything we can to help them. To go on a international games need a lot of supports, so they can accomplish these race.

In addition to supporting our athletes to win the games, we need to make sure athleltes have everything they need. For example, a athlelt need coaches, money...etc. To train a good athlet is not easy, and that is why our covern

In conclusion, I agree that our government should helps our athletes. When athletes feel the support from their country, they become more willing to fight for our country. Therefore, they can win the sports.

Practice

## LAT Essays

ID: 17398

30 Minute Essay

Word Count (excluding prompt) 322

848317398

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

What happen if you are the most talent athlete but you dont have the resourse to train like other athletes? that it unfair, righth? Well this is the rality of many athletes that compete in the olimpics. Exists thousand athletes when his countries dont financial to them, and they cant compete in the olimpics. I think that is incorrect, and i goin to explain why.

First, for me will be exists in the olimpics comite, a way to rule all the contries and they sure that each country have the sames oportunities for his athletes. If there is a country that cant to financial his atletes, so this country cant participate in the olimpics. So, for me it is so important to rule the diferents countries to help his atlethes.

I feel that we dont understand the benefics that give the olimpics to bring to the world. It is a beautiful experiencie to give peace and when to now diferents cultur. So i think that how today we have a globalizate world, and each country have a diferents organization like the ONU or OEA, this organization can obly to every country to participate in the olimpics for the beneficts that give.

I belive that every country have enought resurces to help the athletes. In the every buget conuntry can be a area whit the name OLIMPICS and destinate money for the buget to help the diferent athletes. And the diferent mundial organizations can be ruler this thing.

For this reason, for me, i think that every country have the responsibility for his own athletes. It is not possible that other sourse like Nike or McDonalds make the job that correspond to every country. When a Athletes won, the country that he represent won, and all the world can see that. For that reason, for the prestige of every country it is so important that each country financially and supports his own athletes.

Practice

## LAT Essays

ID: 17865

30 Minute Essay

Word Count (excluding prompt) 273

996117865

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

A zoo is very important in a country or anywhere else because it shows how the animal live in the environment. I disagree with the statement a zoo has no useful purpose for three reasons, showing importance of the animals, fun or hobby, and financial.

First, zoo is important by showing people the importance of the animal. I remember in my country, I kill an animal that I have never seen befor. I did not have any information about it. My parents though it was a superstition. Last month I went to a zoo in Carlfontia, I saw an animal looks like the same tha I killed. The zookeeper said to me that type of animals is very important for the nature because they eat some hawful insect that can distroy the the plant or farm.

Second, zoo is some what necessary in a country or city. There are some people who have as habit going to the zoo to see animals. If here is no zoo, these people will not feel comforatble about their habit.

Finally, zoo is a good activity that can help the economy to flourish or increase. There are many people who work for a zoo. this mean there are some who hunting animal and bring them to the zoo. Also, there are some pople who take care of these animals after hunting. By the way, there are many job opennings in a zoo because there are always things to do.

To conclude, for these three major reasons, importance of the animals, hobby, and economic I say yes, I disagree with the statement a zoo has no useful prupose.

Practice

## LAT Essays

ID: 20487

30 Minute Essay

Word Count (excluding prompt) 424

706920487

*Prompt: Do you agree or disagree with the following statement? It is better to be a member of a group than to be the leader of a group. Use details and examples to support your opinion.*

The current president of Mexico, Felipe Calderon is one of the richest person in all the country. Nowadays, being a leader means to have more money and have more responsibilities and experiences that other people cannot have. Being a leader has a lot of positive consequences in order to be a superior human being. Additionally, being a leader is not an easy job for everyone because it needs a lot of preparation and education. Individuals who are leaders have more possibilities to have a better and prosperous future and to develop skills that a simple member of a group cannot develop.

Nowadays, there are many leaders in this society who have increased possibilities of having a brilliant future. For example, the richest person all over the world, Carlos Slim. He is the major example of leadership because without his companies millions of people in Mexico could not work. Carlos Slim is the manager of TELMEX, the second biggest industry in this country, and relevant companies like Sanborns and Sears as well. Additionally, it was not easy for him to create all these companies because it took to him many years and hard work to create them. Just like Slim, there are more leaders who are increased the value of their life. For example, Bill Gates the second richest man. He started studying at Harvard and now he is the chief of Microsoft. In addition, these individuals have created the best future for their families. Their children are going to have the best education and accebilities that other children cannot have. Another important point about being leaders is that them can increase and develop more skills than a normal member of a group cannot increase. For example, leaders can increase their manager abilities because they have a lot of responsibilities. Leaders of this society can develop leadership abilities in order to help people. For example, Bill Gates has created non-profit organizations in order to help poor individuals from Africa or America. He has the vision of helping and caring human beings. Other abilities that leaders can develop are honesty, and integrity.

Finally, individuals of this society need to know that life is more than being a part of something. Life means to develop new skills, and being a leader is going to help to increase them. Also, life means to learn new things and leadership is a superior way to do it. For example, leaders like presidents or major chiefs of companies know how to use their power for the greater well of the society.

Practice

## LAT Essays

ID: 25722

30 Minute Essay

Word Count (excluding prompt) 339

374625722

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

To be athletes and represent your country attend Olympics is very special for all citizens. The government should support the financial for athletes because it increase the citizen unity philosophy, the other countries will more respect your countries, and the government and the corporate can support athletes together.

First, the government support the athletes can increase unity philosophy of citizen. Because of the unity, people would donate the money for the athletes who represent the country. It also cause many people willing invest athletes. The nation also can have change work with government. For example, when government have financial aid for the athletes, the citizens will have chance to know those athletes, and citizens and government will have more unity together.

Second, Other countries would more respect your country. In 2010, Olympics in Londong, Taiwan's flag was taken out after hang out three day. If our government can support our athletes, and the nation can unite together, may be other countries can more respect you. Also, the country will know Taiwan is the belong to China any more. Becuase of the financial aid, people will know the economic in Taiwan is very good. Indeed, if the government can support the financial aid for the athletes, other countries will know the economic of this country is very good.

Finally, if the government and the corporate can combine together, it will be a good example and investment for all of citizens. Because of this, the athletes can proud of their country and represent for the country. If athletes do not have financial aid form their own country, the athletes will go other countries who can accept them that will cause no one want to be an athletes for the country. Also, people will not have chance to service their own country. Therefore, if the government and corportate can work together, the athletes will feel respec and proud of their own country.

In conclusion, the government should support the financial aid for the athletes. It can let others to know our country.

Practice

## LAT Essays

ID: 27576

30 Minute Essay

Word Count (excluding prompt) 436

055827576

*Prompt: You are the freshman counselor at a university. You are writing an essay to all incoming freshman that answers the following question: What are the 3 most important subjects that students should study in college to prepare them for the future? Why?*

The college life helps you have much better lives by providing you with many academic subjects that will be precious in your future career. As you know, university gives you many options to be exposed to various academic world. Some students are going to focus on just their major subjects. Other students may study interesting subjects that they can enjoy learning. However, in my opinion, learning the three critical subjects - English, foreign language, and communication skills - are really important for your future career lives. This is because those subjects are basically applicable to every area of your career, and they are essential when you play a role in the society as a part.

First of all, as long as you live in American society, learning English plays a pivotal role in every part. Reading in English is really important when you pursue to study here since there is generally the great amount of reading assignment in university. If you have trouble with reading fast and appropriately, you will have many obstacles to be succeeded. Also, writing skills are the supreme factor for successful lives in America. Writing well means that you can persuade others with your writings that many parts of the society require. Those two reasons are that I recommend that you study English.

On top of that, I am absolutely convinced that learning foreign language will take you into successful life in the future. This modern society is not limited with just your own countries. There are numerous chances to extend your career and requirements since it is a global society. Expanding your linguistic limitations means that you have more opportunities to be succeeded in the society than others who do not have any extra language abilities.

Lastly, learning communication skills is undisputedly crucial if you are preparing for your future. The society consists of various and different people. You can encounter people who are not matched your characteristics, and you will have many difficulties to communicate with them. Once you jump into the society, you no longer can act with your own thoughts, you should frequently compromise and discuss things with others. If you learn communication skills when you are college students, you will be better for treating and interacting with others that make you communicate with others appropriately in the future.

To conclude, you have many options to study once you become college students. As college life is really important to prepare your successful future, you are required to contemplate when you choose subjects that you will study. I strongly encourage you to take English, foreign languages, and communication skills.

Practice

## LAT Essays

ID: 29066

30 Minute Essay

Word Count (excluding prompt) 145

625229066

*Prompt: Do you agree or disagree with the following statement? A zoo has no useful purpose. Use specific reasons and examples to explain your answer.*

Zoo has many animals. people are go to zoo with their family. people like animals. They have interesting. some children like go to there. some animals are danger. they are eat meat. They are danger to people. some animals like paint . They are small or big. they live in cage. also they live each cage. They live same kind animals. some people watch them front cage. people give some food to them. animal like that. but ,we can not give many food , because it is not good to animals. I think some kid like monkey. I think they look like person. so, I think they are purple. They can do show. They have trainer. They understand their trainer language. they have practice. they can do their show. people like their show. many people cross animals. if we want see them. we can go to zoo.

Practice

## LAT Essays

ID: 29541

30 Minute Essay

Word Count (excluding prompt) 337

987629541

*Prompt: Do you agree or disagree with the following statement? All students should be required to study art and music in secondary school. Use specific reasons to support your answer.*

Arts have been always taught in every time and through generations. People become better when they know arts. In history towns always have had arts, and nowadays we can know those civilizations because of rests of their arts. In this century, it is the same as old times. children need to be taught in types of art in order to help them to develop skills and to help them to open doors at future times. Although, some people do not like arts, it is needed to teach art to improve life style and to increase possibilities to be hired.

Arts must be taught to improve life style. Happiness can be found in arts everytime. For example, a 15 years old korean girl Chin Chin use to play piano in christmas day, so she helps to make that enviroment happier, and her family loves her to play the piano. Therefore, arts can help people to find happiness every palce and situation. Another reason is, arts improve life stile because education can be shown by art skills, and studied people have more opportunities to open laboral fields. They are more able to find different kinds of work.

Arts increase possibilities to be hired. People with more abilities than other people are chosen to perform activities in a work. If chances to find a job are closed, it is possible to get a job because your arts abilities. For example, Edith Maria del Carmen a mexican citizen had a har time finding job; However she was able to find a job teaching drawing because she was studing drawing before finishing school. Therefore, due to the Edith's example it is possible to undertood the helpfulness of having arts skills. In addition, arts skills help you to do a better job, even when those art skills are not well used For instead, Carlos from USA was working as a medical assistant, but because of his abilities in drawing, his payment increase beacuse he was helping doctors to draw organs, and so forth.

Practice

## LAT Essays

ID: 39493

30 Minute Essay

Word Count (excluding prompt) 377

987439493

*Prompt: You are the freshman counselor at a university. You are writing an essay to all incoming freshman that answers the following question: What are the 3 most important subjects that students should study in college to prepare them for the future? Why?*

Being a freshman could be very hard for some people because university's life is a new world for every student. Because of that, freshman can find counselors at a university who can help them to be successful. In order to be prepared for the future, freshman should study in the college the following subjects.

First, freshman ought to study foreign languages to be more competitive to find a good job in the future. Because of the globalization, the world is more competitive now. In consequence, every student has to develop strong communication skills in order to be more attractive for different kinds of companies. Companies want to establish operation in different parts in the world, and students have to be prepared to communicate in other languages with the purpose of helping companies with their goals of catching new markets. Knowing this subject in the college, freshman will be able more competitive at the moment to find a high position in any job in the future.

Second, freshman should study marketing. This subject will teach them how to promote and sell any kind of product. Even though, they do not develop this career in the future, knowing marketing strategies can help them to promote and sell their ideas, and they also know how to persuade others. In this way, freshman will cultivate a strong personality that will help to show their clear thoughts at any moment. Studying marketing is an excellent recourse to persuade others about their ideas.

Finally, freshman ought to take time to study business. This world is a big company where everybody plays an important part in it. If freshman study business, they will be more capable to understand and comprehend how to take an important place in this world. They can develop networks, and relationships that could help them to earn more profit in their business. Knowing the strategies of business as a subject, freshman can take some advantages in their future.

All the subjects in the university are important, but if any freshman wants to be very successful in their future, they will have to study foreign languages, marketing and business. These three subjects will teach them how to be more competitive in this difficult world. These are my advice as a counselor.

Practice

## LAT Essays

ID: 49343

30 Minute Essay

Word Count (excluding prompt) 393

263049343

*Prompt: Do you agree or disagree with the following statement? It is better to be a member of a group than to be the leader of a group. Use details and examples to support your opinion.*

Be the Leader of a Group

Be a member of a group is easier than be a leader, however be a leader is better. Although is more difficult, when you are a leader of the group, the group do what you thing is better, you recieve more credits when the group achieve the goal and bigs companies prefer this kind of professional.

First, one of the responsibility of the leader is make decision. A good leader usually listen all the group before to make important decisions, but the person that will make the decision will be the leader. This is a good point, because I like to listen different opinions before to make a decision, but I like to make my decision. When you are a leader this happen, and it is one of the point that explain why be a leader is better then be a member of a group.

Second, people work in group because they want to achieve a goal. In my opinion, all group have some goals to achieve. To achieve this goal all members is necessary and important, however the leader of the group recieve more credits when the goup finish the goal. This happen because the leader needed make important decisions, control the plans, and usually spend more time then others thinking about the problems. In this case, the leader will work more, but will receive more credits. Then, is another good point to be a leader.

Finally, if you think in your career you will prefer be a leader too. The biggest and best companies prefer the people with leader skills. Because of that the best jobs, with good salaries need this skill. Need a person that know how be a leader, how motivate people, how make decisons and other characteristics that only a learder have. If you think to have a career with a good future, you need to know that in this case be a leader is the best option.

As a conclusion, if I could choose between be a member of a group or be the leader of the group I am sure what I will choose. Although if more difficult, I think that if we choose be a leader we can have more pleasure. Make decisions, recieve credits and be a value professional to the companies are excellent examples that is true.

Practice

## LAT Essays

ID: 50685

30 Minute Essay

Word Count (excluding prompt) 359

082550685

*Prompt: Do you agree or disagree with the following statement? All students should be required to study art and music in secondary school. Use specific reasons to support your answer.*

Many people think art and music are important. Sometimes we are going to art museum, or music concert. Also, we see many visual arts in the subway, or other places, and we listen popular songs. We use art and music things after we graduated to school. To support my idea, I have three points: easy to school subject, useful, and they will be clever.

First, students use art and music every day. When art class or music class, they easy to understand. For example, when I was a middle school student, I learned how to play the piano. I did not like play the piano. One day, our music teacher explained about tempo, and how can us keep the tempo. Other friends, who were did not learn music, they were very hard to understand. However, I already learn in my piano class, so I could very well, and I got a good score.

Second, art and music always use when we graduate school. We can see many visual arts, and we can listen song every where. For instance, when I was a high school student, I learned art. I liked art, so I could learn very fun. After I graduated high school, I did not know using my art skills. When I taught to children in my church, I always used visual resoureces. One day, I made visual picture myself, and children liked it. I was proud of myself.

Third, art and music make students to bright. When students learn art or music, they can increase their thinking. They can creative thinking. For example, my brother leared piano when he was young. He learned piano almost 10years. He can play the piano very well. He can think very unusal things. Also, he did study very well. He always got good scores.

As we have seen, if students learn art and music, it is good to students. They will get good scores in art, or music test, and they can use their skills or information after graduated school. Also, they will be smart. Some people say art and music education are very expensive. However, there are good to student's future.

Practice

## LAT Essays

ID: 57252

30 Minute Essay

Word Count (excluding prompt) 207

039457252

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

I know some countries financially support their athletes in the Olympics, but other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsor, such as Nike, McDonald's, Addida, Puma, etc. I agree this good way. It is better for athletes. My country so many good athletes, and so many big company support them. Let's they can do well. My goventment always support their athletes in the Olympics or another sports.

Other countries don't give money to their althletes. They have to pay for their own training or get corporate sponsors, such as big company, or famous company. It's good idea too. My country some one athlete. He is poor athlete. Some day I watched him on TV. He said " He want to attend one sport but he not more money. So he want to goventment or big company support him. After he went to sport, he got a lot money. When he own some sport, he was getting married three years ago. It's good news.

So, I opinion every body enjoy for Olympics. You have to support your country athletes. If you are famous company of boss support your country athletes more. Let evey body enjoy Olympics

Practice

## LAT Essays

ID: 57753

30 Minute Essay

Word Count (excluding prompt) 492

424357753

*Prompt: Do you agree or disagree with the following statement? It is better to be a member of a group than to be the leader of a group. Use details and examples to support your opinion.*

In a group always there are going to be a leader that try to guide the group in their own activities. Oftenly with every position into the group have some pros and counters, that it will help us to know what is better. Next we are going to analyse the members of the group, what are their duties or benefits. That way reader should be able to know what are their rol in their own group.

Firstly, the fundamental leader. Someone that is present in all human organizations, such as: Churches, government, companies, criminals, etc. If it doesn't matter what are they doing, if is good or bad, everyone have a different perception of the life. Being a leader of a group, is something natural, one of the members should be leader and should guide to the group. With that charge, become a responsibility of take care of the members of the group. In addition being a leader isn't easy, because members should choose the leader, and they have to respect that position and the person who are going to guide the group.

Second, being a member is important in several ways, because you play a role into the group. A member do not have the same responsabilites than a leader, so they are not responsible for acts that do the group, also it depends of the organization or group, but probably they do not have the same respect that others feel for their leader. In addition, benefits are not going to be the same as a leader. Probably you will have some of this but not all of them. So in other words is very important a member, but at the same time is not important.

In conclusion, a leader should be responsable for their his or her group, they should take care of every member of group also being secure that everyone keep respect, admiration and confidence for him or her. Because, members don't have the same privileges and they are not so important such as leader. But at the same time they are fundamental for the group, because without members is not a group. One of the best examples that I know is in Venezuela, the actual president of Venezuela was a soldier, and have a high range, but he was working to derrocate the president and introduce his ideas to the country. So he did it, he derrocate the president, knowing so well how the country works and getting respect of several people that in the future will help him. Actually, he has fourteen years in the presidence, and he is respected but several people in his country, in addition he introduce the socialism to Venezuela. Some things are good such as help to the poor people or something like that, but on the other hand he is so corrupt and have some cases of violence to his oppositors, also the economy of Venezuela are decreasing very fast.

Practice

## LAT Essays

ID: 62695

30 Minute Essay

Word Count (excluding prompt) 327

180262695

*Prompt: You are the freshman counselor at a university. You are writing an essay to all incoming freshman that answers the following question: What are the 3 most important subjects that students should study in college to prepare them for the future? Why?*

Welcome to choose Biology Chemistry as your major. I am your freshman counselor, Min-Heng Chen. There are a lot of subjects you will choose in the next four years, but there are three most important subjects you should work extra hard on. They are Experimental Chemistry, Biology Chemistry, and English.

First, the Experimental Chemistry is the most important to students who study Chemistry at any fields. Experimental Chemistry is the extend of Basic Chemistry, and the applications of all fields of Chemistry. It is a subject about every instruments of chemistry, such as AA and GCMass. If you do not expert at Experimental Chemistry, you will not easily operate the instruments to finish your tests and experiments. Most people will think the Experimental Chemistry is easy because it talks a lot about how to use the instruments and reffers little principals, however, the principals is the most important thing you should remember.

Next subject you should working on is Biology Chemistry. Biology Chemistry which is the major you take, and it is very hard for most students because Biology Chemistry is the combination of Biology and Chemistry, and it is very different from both of them.

Finally, English is also an important subject. You may doubt it is a joke, but actually I am not joking about it. English is the most popular language now and before. When you study all of the subjects of this major, the context book you will use, the paper you need to read, the names of the chemicals, ect., are all written by English. Plus, you have to pubilsh your paper some day, and guess what, you have to write it in English as well.

All of the subjects you take is important and worth to do the efforts to working on them, and please spend your time wisely, and try your best on studying Experimental Chemistry, Biology Chemistry, and English. It will prepare you perform extrodinarily outstanding in the future.

Practice

## LAT Essays

ID: 64941

30 Minute Essay

Word Count (excluding prompt) 116

969864941

*Prompt: Do you agree or disagree with the following statement? All students should be required to study art and music in secondary school. Use specific reasons to support your answer.*

i am not agree to fonce people because for my when you fonce people to do someting you are like using your will and for me taht is not feare  
the thing is that i now how those it feels to be like force to do someting i dont like that fealing none of us like that so if the y want to required to all student to take that clases and be like force to take music and art to my is not fear and am not saing that art and music is not good but i just now that everithing has to be do whit you agree and someting that you like thats all

Practice

## LAT Essays

ID: 68121

30 Minute Essay

Word Count (excluding prompt) 424

296168121

*Prompt: Do you agree or disagree with the following statement? All drugs should be legal for people over 18. Use details and examples to support your opinion.*

In many countries the government is trying to combat the use of drugs. Many people believe that drugs are an easy way to forget or solve problems. In fact, when people are addicted to drugs they can face many problems in their lives. In today's world it is easy to find people using drugs in many public places and unfortunately is being a normal behavior around the world. Even though drugs have benefits to help people with diseases, for example, could be a big world problem if they legalize the use of drugs over 18. Because people at this age is very young they could cause many problems to their families, to their lives and to their government.

The first reason that the government should not legalize the use of drugs over 18 is because people could have many problems in their families. It is common in many families to hide the person who are addicted to any drugs. For example, people that have problems with drugs are able to sell their parents stuffs just to have enough money to buy their drugs. It is a fact that they can destroy their families because they could bring the addictive substance to their home and certainly problems will come.

The second reason is because people who have this kind of problem do not care about their health. In many cases they just use drugs to feel better or because they are trying to satisfied their own desires. Many people who have this problem damage their lives through the drugs and could have other disease such as aids, cancer and others.

The last reason is because they could bring many government problems. In places that people have addiction they do not care about their jobs, in almost cases they lose their jobs. Drugs is a easy way to earn money and if the government allow people use drugs over 18 may be a good reason to invest more money and damage the economy of a state or country.

In conclusion, the drugs should not be legal for people over 18. The problems that families and the government could face could be extremely dangerous because they do not know the bad consequences. In addition, people are able to live in a better world if the drugs are controlled. Families could live with more peace and harmony, people could be healthy and the government might be protected from economy crises. Because there are countries who are legalizing the use of drugs, people should be extremely aware about them.

Practice

## LAT Essays

ID: 99846

30 Minute Essay

Word Count (excluding prompt) 111

534299846

*Prompt: Some countries financially support their athletes who compete in the Olympics. Other countries don't give money to their athletes, so they have to pay for their own training or get corporate sponsors, such as Nike or McDonald's. Which way do you think is better? Give specific reasons and examples to support your opinion.*

I think so is better some countries financially support their athletes because is very important for your countries. If my country patrocination help because with support for athletes the peoples have more time for work, for example long time for trainnes, little by little their athletes very happy. Maybe the peoples your countries like in special for gold medal. But a more important is hope the governaments guess a little after. Because a population like the Olympics but not is all world opportunities. Maybe more important opportunities the peoples. I think important for peoples incentive and elevation your contries an special their governants. Conclusion, maybe is very important opportunities for people.

## Appendix B

Handout 1: Benchmark Training

### Writing Rater Practice Form

Name \_\_\_\_\_

Use the Writing Rater Training Binder to complete this form.

Page	ID	Rating (circle the score)	Evidence for Rating (List at least three things that helped you reach your decision. Use sentences or phrases from the rubric wherever possible to support your decisions.)
34	13554	0 1 2 3 4 5 6 7	
35	17398	0 1 2 3 4 5 6 7	
38	25722	0 1 2 3 4 5 6 7	
39	27576	0 1 2 3 4 5 6 7	
41	29541	0 1 2 3 4 5 6 7	
42	39493	0 1 2 3 4 5 6 7	
43	49343	0 1 2 3 4 5 6 7	
44	50685	0 1 2 3 4 5 6 7	
48	64941	0 1 2 3 4 5 6 7	
50	99846	0 1 2 3 4 5 6 7	

## Handout 1: Benchmark Training

**Writing Rater Practice Form**

Name \_\_\_\_\_

Use the Writing Rater Training Binder to complete this form.

Page	ID	Rating (circle the score)	Evidence for Rating (List at least three things that helped you reach your decision. Use sentences or phrases from the rubric wherever possible to support your decisions.)
34	13554	0 1 2 <b>3</b> 4 5 6 7	Weak organization but multiple paragraphs. Better sentence structures than a 2. Small vocabulary range. Compound complex sentences. Conversational style writing. Present tense. Basic errors.
35	17398	0 1 2 <b>3</b> 4 5 6 7	Organization is weak with multiple paragraphs. Small range of correct vocabulary. Some success with compound and complex sentences. Basic errors in grammar, word choice, punctuation and spelling
38	25722	0 1 2 3 <b>4</b> 5 6 7	Evidence of multi-paragraph essay organization. Clear paragraph markers (first, second, finally). Concrete topics and examples are given. Grammatical accuracy and spelling errors frequent, but did not completely hinder meaning.
39	27576	0 1 2 3 4 5 <b>6</b> 7	Multiple paragraph, clear organization. Thesis and topic sentences present with various sentence structures. Not perfect but doesn't distract from meaning. Vocabulary is very high.
41	29541	0 1 2 <b>3</b> 4 5 6 7	Multiple paragraphs with basic organization. Almost all simple sentences, the majority are short. Uses concrete examples without completely integrating or relating them, but in general makes efforts to be abstract. Vocabulary is more than basic and predictable. Basic errors in grammar and spelling.
42	39493	0 1 2 3 4 <b>5</b> 6 7	Clear organization on essay level with multiple paragraph markers. Good concrete ideas with some attempt at abstract ideas. High vocabulary. Mostly correct grammar. Attempted complex sentences.
43	49343	0 1 2 3 <b>4</b> 5 6 7	Good organization of multiple paragraphs with a clear purpose for each paragraph related to the topic and transitions. Attempts to be abstract but brings topic back to personal experiences (many references to 'I'). Shorter sentences but a variety of sentence types. Accuracy errors, but meaning is clear.
44	50685	0 1 2 3 4 <b>5</b> 6 7	Multiple paragraphs with clear purpose and transitions in each paragraph. Talks about abstract applications but always with concrete examples. Some advanced vocabulary ('visual resources,' 'tempo'), but not enough (overuse of 'good'). Narrates in past and present with simple errors. Looks like oral discourse.
48	64941	0 <b>1</b> 2 3 4 5 6 7	Poor basic sentence level constructions. Some simple sentences. Purpose not completely clear. Mostly present tense. Formulaic language. Many basic errors. Common constructions repeated.
50	99846	0 1 <b>2</b> 3 4 5 6 7	Two simple sentences, attempts complex sentences. No organization. Unclear message. Limited vocabulary, though a few high words like 'opportunity,' 'population,' 'financially.'

Handout 2: Borderline Training

**Writing Rater Practice Form**

Name \_\_\_\_\_

Use the Writing Rater Training Binder to complete this form.

Page	ID	Rating (circle the score)	Evidence for Rating (List at least three things that helped you reach your decision. Use sentences or phrases from the rubric wherever possible to support your decisions.)
31	02112	0 1 2 3 4 5 6 7	
32	05177	0 1 2 3 4 5 6 7	
33	10726	0 1 2 3 4 5 6 7	
36	17865	0 1 2 3 4 5 6 7	
37	20487	0 1 2 3 4 5 6 7	
40	29066	0 1 2 3 4 5 6 7	
45	57252	0 1 2 3 4 5 6 7	
46	57753	0 1 2 3 4 5 6 7	
47	62695	0 1 2 3 4 5 6 7	
49	68121	0 1 2 3 4 5 6 7	

## Handout 2: Borderline Training

**Writing Rater Practice Form**

Name \_\_\_\_\_

Use the Writing Rater Training Binder to complete this form.

Page	ID	Rating (circle the score)	Evidence for Rating (List at least three things that helped you reach your decision. Use sentences or phrases from the rubric wherever possible to support your decisions.)
31	02112	0 1 2 3 4 <u>5</u> 6 7	<p><u>Why is it a 5?</u> Multiple paragraphs, organized, mastery of grammar with simple sentences, and abstract ideas are discussed but meaning is sometimes unclear. More complex sentences with errors. Length, transition words.</p> <p><u>Why could it be confused for a 4?</u> The organization is more formulaic. The vocabulary and use of connectors are not as varied as level 5. Some explanations could have been clearer, but overall meaning is fine. Some basic grammar errors. Multiple paragraphs, but organization is weak (thesis and supporting points don't exactly match).</p>
32	05177	0 1 2 <u>3</u> 4 5 6 7	<p><u>Why is it a 3?</u> The student uses paragraphs, but the organization at the essay and paragraph levels is weak. When the stu attempts complex/compound sentences there are errors. Overly conversational. Basic errors in word choice, grammar, spelling, punctuation. Lack of control in linking and connecting sentences.</p> <p><u>Why could it be confused for a 4?</u> Author gives strong examples to support his reasons; paragraph is very cohesive and follows a logical, sequential outline; strong use of appropriate vocabulary and highly accurate. The topics are concrete. There are errors in accuracy, but the meaning is generally not obscured. Paragraphs aren't long, but are organized.</p>
33	10726	0 1 2 <u>3</u> 4 5 6 7	<p><u>Why is it a 3?</u> The paragraph and essay level organization needs improvement. The student uses a small range of vocabulary. Short simple conversational style writing. Related to personal information. Some form of organization that resembles paragraphs. Mastery of simple sentence structure. Organization is weak with multiple paragraphs.</p> <p><u>Why could it be confused for a 2?</u> Student never attempts tense other than present. Most paragraphs are not true paragraphs. Word dropping and spelling issues sometimes create confusion. Lacks correct compound sentences.</p>
36	17865	0 1 2 <u>3</u> 4 5 6 7	<p><u>Why is it a 3?</u> Lack of development on the paragraph level. Solid writing in conversational sentences, but with more formality, accuracy decreases. Second body paragraph and introduction have weak organization. Some tense related errors, and basic errors in grammar, word choice, punctuation and spelling. Short paragraphs with mostly simple sentences.</p> <p><u>Why could it be confused for a 4?</u> Multiple paragraphs with basic organization at essay level. Complex sentences are generally accurate. Student discusses concrete topics. Paragraphs are thematically connected and connectors are used.</p>
37	20487	0 1 2 3 4 5 <u>6</u> 7	<p><u>Why is it a 6?</u> Essay level organization. Good academic vocabulary. Strong development of ideas and examples. Few grammar errors. Uses abstract and concrete language well. A variety of concrete devices, but redundant with some. They address the question in an abstract way without referring to their own experiences. Thesis is well stated abstractly.</p> <p><u>Why could it be confused for a 5?</u> Shows markers of essay organization, but is unfinished. Some grammar and spelling errors. Organization of body paragraphs is not solid.</p>

## Handout 2: Borderline Training

40	29066	0 1 2 3 4 5 6 7	<p><u>Why is it a 1?</u> This student could write simple sentences in the present tense, but lacked other tenses (present progressive). The majority of the sentences were very formulaic. Limited vocabulary and personal context. Short and simple sentences. Weak use of location preposition. No real sequencing of sentences. No attempt to bridge idea between sentences.</p> <p><u>Why could it be confused for a 2?</u> The student attempted to use some connectors. They were also able to write several sentences that could form a paragraph. Close, personal writing. Lacks organization. Length, quantity of distinct words.</p>
45	57252	0 1 2 3 4 5 6 7	<p><u>Why is it a 2?</u> Very little coordination or subordination of ideas. Block repurposing of words and phrases directly from the prompt suggests formulaic or mechanized approach to writing. Not successful beyond present tense. Attempts to give the big picture. Some complex/compound sentences.</p> <p><u>Why could it be confused for a 1?</u> Mostly simple sentences. Fragmented sentences. Sometimes ideas are unclear.</p> <p><u>Why could it be confused for a 3?</u> Length, multiple paragraphs emerging. Attempts at past narration albeit confusing and unsuccessful. Conversational tone.</p>
46	57753	0 1 2 3 4 5 6 7	<p><u>Why is it a 4?</u> Some mastery of different tenses and problems with accuracy, but overall meaning is understood. There were multiple paragraphs with organization. Varied vocabulary, but sometimes word choice was off. Complex sentences successfully completed.</p> <p><u>Why could it be confused for a 3?</u> Paragraphs and semblance of essay form. Good sentence structure and transitions, without grammar interfering too much. However, there is some extra info and paragraphs that aren't super.</p> <p><u>Why could it be confused for a 5?</u> The student talked about some abstract ideas. They used complex sentences and transitions, but with some errors that made meaning a bit unclear. Paragraphs lack support/development. Multiple paragraphs with markers on the essay level.</p>
47	62695	0 1 2 3 4 5 6 7	<p><u>Why is it a 3?</u> Interparagraph structure is weak with confusing sequencing of supporting ideas and some logic breakdown with support present. Structure and meaning predictable. Some evidence of present and future tense, but no past tense mastery. Highly related to personal meaning/conversational. Basic grammar errors. Run-on sentences.</p> <p><u>Why could it be confused for a 4?</u> Quantity of language and use of expected academic signals. Thesis statement, basic essay organization, attempts to extend discourse beyond paragraph level, context of personal experiences, success in some complex sentences.</p>
49	68121	0 1 2 3 4 5 6 7	<p><u>Why is it a 4?</u> Multiple paragraphs, basic essay organization. Organization is just okay at paragraph level. Some examples unclear. Grammatical accuracy. Concrete topic with examples, but explanations could be clearer. Complex sentences are common and often accurate.</p> <p><u>Why could it be confused for a 5?</u> Multiple paragraphs with essay level organizational markers (thesis statement, transitions). Academic vocabulary, favors concrete ideas, but brings in some abstract concepts. Organization isn't formulaic.</p>

## Appendix C

### Handout A (gap)

### Writing Rater Practice Form

Name \_\_\_\_\_ Email \_\_\_\_\_

Use the Writing Rater Training Binder to complete this form.

Page	ID	Rating <small>(circle the score)</small>	Evidence for Rating <small>(List at least three things that helped you reach your decision. Use sentences or phrases from the rubric wherever possible to support your decisions.)</small>
30	02112	0 1 2 3 4 5 6 7	
31	05177	0 1 2 3 4 5 6 7	
32	10726	0 1 2 3 4 5 6 7	
34	17398	0 1 2 3 4 5 6 7	
39	29066	0 1 2 3 4 5 6 7	
40	29541	0 1 2 3 4 5 6 7	
42	49343	0 1 2 3 4 5 6 7	
43	50685	0 1 2 3 4 5 6 7	
45	57753	0 1 2 3 4 5 6 7	
49	99846	0 1 2 3 4 5 6 7	

**Handout A (filled)**

**Writing Rater Practice Form**

Name \_\_\_\_\_

Email \_\_\_\_\_

Use the Writing Rater Training Binder to complete this form.

Page	ID	Established Score	Evidence for Rating <small>(List at least three things that helped you reach your decision. Use sentences or phrases from the rubric wherever possible to support your decisions.)</small>
30	02112	0 1 2 3 4 <b>5</b> 6 7	<u>Why is it a 5?</u> Multiple paragraphs, organized, mastery of grammar with simple sentences, and abstract ideas are discussed but meaning is sometimes unclear. More complex sentences with errors. Length, transition words. <u>Why could it be confused for a 4?</u> The organization is more formulaic. The vocabulary and use of connectors are not as varied as level 5. Some explanations could have been clearer, but overall meaning is fine. Some basic grammar errors. Multiple paragraphs, but organization is weak (thesis and supporting points don't exactly match).
31	05177	0 1 2 <b>3</b> 4 5 6 7	<u>Why is it a 3?</u> The student uses paragraphs, but the organization at the essay and paragraph levels is weak. When the stu attempts complex/compound sentences there are errors. Overly conversational. Basic errors in word choice, grammar, spelling, punctuation. Lack of control in linking and connecting sentences. <u>Why could it be confused for a 4?</u> Author gives strong examples to support his reasons; paragraph is very cohesive and follows a logical, sequential outline; strong use of appropriate vocabulary and highly accurate. The topics are concrete. There are errors in accuracy, but the meaning is generally not obscured. Paragraphs aren't long, but are organized.
32	10726	0 1 2 <b>3</b> 4 5 6 7	<u>Why is it a 3?</u> The paragraph and essay level organization needs improvement. The student uses a small range of vocabulary. Short simple conversational style writing. Related to personal information. Some form of organization that resembles paragraphs. Mastery of simple sentence structure. Organization is weak with multiple paragraphs. <u>Why could it be confused for a 2?</u> Student never attempts tense other than present. Most paragraphs are not true paragraphs. Word dropping and spelling issues sometimes create confusion. Lacks correct compound sentences.
34	17398	0 1 2 <b>3</b> 4 5 6 7	Benchmark
39	29066	0 <b>1</b> 2 3 4 5 6 7	<u>Why is it a 1?</u> This student could write simple sentences in the present tense, but lacked other tenses (present progressive). The majority of the sentences were very formulaic. Limited vocabulary and personal context. Short and simple sentences. Weak use of location preposition. No real sequencing of sentences. No attempt to bridge idea between sentences. <u>Why could it be confused for a 2?</u> The student attempted to use some connectors. They were also able to write several sentences that could form a paragraph. Close, personal writing. Lacks organization. Length, quantity of distinct words.
40	29541	0 1 2 3 4 5 6 <b>7</b>	Benchmark
42	49343	0 1 2 3 <b>4</b> 5 6 7	Benchmark
43	50685	0 1 2 3 4 <b>5</b> 6 7	Benchmark
45	57753	0 1 2 3 <b>4</b> 5 6 7	<u>Why is it a 4?</u> Some mastery of different tenses and problems with accuracy, but overall meaning is understood. There were multiple paragraphs with organization. Varied vocabulary, but sometimes word choice was off. Complex sentences successfully completed. <u>Why could it be confused for a 3?</u> Paragraphs and semblance of essay form. Good sentence structure and transitions, without grammar interfering too much. However, there is some extra info and paragraphs that aren't super. <u>Why could it be confused for a 5?</u> The student talked about some abstract ideas. They used complex sentences and transitions, but with some errors that made meaning a bit unclear. Paragraphs lack support/development. Multiple paragraphs with markers on the essay level.
49	99846	0 1 <b>2</b> 3 4 5 6 7	Benchmark

**Handout B (gap)**

**Writing Rater Practice Form**

Name \_\_\_\_\_

Email \_\_\_\_\_

Use the Writing Rater Training Binder to complete this form.

Page	ID	Rating <small>(circle the score)</small>	Evidence for Rating <small>(List at least three things that helped you reach your decision. Use sentences or phrases from the rubric wherever possible to support your decisions.)</small>
33	13554	0 1 2 3 4 5 6 7	
35	17865	0 1 2 3 4 5 6 7	
36	20487	0 1 2 3 4 5 6 7	
37	25722	0 1 2 3 4 5 6 7	
38	27576	0 1 2 3 4 5 6 7	
41	39493	0 1 2 3 4 5 6 7	
44	57252	0 1 2 3 4 5 6 7	
46	62695	0 1 2 3 4 5 6 7	
47	64941	0 1 2 3 4 5 6 7	
48	68121	0 1 2 3 4 5 6 7	

## Handout B (filled)

### Writing Rater Practice Form

Use the Writing Rater Training Binder to complete this form.

Page	ID	Established Score	Evidence for Rating <small>(List at least three things that helped you reach your decision. Use sentences or phrases from the rubric wherever possible to support your decisions.)</small>
33	13554	0 1 2 <b>3</b> 4 5 6 7	Weak organization but multiple paragraphs. Better sentence structures than a 2. Small vocabulary range. Compound complex sentences. Conversational style writing. Present tense. Basic errors.
35	17865	0 1 2 <b>3</b> 4 5 6 7	<u>Why is it a 3?</u> Lack of development on the paragraph level. Solid writing in conversational sentences, but with more formality, accuracy decreases. Second body paragraph and introduction have weak organization. Some tense related errors, and basic errors in grammar, word choice, punctuation and spelling. Short paragraphs with mostly simple sentences. <u>Why could it be confused for a 4?</u> Multiple paragraphs with basic organization at essay level. Complex sentences are generally accurate. Student discusses concrete topics. Paragraphs are thematically connected and connectors are used.
36	20487	0 1 2 3 4 5 <b>6</b> 7	<u>Why is it a 6?</u> Essay level organization. Good academic vocabulary. Strong development of ideas and examples. Few grammar errors. Uses abstract and concrete language well. A variety of concrete devices, but redundant with some. They address the question in an abstract way without referring to their own experiences. Thesis is well stated abstractly. <u>Why could it be confused for a 5?</u> Shows markers of essay organization, but is unfinished. Some grammar and spelling errors. Organization of body paragraphs is not solid.
37	25722	0 1 2 3 <b>4</b> 5 6 7	Evidence of multi-paragraph essay organization. Clear paragraph markers (first, second, finally). Concrete topics and examples are given. Grammatical accuracy and spelling errors frequent, but did not completely hinder meaning.
38	27576	0 1 2 3 4 5 <b>6</b> 7	Multiple paragraph, clear organization. Thesis and topic sentences present with various sentence structures. Not perfect but doesn't distract from meaning. Vocabulary is very high.
41	39493	0 1 2 3 4 <b>5</b> 6 7	Clear organization on essay level with multiple paragraph markers. Good concrete ideas with some attempt at abstract ideas. High vocabulary. Mostly correct grammar. Attempted complex sentences.
44	57252	0 1 <b>2</b> 3 4 5 6 7	<u>Why is it a 2?</u> Very little coordination or subordination of ideas. Block repurposing of words and phrases directly from the prompt suggests formulaic or mechanized approach to writing. Not successful beyond present tense. Attempts to give the big picture. Some complex/compound sentences. <u>Why could it be confused for a 1?</u> Mostly simple sentences. Fragmented sentences. Sometimes ideas are unclear. <u>Why could it be confused for a 3?</u> Length, multiple paragraphs emerging. Attempts at past narration albeit confusing and unsuccessful. Conversational tone.
46	62695	0 1 2 <b>3</b> 4 5 6 7	<u>Why is it a 3?</u> Interparagraph structure is weak with confusing sequencing of supporting ideas and some logic breakdown with support present. Structure and meaning predictable. Some evidence of present and future tense, but no past tense mastery. Highly related to personal meaning/conversational. Basic grammar errors. Run-on sentences. <u>Why could it be confused for a 4?</u> Quantity of language and use of expected academic signals. Thesis statement, basic essay organization, attempts to extend discourse beyond paragraph level, context of personal experiences, success in some complex sentences.
47	64941	0 <b>1</b> 2 3 4 5 6 7	Poor basic sentence level constructions. Some simple sentences. Purpose not completely clear. Mostly present tense. Formulaic language. Many basic errors. Common constructions repeated.
48	68121	0 1 2 3 <b>4</b> 5 6 7	<u>Why is it a 4?</u> Multiple paragraphs, basic essay organization. Organization is just okay at paragraph level. Some examples unclear. Grammatical accuracy. Concrete topic with examples, but explanations could be clearer. Complex sentences are common and often accurate. <u>Why could it be confused for a 5?</u> Multiple paragraphs with essay level organizational markers (thesis statement, transitions). Academic vocabulary, favors concrete ideas, but brings in some abstract concepts. Organization isn't formulaic.

## Appendix D



---

**Default Question Block**

You are being invited to participate in this research study of a New Writing Rater Training. I am a graduate student at Brigham Young University and I am conducting this research under the supervision of Dr. Norman Evans from the Linguistics and English Language Department. I am interested in finding out about the effectiveness of training new raters of writing by the use of experienced rater feedback.

Your participation in this study will require the completion of the attached survey. This should take approximately 20 minutes of your time. You will not be paid for being in this study. This survey involves minimal risk to you. The benefits, however, may impact society by helping increase knowledge about rater training.

You do not have to be in this study if you do not want to be. You do not have to answer any question that you do not want to answer for any reason. We will be happy to answer any questions you have about this study. If you have further questions about this project or if you have a research-related problem you may contact me, Brittney Greer at bgreer@byu.net or my advisor, Norman Evans at norman\_evans@byu.edu.

If you have any questions about your rights as a research participant you may contact the IRB Administrator at A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu; (801) 422-1461. The IRB is a group of people who review research studies to protect the rights and welfare of research participants.

Pressing "Continue" implies your consent to participate. If you choose to participate, please complete the attached survey by 5pm, Tuesday, March 19, 2013. Thank you!

---

Continue



1.

After the practice assessment of writing that you did as homework in the 611 class and discussed with Brittney Greer, do you feel more or less confident about your abilities to rate timed writing?

- I feel more confident
- I do not feel that my confidence has changed.
- I feel less confident

Please explain why you answered the way you did.

2.

Which set of writing samples was more difficult for you to rate?

- The first set of samples given on Monday as homework (Handout A)
- The second set of samples given on Wednesday as homework (Handout B)

Why do you believe that particular set of samples was more difficult for you?

3.

Were the BENCHMARK samples useful and/or helpful in your rating? Why or why not?

(The benchmark samples were provided in your packet on Monday, and showed clear examples of what a 1, 2, 3 etc on the rubric would look like in essay form)

4.

Were the THRESHOLD samples useful and/or helpful in your rating? Why or why not?

(The threshold samples were provided on Wednesday, and showed examples of writing that could be considered as one of two bordering scores in the rubric)

5.

Was the EXPERIENCED RATER FEEDBACK useful and/or helpful in your rating? Why or why not?

(This was provided on Wednesday, and consisted of comments defending the scores given by experienced raters)

6.

Be honest (please don't tell me what you think I want to hear).

Rate the following five items from high (1) to low (5) according to how much you relied on each during the rating process. (ie. Benchmark samples go first if you relied on them the most in your rating, experienced rater feedback will go last if you did not refer to or rely upon it in your rating.)

Rubric

Benchmark Samples

Threshold Samples

Experienced Rater Feedback

Your own intuition

7.

Following the class discussion with Brittney on Wednesday, did you feel more prepared or less prepared to rate more essays? Why?

8.

In the training, you viewed feedback from experienced raters in two forms.

1. In the training packet Part 2, The rater comments were separated.

Example: Rater 1: good structure. Writer used concrete and abstract examples. Rater 2: While the writer had a good structure on the paragraph level, the essay level organization was lacking... etc.

2. In the pink and yellow handouts, the rater comments had been combined.

Example: Why was this rated a 3? Writer had many basic errors, had fair paragraph level organization, but not enough examples for a 4. Why could this have been rated a 4? Writer used some abstract examples, had a larger vocabulary.

Did you have a preference between these two organizational structures of feedback? Why?

9.

Please write your name. This is ONLY to verify completion. Your responses will not be connected to you, and your anonymity in any future publication will be preserved.