



2017-12-01

Testing the Test: Expanding the Dialogue on Workplace Writing Assessment

Lindsay Elizabeth Tanner
Brigham Young University

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

 Part of the [English Language and Literature Commons](#)

BYU ScholarsArchive Citation

Tanner, Lindsay Elizabeth, "Testing the Test: Expanding the Dialogue on Workplace Writing Assessment" (2017). *All Theses and Dissertations*. 6616.

<https://scholarsarchive.byu.edu/etd/6616>

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, ellen_amatangelo@byu.edu.

Testing the Test: Expanding the Dialogue on
Workplace Writing Assessment

Lindsay Elizabeth Tanner

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

Jon Balzotti, Chair
Kristine Hansen
Grant Boswell

Department of English
Brigham Young University

Copyright © 2017 Lindsay Elizabeth Tanner

All Rights Reserved

ABSTRACT

Testing the Test: Expanding the Dialogue on Workplace Writing Assessment

Lindsay Elizabeth Tanner
Department of English, BYU
Master of Arts

This project is a case study of writing assessment practices in a particular workplace called “High Hits,” a local search engine optimization (SEO) company. The writing tests given to new hires serve a parallel purpose to academic placement exams, in that they are a high-stress, high-risk situation that aims to evaluate writer ability rather than the quality of the completed task (Haswell 242, Elbow 83, Moss 110). However, while academic assessment measures ability with the aim to improve the students’ learning, workplace assessment is driven by market forces and is seen in terms of return on investment. This case study used qualitative and quantitative measurements to examine the writing tests of employees; this examination was followed by analyzing a random sampling of subsequent writing tasks of copywriters to determine whether the assessment methods being used by the company to assess the writing tests adequately predicted the writing ability of the copywriters.

Keywords: writing assessment, writing tests, workplace assessment, grounded theory

TABLE OF CONTENTS

TITLE PAGE	i
ABSTRACT.....	ii
TABLE OF CONTENTS.....	iii
LIST OF TABLES.....	v
Introduction.....	1
Review of Literature	4
Writing Assessment Theory and Practice.....	4
Methods.....	11
Participants.....	13
Procedures.....	14
Survey	15
Informed Consent.....	15
Focus Group.....	15
Writing Test	15
Writing Samples.....	16
Rubrics	17
Training Student Raters	18
Interrater Reliability.....	20
Data Analysis	23

Results and Discussion	26
Applications to Workplace Writing Assessment	35
Limitations of the Present Study.....	37
Future Research	38
Works Cited	41
Appendix One: Blog Rubric	44
Appendix Two: Copy Rubric.....	45
Appendix Three: Keyword Codes	46
Appendix Four: Copywriter Writing Test.....	47
Appendix Five: Copywriter Demographic Survey	49
Appendix Six: Participant Recruitment Email.....	51
Appendix Seven: Informed Consent for Focus Group Participants.....	52
Appendix Eight: Informed Consent for Copywriter Participants	54
Appendix Nine: Student Rater Job Requirements	56
Appendix Ten: Student Rater Grammar Test	57

LIST OF TABLES

Table 1. Copywriter Demographic Information	13
Table 2. Exact and Adjacent Inter-Rater Reliability.....	20
Table 3. Causes of Inter-Rater Disagreement.....	21
Table 4. Open Coding Process.....	24
Table 5. Keyword Rankings According to Rubric Tier.....	24
Table 6. Keywords Calculated to Integer Values	25
Table 7. Analysis of Statistical Significance of Test Variables Compared to Writing Sample Variables.....	27
Table 8. Date Trends Evaluating Statistical Significance of Test Scores Compared to Writing Sample Comments and Scores.....	28
Table 9. Date Trends Evaluating Statistical Significance of Writing Sample Improvement Over Time	29
Table 10. Mean Numerical Scores for Writing Tests Compared to Copy and Blogs.....	30
Table 11. Mean Qualitative Comments for Writing Tests Compared to Copy and Blogs	32
Table 12. Writer Rankings by Post-Hire Mean Comments and Mean Scores.....	33
Table 13. Ordered Differences of Rubric Tiers (Copy and Blogs).....	34
Table 14. Analysis of Variance of Comment Values by Score (Copy and Blogs).....	35

Introduction

Since writing is a complex and multi-dimensional task, many employers hiring new writers face difficulties in determining a copywriting candidate's writing proficiency, their ability to meet a standard defined by a discourse community (White 12). Many employers use the direct test method, which asks applicants to solve a writing problem within a short period of time and create an original text. Hiring managers use a variety of strategies to evaluate these texts and determine whether the applicants are capable of meeting a company's requirements. The managers often assume that one successful task completion predicts a writer's ability to consistently produce comparable texts.

However, the evaluation of applicants' writing tests is not as straightforward as it may seem. Evaluation of client-facing writing is often performed with help of rubrics, but initial interviews I completed with hiring managers for this study suggest that rubrics are often ignored or underutilized during the application process. It is also often unclear whether the writing tests that receive high, medium, or low scores correspond with a candidate's actual performance on the job.

In a review of recent scholarship, there is some debate about how effective rubric scoring can be in evaluating writing ability. Peter Elbow argues against a holistic rubric that results in a numeric score. He states that since writing contains innumerable variables that determine quality, it is impossible to reduce any written product to a single metric (122). Pamela A. Moss counters that standardization and quantification in writing assessment are essential for practical purposes. She argues that less standardized forms of evaluation produce results that are difficult to compare and to use in generalizing performance (110). Even if the holistic vs. analytical scoring debate were resolved, studies still show that one writing sample may not predict future scores. John R.

Hayes performed a study to examine student consistency in essay writing. He compared the scores of almost 800 student essays written throughout a single semester and found that students had quite low consistency in scores from essay to essay (5). This research points to a conclusion that hiring managers generally do not take enough data into account to produce an accurate measurement of success when evaluating writers' performance.

However, the research in holistic rubric scoring has been pursued mainly in academic settings, which have a different purpose than that of the marketplace. No studies to my knowledge have been conducted to investigate the effectiveness of writing tests in predicting employee writing consistency. Many employers use intuitive approaches in evaluating their employee's writing, but these methods may or may not yield acceptable, consistent, or reliable results.

While little work has been done to study workplace writing assessment (Beaman, Boettger, Dias et al.), writing assessment in academia has been studied extensively. Decades of research have established a pluralistic approach to writing assessment (Ackerman 143, Kaufer and Young 55, MacDonald 321, Petraglia xii). Researchers have found that writing encompasses a diverse number of activities and genres that cannot be unified under a single standard (Downs and Wardle 556). Instead, writing standards must be negotiated by discourse communities, and such standards often reflect the values of a group rather than any objective standard (Elbow 83). Writing proficiency becomes not a question of objective quality but of a writer's ability to conform to standards established by a community.

In the classroom, teachers use writing assessment as a matter of course to measure progress, to assign placement, to give feedback, and to certify proficiency, among other purposes (CCCC 431). Assessment also plays an integral part in the academic world by educating

“teachers and students alike as to the kinds of work, methods, criteria, and standards which are and ought to be valued” (Wiggins 130). In contrast, however, workplace writing assessment is mostly used for placement, specifically in hiring, and to enforce institutionalized standards (Dias et al. 114). Writers in the workplace environment must resolve the inherent difficulty of identifying values of “good writing” that most likely do not mirror earlier classroom experiences they may have had in the classroom.

The current status of little research on assessing writing in the workplace provides an ideal opportunity for expanding the dialogue on the topic. In order to gain additional insight into workplace writing assessment, I conducted a case study of copywriting employees at “High Hits” (name is a pseudonym), a search engine optimization (SEO) company in Orem, Utah. The company employs a small team of 10-20 copywriters to produce content for small- and medium-sized businesses’ websites in order to help them appear high in the list of “hits” when a potential customer does an Internet search for a particular product or service. The company is an ideal case to study because they use standard practices when hiring copywriters, using a writing test, conducting interviews, and reviewing resumes. The hiring managers use their own knowledge and experiences to evaluate the writing test. After the copywriters are hired, they are trained on company standards (their discourse community) and their daily writing tasks are scored using a holistic rubric. However, despite these measures, the managers and writing teams express frustration that many new copywriters fail to consistently meet company standards, even after passing the initial writing tests. Many companies experience communication and writing deficiency similarly to this SEO company, a situation that creates significant costs and confusion.

This case study attempts to address this frustration: it asks if writing tests evaluated by experienced professional raters predict writing proficiency. Writing tests will be compared to writing samples “High Hits” copywriters wrote after being hired. Holistic rubric scores and student raters’ comments will be used as the data sets to determine whether the company’s current metrics for evaluating copywriter applicants’ proficiency are predictive of the copywriters’ performance after they are hired. A review of the literature on writing assessment theory and practice provides the foundation for the methodology used in this case study.

Review of Literature

Writing Assessment Theory and Practice

Scoring is an integral step in writing assessment. In creating a score that sums up the writer’s test, scorers translate the information they gather from examining the writer’s work into a numeric value that can be compared to other writers’ numeric scores (Moss 110). When multiple raters reach agreement on the scores that should be given, the raters’ scores are said to have high inter-rater reliability and the writing test is considered reliable. The reliability index is used to compare scoring writing samples to meet “the same (or similar) reliability standards achieved by other testing methods” and thereby improve the method’s credibility compared to more objective assessment methods (Elbow and Yancey 94).

The rubrics used during scoring may be general or task-specific (the rubrics in this case study are task-specific). General rubrics are intended to be used to score any writing activity, while task-specific rubrics are altered to more precisely evaluate a particular writing activity (Bean 270). Both kinds of rubrics have strengths and weaknesses. Task-specific rubrics are seen as more accurate and descriptive of specific writing activities. However, a score derived from a specific writing activity may not reveal anything useful about the writer’s performance in a

different type of writing (White 585). General rubrics, on the other hand, cover a wider variety of writing tasks and so can serve as a more generalized measure of writing performance. However, they may be inappropriately applied to all writing activities, whether or not the criteria identified in the rubrics are actually valuable measuring points in the writing activity (585).

Practitioners primarily use some form of the following two methods to translate the rater's reading into a score: analytic and holistic scoring (the rubrics in this case study use holistic scoring). Analytic scoring, or multiple trait scoring, uses separate scales for several criteria and produces a set of scores (Rezaei and Lovorn 19, Bean 270). Analytic rubrics define levels of achievement for each criterion. For example, a rubric might have separate categories for organization, analysis, and grammar. The rubric would produce three numeric scores. Holistic scoring, also called global or impressionistic scoring, produces a single numeric score (Lee, Gentile, and Kantor 391). Holistic scoring emphasizes product rather than process, focusing on the appearance of the final draft rather than the steps taken to arrive at that draft (Rezaei and Lovorn 19). Analytic scoring maintains more detail of writers' strengths and weaknesses, while holistic scoring flattens that variation to produce a single score (Lee, Gentile, and Kantor 391). Many rubrics, holistic and analytic, use a sliding scale with step-down descriptors for different degrees of writing proficiency. For example, a high score could be described with words like "always, fully, high" and low scores are described with words such as "minimally, limited, etc." (Bean 275). Most practitioners use analytic or holistic scoring, or some combination of the two methods, to create rubrics.

The Educational Testing Service first developed holistic scoring in the mid twentieth century to lower costs of assessment and to increase agreement between raters (White 584). In that study, senior researcher Paul Diederich demonstrated that readers could use the same set of

criteria as a guideline to increase consistency between raters and their scores (Diederich 32). This guideline was eventually formalized into a rubric. Diederich also defined a set of procedures for training, grade norming, and calculating inter-rater reliability that are standard practices for holistic scoring today. Since holistic scoring is driven by an emphasis on comparability and agreement, it only functions under the assumption that agreement (between members of a discourse community) and validity (the idea that the test accurately measures what it is intended to measure) are synonymous in practice. This assumption is derived from classical test theory, which states that the score observed by a rater is “an estimate of the true proficiency, or true score, that is exhibited in the sample” (Penny, Johnson, and Gordon 145). The score can only be an estimate because raters imperfectly interpret the rubric guidelines. Rubric training is performed with the aim of improving agreement between raters, called inter-rater reliability (145). However, while agreement in holistic scoring provides practical benefits, scholars have challenged the assumption that agreement always reflects proficiency in those whose writing is being assessed. These challenges reveal possible weaknesses in holistic scoring.

First, assessing a complex skill like writing may result in tradeoffs between reliability and validity. Reliability refers to agreement between raters, consistent scores assigned by the same rater, and consistent scores over time by different writers. Validity refers to the test’s ability to properly measure what it claims to measure. Elbow argues that there is “an inherently inverse relationship between reliability and validity. As reliability improves, validity degrades—and vice versa” (Elbow 88). Inherent to his criticism is the question of whether the reliability produced by agreement is a valid expression of value or simply a display of agreement. Inter-rater reliability is calculated by standard deviations from the mean, so even if multiple raters have comparable mean scores, they could still be misinterpreting the rubric guidelines. Elbow argues that the

significance of agreement is even more tenuous. He claims that any agreement about “value” in writing actually “produce agreement only about conventions” (89). In that case, even a perfect application of rubric guidelines reveals only how well the text adheres to the rubric. Of course, under some circumstances there is value in assessing how well a text follows a rubric’s guidelines, but that assessment does not always correlate with writing proficiency, or a writer’s ability to create texts that meet all the standards defined by a discourse community that may not be included in the rubric.

Second, while holistic scoring assumes that the score represents the true value of a text, rubrics and texts do not always perfectly align. This problem is especially evident in rubrics that have multiple categories. Richard Haswell explains this conundrum well:

My experience has always been that it is nearly impossible to find anchor essays that are true to the scale of quality pictured by the rubric. What is pictured is quality rising ladder-like and unilaterally across all subskills. A ‘4’ essay is better than a ‘3’ essay in every subskill.... Finding anchor essays that show it, however, is a different matter... the problem re-emerges when readers do not know how to score essays that perform at one level in some subskills and at another level in others. (242)

Invariably, the texts encountered by the raters will deviate from the guidelines outlined in the rubric. Raters must navigate those deviations and decide how to categorize a text that defies categorization.

Third, supposed agreement in holistic scoring can mask disagreement between readers and nuance within texts. The holistic score serves as an “average” of all the positive and negative qualities within the text, hiding most variance and complexity. Raters may be responding to very different parts of the text as they assign a holistic score, but there is no way of knowing what

they notice without more information (Lee, Gentile, and Kantor 394). Additionally, if enough raters score a particular text, eventually the text will receive an average score that appears to show agreement. However, the minimum and maximum scores may vary significantly from the average, which raises the question of which scores should be considered to be the “true score” (Haswell 242). Disagreement may occur because a text does not conform to the expectations of a rubric, causing raters to waffle about the scores they ought to give. These factors can potentially complicate the interpretation of holistic scores.

Fourth, holistic scoring assumes that a score derived from a rubric measures a writer’s ability, when there is actually more evidence that it measures performance at a single task. In high-stakes assessment, such as the writing tests given to copywriting candidates by potential employers, the goal is not just to assess the text, but to place the writer herself into a category (244). While classical test theory argues that the text reflects the writer’s abilities, Elbow argues that an assessor must look “through the text to try to see the writing abilities of the writer behind it” (Elbow 83). However, it is far easier to measure a text than to measure the person who wrote the text. Requiring a writing test to measure the writer as well as the text adds a level of complexity (and therefore uncertainty) that may influence the reliability and validity of a writing test.

Finally, there is debate among researchers as to how raters use rubrics to read writing tests. Some researchers argue that rubrics cause raters to read unnaturally, looking for “adherence” rather than “value.” Many believe that a reader who encounters a text without intending to assign a score has a distinctly different reaction to a text than does a trained rater who approaches it using a rubric. Bean writes that raters are “trained to read in an unnatural way in order to apply negotiated criteria that do not, in any holistic or meaningful sense, belong to the

actual reading practices of real readers” (Bean 277). Instead of “creating a reading,” the rater is forced to “find a meaning that is expected in the schema encoded in the guide” (Elbow and Yancey 97).

Some researchers argue that holistic rubrics are intended to help raters focus on strengths, rather than just deficiencies (Lee, Gentile, and Kantor 394). However, the opposite may also be true: To raters reading with the intent to assign a score, errors are generally more visible than positive elements, so raters are more inclined to notice problems than they do a text’s strengths (Haswell 241). Since rubrics generally list both positive and negative descriptions, they can prevent raters from weighting negative aspects of performance too highly. In this case, a natural reading—i.e., one unaided by a rubric—may actually be an unbalanced reading because many readers may focus more heavily on elements that are naturally more visible.

Richard Haswell argues that a rater’s reading depends more on the rater’s individual experience than on the rubric being used. He proposes three models for assessment: classical, prototype, and exemplar categorization. Classical categorization follows classical test theory; it “assumes that people categorize by grasping the non-accidental properties of a new instance and matching them with the unique set of properties that define the correct category” (Haswell 245). This model assumes that categories have clear boundaries and that raters are perfectly capable of recognizing those boundaries from a generalized description in a rubric and applying them to specific texts. However, in practice, classical categorization is actually the most difficult to train raters to understand, since it requires raters to continually conceptualize generalizations and apply them to specific texts.

Prototype categorization “assumes that people categorize by judging how similar the yet-to-be-categorized instance is to abstract schemas they have of the best example or most representative member (prototype) of possible categories” (246). Grade norming and anchor essays often provide the basis for prototypes. The prototype model differs from the classical model because the raters can rely on a more concrete example to create the boundaries of categories. Raters who have some experience often rely on the prototype model to create readings and assign scores to texts.

Exemplar categorization requires raters to have extensive experience with the kind of texts being evaluated. Raters refer back to “exemplars,” or ideal examples they have read before, and use a “gestalt-like pattern recognition,” depending on “a flock of contextual contingencies, including the categorizer’s previous encounter, subsequent experience with it, and current motivations” to categorize a text (247). However, the exemplar model is the most problematic, because “the rules governing exemplar membership decisions are definitely not deterministic or probabilistic but rather heuristic, norm-based, or interpretive” (247). If experienced raters use only an exemplar model in evaluating texts, it may be difficult to use practices like grade norming and training to standardize rating practices and achieve acceptable inter-rater reliability.

However, if raters use some combination of classical, prototype, and exemplar models in their scoring, it may be possible to achieve a more approximate (albeit not a perfect) measure of writer ability. Haswell’s exemplar evaluation can also be called a “gut reaction,” or an unconscious reading of a text (Elbow 85). In a study performed by Brian Huot, raters who had a background in writing, or an educated “gut,” and who were provided an hour of rubric training were much more successful in achieving inter-rater agreement than a group of raters who had no background in writing (Huot 206). The training (grade norming, prototype model training)

provided a “grounding” to the raters; it allowed them to focus their reading according to the values outlined by the rubric. Elbow calls this grounded reading an “empirical reading,” or a reading that takes into account the raters’ past experience and the requirements outlined by the rubric (Elbow and Yancey 99). Huot reports that this strategy, training raters with expertise, resulted in more consistent scores than training or expertise alone.

Classical test theory assumes that a holistic score can reveal actual proficiency, with some allowance for rater and rubric error. A natural extension of this assumption is that a writer’s proficiency should be consistent from one writing sample to the next. However, the model requires the assumption that not only can raters identify a “true score” for a text, but that that “true score” also represents a writer’s ability, rather than just the proficiency located within the particular text. I test these assumptions in my project by rating the high-stakes assessment (i.e., the writing tests given to job applicants) exactly the same as the writing samples the copywriters created after being hired, using the company rubric. I had previously worked as a hiring manager at “High Hits,” so I had even more interest invested in the research question. Since researchers offer no advice on how exactly to “look through” a text to see the writer’s abilities, I attempted to evaluate the text itself (both the writing tests used during hiring and the writing samples created on the job) as a proxy for ability and see how well that evaluation predicts ability over time.

Methods

This case study uses qualitative and quantitative measurements to examine the writing tests administered to would-be employees at the time they applied for a job at “High Hits,” the SEO company I studied. This examination was followed by analyzing a random sampling of subsequent writing tasks of copywriters after they were hired. This project evaluated writing

assessment for the two types of SEO content that “High Hits” creates for clients: blogs and copy. Blogs provide general information specific to a client’s industry. Blogs serve two SEO purposes: they provide continual “fresh” content to a website’s blog pages, showing search engines that a company is providing content of value to readers, and they demonstrate a company’s expertise within the industry. Copy is static content that appears on a client’s main website, such as the home page. The purpose of copy is to actively sell a client’s products or services to their customer base. Ideally the blog and copy content results in a website’s increased SEO presence and in increased sales.

The company used rubrics they developed in-house to score all blogs and copy created by copywriters on a day-to-day basis. They used the rubrics to enforce industry best practices and maintain a consistent standard of quality to their clients. The company used two distinct rubrics to score the two types of content, with genre-specific requirements outlined in each rubric. However, the majority of copywriters who participated in this study were not evaluated by the company rubrics during the hiring process because the rubric standards had not yet been finalized. Instead, hiring managers relied on an exemplar, or empirical reading of the writing tests, using their past experience with copywriting to identify candidates they thought would succeed in the company.

The purpose of examining both the writing tests and the content written after the copywriters were hired is to determine whether the assessment methods being used by the company to assess the writing tests adequately predicted the writing ability of the copywriters after they were hired. I also aimed to test an alternative assessment practice, using the company rubrics to score the writing tests, to see if the rubrics could be used during the hiring process as a better predictor of copywriter scores after hiring. This project was approved by the Internal

Review Board (IRB) on August 30, 2016. Generous funding has been provided by the Brigham Young University Graduate Research Fellowship award to assist in research expenses, including hiring three student raters to do the actual scoring.

Participants

This study involved the participation of two groups: copywriters and managers. The copywriters worked on a copywriting team at “High Hits” of between 10 and 20 employees. Some copywriters had been full-time employees and some were freelance employees. The freelance employees received payment based on the word count of the writing they produced. Most copywriters had worked for the company for a year or less at the time that writing samples were collected. For many copywriters, this was their first full-time position after graduating from college. Copywriters were recruited through an email sent by the manager of the team. They were given a \$10 gift card as nominal compensation for participation.

Table 1. Copywriter Demographic Information

Gender		Age		Education		Employment Status (As of Sept. 2016)		Household Income	
Female	9	18-24	3	4-year degree	9	Full time	5	< \$20,000	3
Male	1	25-34	7	Masters degree	1	Part time	4	< \$30,000	3
						Unemployed	1	< \$40,000	3
								< \$60,000	1
Total	10		10		10		10		10

Time Employed by Company		Marital Status		How Employee Discovered Job		Reason for Leaving	
1.5 years	3	Never married	4	Employee recommendation	5	Current employee	5
2 years	7	Married	6	Online job board	3	Insufficient pay	3
				Company website	2	Discomfort in work environment	2
				University connections	1	Let go	1
				Job fair	1		
Total	10		10		11*		11*

*Some participants included multiple responses

The second group of participants were managers of the copywriters. They defined and supported the institutional standards given to the copywriters. They also interpreted the feedback given by clients and employees from other departments regarding blogs and copy. The managers in this group had varying roles regarding the written products: two employees managed the entire copywriting department and had limited day-to-day interaction with copywriters, while the other three employees directly managed copywriters and also participated in implementation of the standards. Managers were recruited through an emailed invitation from the principal investigator. They were also given a \$10 gift card as nominal compensation for participation.

Procedures

After the project attained IRB approval, I sent a short demographic survey, hosted by Qualtrics, to twenty current and past copywriters of “High Hits.” Ten copywriters filled out the survey and signed the informed consent form attached to the survey. After gaining consent from the copywriters, I gathered necessary documents from “High Hits,” including the copywriters’ writing tests, all blogs and copy available, the company rubrics, and relevant training materials that I used to train my raters.

I also began the process of hiring my own student raters to do the actual scoring. A job description was posted on the student employment website (See Appendix 9 for the job description text). I hired student raters because they shared many traits with “High Hits” copyeditors, who were recent college graduates themselves. I chose three raters who had previous experience copyediting, had satisfactory writing samples, and passed a grammar test with a score of 90% or better (See Appendix 10 for the full grammar test given to student raters). I hired three student raters: two undergraduate seniors who were in the process of completing an editing minor and one first-year graduate student who had experience grading student work. I

trained the raters using company training materials that I had previously used as a manager at “High Hits” training new employees.

Survey. Participants filled out a short survey detailing background information, including gender, age, employment status (as of September 2016), education, household income, and the method they used to find out about the copywriting position. The demographics of the participants are very similar. It is unlikely that demographic factors had a statistically significant effect on findings. However, unfortunately I did not gather data about previous copywriting experience. From anecdotal evidence, I do know that some of the writers had previous experience writing copy and others did not, but without information from the whole group, I cannot draw any definitive conclusions about how previous experience affects performance.

Informed Consent. Copywriters were provided with documentation for informed consent for their writing to be viewed by student raters after being de-identified. Copywriters also filled out a short demographic survey. The managers were provided with documentation for informed consent for their focus group participation.

Focus Group. Since the standards for SEO content quality are continually being defined, focus group interviews were conducted with the copywriting team lead and the heads of the copywriting department to determine what company writing standards. The main topic of discussion was the current company rubrics that the copywriting team used to score the SEO content. This meeting confirmed that the company was happy with the current rubrics and the standards the rubrics enforced. Based on this feedback, I felt confident using the company rubrics in the current project to measure company standards.

Writing Test. Like academic placement tests, the writing test at “High Hits” seems to be guided under the classical test theory assumption that one successful task completion predicts a

writer's ability to consistently produce comparable texts. The writing test was designed to predict consistent writing proficiency. One measure of the effectiveness of a writing test is the authenticity of the task (Wiggins 130). This test is as similar as possible to the day-to-day writing on the job. It contains two prompts: one for a blog and another for copy. The blog prompt asks the candidate to write a 250–350-word article for one of three clients. It gives instructions that are remarkably similar to the requirements in the rubric: keep the audience in mind, use formatting to make the text easier to scan, and write a compelling title. The copy prompt also mimics a common task on the job—rewriting a poorly written web page for a client. The prompt provides minimal information from the client as well as instructions that reflect rubric requirements: be persuasive, speak in the company's voice, use the keywords, and include a call to action. The applicants were not given a copy of the rubric during the application process, since the rubric text had not been finalized until 2016, after all the participants had been hired as copywriters (all were hired in 2014 or 2015). (See Appendix 4 for the full text of the company writing test.)

“High Hits” was not entirely consistent in administering writing tests to copywriter applicants. Six copywriter participants received one writing test each, with prompts for both copy and a blog. Two copywriters took two writing tests, with prompts for both copy and blog. One participant's writing test did not include a copy prompt, and another participant's writing test did not include a blog prompt. Each test was graded twice by two different student raters. In total, 18 blog and copy prompts within the writing tests were collected; there were 36 readings of the tests.

Writing Samples. The company began keeping all original versions of the copywriters' writing samples after August 2015. I selected a random sampling of copywriters' copy and blogs

written between August 2015 and August 2016. Some employees did not work for the full year that samples were collected, resulting in an uneven number of samples from each copywriter. One copy text and one blog were collected per month per copywriter. Each text was graded twice by two different student raters. In total, 221 writing samples were gathered; there were 442 readings of the texts. All identifying information was removed from the writing tests and writing samples and replaced with numerical markers before writing samples reached the undergraduate raters.

Rubrics. “High Hits” used two holistic, task-specific rubrics to evaluate two different writing products: blogs and copy (See Appendix 1 and Appendix 2 for the complete rubrics). The rubrics were solely used in day-to-day workplace writing, not to evaluate applicants’ writing tests. The rubrics used by “High Hits” are holistic rubrics, but instead of describing multiple subskills on a sliding scale (e.g. “Is the organization excellent, acceptable, or substandard?”), the rubrics use ranked tiers. Each tier describes a set of requirements that must be met in order to move on to the subsequent tier. For example, Tier 1 describes the qualities that are absolutely essential for copywriting: originality (no plagiarism), correct audience, accurate information, and a positive attitude. These terms occur only in Tier 1; there are no “degrees” of proficiency in higher tiers. If a copywriter addresses an incorrect audience, the text automatically receives a Tier 1 score and cannot receive any other score. Requirements terms are not repeated between tiers. If a text lacks any of the required qualities in Tiers 1 through 3, it automatically receives a failing score and is returned to the copywriter for revision. It does not matter if the text contains any of the qualities from higher tiers; the writer must meet the most basic requirements and make changes before an editor will look at it again. If a writer meets all the requirements of Tiers 1 through 3, they can earn a Tier 4 score, which is considered passing. Tier 4 contains qualities that

are no longer requirements, but that indicate proficiency according to the company standards.

Although these workplace rubrics do share many purposes with academic rubrics, such as to give feedback and aid in writer improvement, the main purpose of this workplace rubric is to sift between acceptable and unacceptable texts (CCCC 431). This rubric is even more product-oriented than classroom rubrics. In their role as professional copywriters, the writers are expected to consistently produce texts that climb the tiers of the rubric to Tier 4. Revisions and rewrites translate to lost profits by their employer. While professional development is important to the company, consistent proficiency (passing scores) is expected.

Training Student Raters. I relied on the theories proposed by Huot, Haswell, and Elbow in developing my training methodology. I hired raters with previous experience in editing and gave them two one-hour trainings to give them guidelines for evaluating SEO copywriting. As a former copywriter and hiring manager employed by “High Hits,” I used my experience training new copywriters to give my student raters the information they needed to use the company rubrics. Raters were trained to use the rubrics and to grade the writing samples using online software created for this project by an undergraduate Information Technology student at BYU. The online software presented one writing sample at a time to the student raters, with a text box for inputting the rubric score and a text box for writing a paragraph of qualitative feedback about the writing sample. The rater would read the text, evaluate the text according to the training and rubric she received, input the score and commentary, and click a “next” button to populate a new page with the next writing sample. Each page also provided background information written by the copywriters about the writing sample, including the client’s URL, a short audience profile, content strategy, the copywriter’s intended tone, the type of page being written, and references (when relevant). Raters took this background information into account when evaluating the

writing samples.

I performed grade norming with the student raters four times total. First, they were given ten practice samples of blogs to grade that were not used in the study. Then I gave them feedback for each score and commentary based on how well they utilized the rubric standards. Two weeks later, I reviewed a random sampling of their scores and comments and provided more feedback. I repeated the process when I gave them copy tasks to rate. If raters had any more questions about using the rubric, they contacted me for further feedback. However, they did not always contact me when they had questions and noted in the qualitative comments when they were unsure of the scores they gave, which resulted in some uncertainty in some of the rubric scores.

I also instructed the student raters to transcribe qualitative comments to explain the rubric scores they assigned. I asked the raters to focus on specific details that reflected the rubric score, but to also include other feedback based on their own experience with editing. In order to be as consistent as possible, I had the raters include as much detail as possible in their comments. Most comments were between two and three sentences long. These comments were meant to supplement the rubric scores and provide more insight into why the raters assigned particular scores to particular texts. The comments were also meant to reflect Elbow and Haswell's theory that raters' past experience, or gut intuition, often affect how raters interpret texts and apply rubrics. Since the comments were created in the student raters' own words about elements of the text they noticed organically, these comments ought to be more reflective of their experience and intuition than the rubric scores they assigned to the writing samples. The numerical scores and qualitative comments provided by the student raters served as the data to test my study's hypothesis that writing tests evaluated using the rubric would be more predictive of writing samples created after the copywriters were hired.

Inter-Rater Reliability. Since holistic scoring relies on agreement to ensure reliability, I calculated the inter-rater reliability of the three raters using percent agreement. Table 2 illustrates the average inter-rater reliability for all three raters, broken down by exact agreement (raters assigned exactly the same score) and exact and adjacent agreement (raters assigned the same score or a score that differed only by one point). The exact agreement percentage was 71%, while exact and adjacent agreement was 87%. Additional training and better communication between my student raters and myself would have served to improve these percentages, but the raters' busy student schedules often prohibited me from giving them as much guidance as I would have wished about using the company standards and interpreting the rubric.

Table 2. Exact and Adjacent Inter-Rater Reliability

	Raters 1 & 2	Raters 1 & 3	Raters 2 & 3	Average
Percent Exact Agreement	75%	63%	76%	71%
Percent Exact and Adjacent Agreement	96%	78%	87%	87%

Since inter-rater reliability was an issue in this study, I investigated the raters' comments for each pair of scores to discover the causes of disagreement. Table 3 illustrates my findings. The disagreements fit into three categories: one rater assigned a failing score (Tier 1, 2, or 3) while the other rater assigned a passing score (Tier 4 or 5), the raters assigned two different failing scores (Tier 1, 2, or 3), or the raters assigned two different passing scores (Tier 4 or 5). Since the rubric is divided between pass and fail scores, the degree to which a copywriter passes or fails is not as important to the stakeholders of the company (although it is important to the copyeditors) because only a failing score must be rewritten, causing a financial cost to the company. While all forms of disagreement are concerning because they represent a possible breakdown in the validity of the rubric, the most concerning category of disagreement was also the largest category: disagreement about whether the sample passed or failed the company's

standards. This type of disagreement accounted for 20% of all disagreements.

Table 3. Causes of Inter-Rater Disagreement

	Raters 1 & 2		Raters 1 & 3		Raters 2 & 3		Overall Disagreement	
Pass and fail scores both assigned	15	18%	24	24%	13	16%	52	20%
Degree of failure disputed (Tier 1, 2, or 3)	1	1%	3	3%	2	2%	6	2%
Degree of passing disputed (Tier 4 or 5)	5	6%	9	9%	5	6%	19	7%
Total number of samples	85		98		83		266	

I looked more closely at the comments provided for the 52 readings that were assigned both pass and fail scores to discover any possible reasons for these discrepancies. Of the 52 readings, twenty readings, or 38% of the total, were what I would consider to be rater mistakes, or misreadings, of the rubric. Thirteen of those readings (25% of the total) resulted from confusion about the company’s keyword practices (a Tier 2 issue). During the course of the year from which I collected data, “High Hits” changed their keyword practices, so copywriters formatted keywords differently in about half of their writing samples. My student raters did not always recognize which keyword practices were being used in each writing sample and occasionally assigned a score using the incorrect keyword practice. The remaining seven divergent scores were caused by a student rater’s overly strict interpretation of the rubric. For example, Rater 3 assigned a piece of copy a Tier 3 score for poor readability because the first two paragraphs of the copy did not “get to the point” quickly enough (per the guidelines in the rubric). However, Rater 2 noted in her comments of the same sample that since the text went significantly over word count, a company editor could easily delete the two paragraphs and fix that issue. Rater 2 assigned the copy sample a Tier 4 passing score, which was a more appropriate score.

In the remaining thirty cases (58% of the total), the variation of scores was justified by the student raters' interpretations of the rubric. For example, one writer, Ken, wrote a blog about preventing coal mining accidents for a law firm that specialized in worker compensation. Both student raters noticed that the blog addressed both miners and their supervisors, which made the intended audience of the blog unclear. Both raters agreed that the blog needed to be reworked before it could be sent to the client, but they had different opinions about how to fix the blog. Rater 1 gave the writer a Tier 2 rejection for misidentifying the audience and recommended that the writer focus the whole blog on coal mining supervisors, since they seemed to have the most authority to fix the problems listed in the blog. However, Rater 2 thought that the blog could be arranged into two sections with advice for coal miners and their supervisors. She assigned the blog a Tier 4 passing score and recommended reorganization rather than a total rewrite. Both raters provided viable solutions to edit the blog, and both solutions corresponded with the scores they assigned from the rubric.

Taking these considerations into account, the disagreements between scores are actually not so worrying as they initially appeared. 58% of the disagreements that the student raters had between passing and failing scores were actually justified by different interpretations of the rubric. The disagreements between degrees of failure and degrees of passing (9% of total disagreement) do not pose a problem for the company stakeholders, since a sample that receives a failing score of any number must always be rewritten and a passing score of any number is always passed on to the client. The question of pass or fail is not at question in that type of divergent scores. Overall, although better training and communication would have improved the student raters' use of the rubric (especially regarding keyword practices), the student raters were mostly satisfactory in their scoring.

Data Analysis. The analysis of the took two forms: open and axial coding based in grounded theory of the qualitative comments and statistical analysis of the numeric rubric scores. My coding analysis was based in Joyce Neff’s explanation of grounded theory, in which the researcher performs a series of passes over textual data to code the data and “examine conceptual relationships and to generate theory” (Neff 125). Neff’s grounded theory uses three kinds of passes (open, axial, and selective) to translate the specific information in a study to more generalizable theory. I utilized only open and axial coding in my analysis. Open coding disassembles the text into keywords and creates categories from the keywords. Axial coding organizes the categories into groups based on features and relationships (Neff 128).

I used open coding on the qualitative comments. Since my research focused on evaluating the rubric as an assessment tool, I used language from the rubric as a lens through which to generate the keywords. For example, if a rater noted that a text had useful information for the reader, I would use the keyword “use value,” since that language corresponded to a requirement in Tier 4 of the blog rubric. Then I noticed that the comments my raters provided could be easily divided into positive and negative feedback, so I created two columns that I labeled positive and negative feedback. I placed a keyword in the positive column to indicate that the quality was executed well and in the negative column to indicate that the quality was executed poorly. With the separate columns, I could use the same keywords for both columns rather than defining specifically positive or negative keywords. Table 4 shows how I extracted rubric-specific keywords from a comment provided by a rater.

Table 4. Open Coding Process

Copy ID	Score	Comment	Positive Keywords	Negative Keywords
125	4	This copy was accurate and positive. It uses best keyword practices and also has a good conversion funnel (CTA). The copy used specific details and emphasis.	accuracy, positivity, keywords, conversion funnel, CTA, specific, details	
125	4	The copy does a good job of addressing the audience and maintains a professional and friendly tone the embodies the persona of the client. There were a few sentences that could have been placed more logically (5.3.1), but overall it was well done.	audience, tone, client branding	logical

In order to mathematically express the writers’ performance (and use this qualitative data in statistical analysis), each keyword was given a point value matching the rubric tier to which it belonged (e.g. since the keyword “logical” belonged in Tier 3 of the rubric, it was assigned the value 3). Each keyword corresponded to only one tier of the rubric, and so each keyword was assigned a ranked value. Table 5 illustrates the point values that corresponded to the positive and negative keywords extracted from the qualitative comments.

Table 5. Keyword Rankings According to Rubric Tier

	Tier 1	Tier 2	Tier 3	Tier 4	Tier 5
Positive keywords	1	2	3	4	5
Negative keywords	-5	-4	-3	-2	-1

Positive keywords received a ranked value that corresponds to tier level. Tier 1 keywords have a value of one, because receiving a positive comment about fulfilling a basic requirement is not a very impressive indicator of a writer’s proficiency. Tier 5 keywords have a value of five, because receiving a positive comment about achieving excellence is more impressive. To

contrast, negative keywords have an inverted ranking with negative values. A negative comment with a Tier 1 keyword is a much more problematic indicator of writer performance than a negative comment using a Tier 5 keyword.

I counted the number of times each keyword appeared in my open coding for each writing sample and multiplied the number of keywords by the ranked values of the corresponding rubric tiers. I completed this process for each writing sample, resulting in an integer value for the positive keywords and the negative keywords. The positive and negative values were also added together to create one value that could be compared to the numeric rubric scores. This process of ranking and quantifying keywords constituted axial coding. See Table 6 for an example of the calculation derived from the example in Table 2.

Table 6. Keywords Calculated to Integer Values

Copy # 125	Positive Keywords	Count	Rank	Positive Comment Value	Negative Keywords	Count	Rank	Negative Comment Value	Total Comment Value
Rater 1	accuracy	1	1	1					
	positivity	1	1	1					
	keywords	1	2	2					
	conversion funnel	1	3	3					
	CTA	1	3	3					
	specific details	1	4	4					
		1	3	3					
Total				17				0	17
Rater 3	audience	1	1	1	logical	1	-3	-3	
	tone	1	2	2					
	client branding	1	3	3					
Total				6				-3	3

In Table 6, Copy #125 produced two positive comment values (17 and 6), two negative comment values (0 and -3), and two total comment values (17 and 3). The separate positive and negative comment values were used in statistical analysis of a possible correlation between the writing tests and the writing samples to test the hypothesis of this case study. The total comment

values were used in statistical analysis comparing a possible correlation between the comment values and the numeric rubric scores to test the rubric's ability to mimic the student raters' empirical readings of the writing samples.

This application of grounded theory resulted in numeric values that could be used in statistical and comparative analysis, which is not a traditional use of keyword coding. However, the purpose of grounded theory is to closely examine the assumptions of the research project. The main assumption being examined here is that the rubric properly measures the standards of "good writing" as defined by the company. I created numerical values for the comments attached to each writing sample using that assumption. Statistical analysis was then used to examine that assumption.

The values derived from the student rater comments, numeric rubric scores assigned by the student raters, dates the samples were written, and author designations were used in statistical analysis.

Results and Discussion

Statistical analysis system (SAS) software was used to run statistics tests on the numerical comment values and rubric-generated scores to determine if a statistical correlation existed, between the comments and scores assigned to the writing tests given to copywriters as job applicants and the comments and scores assigned to the writing samples created by copywriters as employees. The results of the statistical analysis are summarized and interpreted below.

First, I determined a statistical correlation between the values derived from the positive and negative comments written by the student raters and the rubric scores. Since the comment values were based on the student raters' observations rather than on a standardized scale, they

could only be compared to the rubric scores if a correlation could be observed. I performed a one-way analysis of variance comparing the comment values to the rubric scores, which yielded a p-value of < 0.0001 for both copy and blogs, or a high probability of statistical correlation between the comment values and numeric rubric scores assigned by student raters. Since a correlation did indeed exist between the comment values and rubric scores, they were included in the data sets used in this study.

A statistical analysis was also conducted to see if any correlation existed between the writing tests and the writing samples. A restricted maximum likelihood approach was used to compare the variables of the writing tests and the writing samples. The writing tests generated the following variables: numerical rubric scores and positive and negative qualitative comments. The writing samples generated comparable variables: numerical rubric scores and positive and negative comments. Tables 7 through 9 display the results of this analysis.

Table 7. Analysis of Statistical Significance of Test Variables Compared to Writing Sample Variables

Test Variable (Pre-hire)	Writing Sample Variable (Post-hire)	Slope	Percent Change	p-Value
Copy test positive comments	Copy positive comments	0.33	4%	0.12
Copy test negative comments	Copy negative comments	0.03	1%	0.69
Copy test numeric score	Copy numeric score	0.05	1%	0.46
Blog test positive comments	Blog positive comments	0.11	1%	0.48
Blog test negative comments	Blog negative comments	0.08	3%	0.06
Blog test numeric score	Blog numeric score	0.29	11%	0.01*

*p < 0.01 is statistically significant

Table 7 shows the results of an analysis of whether test variables have a statistically significant relationship to writing sample variables. This analysis aims to directly answer the study's central question: whether the scores and comments gathered from the tests can predict the scores and comments of writing samples after the writer is hired. According to the analysis, the only variable that has a predictive relationship is blog test numeric scores on blog numeric

scores. That is to say, the score a writer receives on the blog test supposedly correlates to scores a writer is likely to receive after being hired. However, no other variables yielded any statistically significant results.

Next, the effect of writing test variables on writing sample variables over time was calculated, to see if the test variables influence the writing sample variables to change over time. The results are presented in Table 8. The only variables that have a statistically significant effect are blog test scores on blog negative comments. When the blog test is taken into account, the statistical model predicts that overall blog negative comments will decrease by -3.38 points over a year, or a difference of 32% per year. However, none of the other variables are statistically significant when change over time is taken into account. I am inclined to be skeptical of this finding since it only applies to one variable out of many.

Table 8. Data Trends Evaluating Statistical Significance of Test Scores Compared to Writing Sample Comments and Scores

Test Variable (Pre-hire)	Writing Sample Variable (Post-hire)	Yearly Rate of Change	Yearly Rate of Change (Percent)	p-Value
Copy test scores	Copy positive comments	NS	NS	NS
Copy test scores	Copy negative comments	-1.88	2%	0
Copy test scores	Copy numeric scores	NS	NS	NS
Blog test scores	Blog positive comments	0.79	11%	0.35
Blog test scores	Blog negative comments	-3.38	-32%	< 0.0001*
Blog test scores	Blog numeric scores	0.17	5%	0.20

*p < 0.0001 is statistically significant

The third statistical analysis performed ignored the test variables entirely and focused on the writing sample variables to see if there was any improvement over time. The results are presented in Table 9. In this case, no variables were found to be statistically significant. That does not mean that the copywriters did not improve over time, but that the methods I used to measure performance (the rubric scores and the positive and negative comment values) did not show improvement.

Table 9. Data Trends Evaluating Statistical Significance of Writing Sample Improvement Over Time

Writing Sample Variable (Post-hire)	Yearly Rate of Change	Yearly Rate of Change (Percent)	p-Value
Copy positive comments	-1.82	2%	0.21
Copy negative comments	0.94	2%	0.30
Copy numeric scores	NS	NS	NS
Blog positive comments	-1.34	-12%	0.37
Blog negative comments	-3.10	-76%	0.68
Blog numeric scores	-0.33	-8%	0.16

The overall lack of statistical correlation between writing test variables and writing sample variables appears to answer my research question of whether a strict rubric reading of the writing tests would serve as a valid predictor of writing proficiency: the scores and comment values assigned to the writing tests were not statistically predictive of the copywriters' scores or comment values that they received for their writing samples on the job.

However, these statistical results only show whether there is any statistical relationship between the exact numerical values of the scores and comment values assigned to writing tests and writing samples. They do not reveal anything about individual copywriters' writing proficiency or if their rubric scores or comment values improved over time. Tables 10 and 11 show the mean blog and copy rubric scores and comment values broken down by copywriter to see a more detailed picture of the changes between the writing tests and the writing samples created after being hired.

Table 10. Mean Numerical Scores for Writing Tests Compared to Copy and Blogs

Writer	Mean Copy Score (Writing Test)	Mean Copy Score (After Hire)	Difference Between Scores
Tina*	3.00	3.94	0.94
Eileen	2.00	3.93	1.93
Lauren	1.75	3.92	2.17
Ken	4.00	3.91	-0.09
Rita	2.00	3.85	1.85
Cathy	2.00	3.75	1.75
Eva	2.00	3.67	1.67
Naomi	N/A	3.67	N/A
Elena	2.00	3.67	1.67
Mary	4.00	3.41	-0.59
Average	2.53	3.77	1.24

Writer	Mean Blog Score (Writing Test)	Mean Blog Score (After Hire)	Difference Between Scores
Tina	N/A	4.04	N/A
Naomi	4.00	4.04	0.04
Ken	4.00	3.94	-0.06
Eileen	3.50	3.86	0.36
Rita	3.00	3.81	0.81
Eva	2.00	3.55	1.55
Lauren	2.50	3.53	1.03
Elena	2.50	3.50	1.00
Mary	4.00	3.21	-0.79
Cathy	2.00	3.16	1.16
Average	3.06	3.66	0.61

*Names have been changed

Table 10 shows the mean numeric scores for copy and blogs for each copywriter. The table shows that the copy and blog writing tests have much lower scores than the copy and blog writing samples after the copywriters were hired. The difference is likely explained by the copywriters' unfamiliarity with the rubric before they were hired, since they were not provided with a rubric during the hiring process. The difference might also be explained by the most common comment given by the student raters for the copy and blog tests, which was that the copywriter applicant failed to follow the test instructions. Many copywriters wrote blogs in first person despite being told not to and only included the keyword once in the copy when they were asked to include it twice. According to the rubrics, those blogs would receive a Tier 2 rejection for use value (the blog was not useful to the reader) and the copy would receive a Tier 2 rejection for keywords, since keywords are essential for SEO. If the hiring managers had evaluated the writing tests using the rubric as strictly as my student raters did, they might have assumed that the copywriters in this study would frequently fail to meet company proficiency standards.

However, the average difference between writing test and writing sample scores is 1.24 for copy and 0.61 for blogs. In a five-point rubric, these differences represent over a full tier and half tier score difference. The copywriters achieved much more proficient average scores after being hired than they did during the hiring process. Since the hiring managers used an exemplar, or empirical reading method to evaluate the writing tests, relying on their experience rather than a strict rubric, it is possible they identified qualities in the writing tests that were not measured by the rubric, but which better predicted future passing scores.

In addition to rubric scores, I also compared the comment values in the writing tests and writing samples to identify changes that could be observed for individual writers. Table 11 shows the results. Although there is no statistical correlation between writing test comments and the writing sample comments, in the table we see that largely negative comments in the writing test give way to largely positive comments post-hire. This trend might be explained in one of two ways. First, since I had the raters score the tests last, after they had graded all the copy and blogs, they may have been more confident pointing out errors than they were when they were recently trained. Also, since the writing tests all used the same prompt, the student raters were aware that these samples were writing tests rather than everyday writing and they might have felt more obligation to point out problems. Second, the difference might be explained because the copywriters' training and experience on the job helped them create texts that more fully met the company proficiency standards. After being hired, the copywriters underwent multi-week training on company requirements and industry best practices. The samples collected for this study were taken between five months and a year after the copywriters were first hired, so the copywriters had plenty of time to improve their proficiency with training and daily practice. Although the comment data yield no statistically significant findings, the switch from mostly

negative feedback to mostly positive feedback may be evidence of how training and practice can improve writing ability.

Table 11. Mean Qualitative Comments for Writing Tests Compared to Copy and Blogs

Writer	Mean Copy Comments (Writing Test)	Mean Copy Comments (After Hire)	Difference Between Scores	Writer	Mean Blog Comments (Writing Test)	Mean Blog Comments (After Hire)	Difference Between Scores
Tina	3.50	9.00	5.50	Naomi	6.00	9.38	3.38
Eileen	-4.50	7.33	11.83	Eva	-25.00	7.73	32.73
Ken	7.50	7.05	-0.45	Rita	-0.50	6.88	7.38
Elena	-9.00	7.00	16.00	Eileen	2.00	6.76	4.76
Lauren	-6.25	5.19	11.44	Tina	N/A	6.00	N/A
Rita	-2.00	5.11	7.11	Elena	-14.00	5.54	19.54
Naomi	N/A	5.04	N/A	Ken	6.00	4.44	-1.56
Eva	-14.00	4.17	18.17	Cathy	-15.50	3.28	18.78
Cathy	-2.50	3.71	6.21	Lauren	-10.50	1.73	12.23
Mary	5.50	2.94	-2.56	Mary	6.75	0.88	-5.87
Average	-2.42	5.26	7.68	Average	-4.97	5.59	10.56

Since the method I proposed, using a strict rubric reading to evaluate the writing tests, ultimately failed to statistically predict writer proficiency post-hire, I examined other data to see if I could create a better picture of that proficiency. I examined the mean comments and scores of the copywriters' writing samples (both copy and blogs) after they were hired and compared them to the standard deviations of the comments and scores. Table 12 shows the results. As I expected, copywriters who had higher numeric scores also had lower standard deviations (meaning they were more consistent in their performance). However, I did not expect copywriters with lower numeric scores to have higher standard deviations. I expected lower performers to be consistently low performers, which was not the case. Instead, these lower performers were more erratic, sometimes producing passing work and sometimes producing failing work.

Table 12. *Writer Rankings by Post-Hire Mean Comments and Mean Scores*

Ranking (by Mean)	Writer	Mean Comments	Standard Deviation
1	Tina	7.26	6.44
2	Naomi	7.17	7.42
3	Eileen	7.04	7.56
4	Eva	6.47	7.14
5	Elena	6.27	9.63
6	Rita	5.98	7.09
7	Ken	5.95	8.15
8	Lauren	3.93	7.16
9	Cathy	3.51	8.02
10	Mary	1.73	6.78
Average		5.53	7.54

Ranking (by Mean)	Writer	Mean Scores	Standard Deviation
1	Tina	4.00	0.44
2	Eileen	3.89	0.73
3	Ken	3.92	0.54
4	Rita	3.83	0.61
5	Lauren	3.78	0.76
6	Naomi	3.85	0.57
7	Elena	3.58	1.07
8	Cathy	3.47	1.03
9	Eva	3.59	0.94
10	Mary	3.29	1.10
Average		3.72	0.78

Comparing the rankings of the copywriters by mean comments versus mean scores, the rank order of the copywriters also shuffles around. The high performers and low performers stay in relatively consistent positions, but the middle of the group varies. For example, Ken appears in rank 7 for comment values but rank 3 by score. Eva appears at rank 4 for comments, but rank 9 for scores. Since comments are meant to correspond to an exemplar or empirical reading and the rubric scores are meant to correspond to a classical or prototype reading, the different rank orders seem to show that the rubric and the student raters' intuition measure the writing samples differently. At this point there is no evidence to trust one form of measurement as more valid over the other, but this lack of agreement is troubling in the context of holistic scoring, which requires agreement to be considered valid.

To discover more about how comment values measure writing samples in comparison to the rubric scores, I used a student's t-test to conduct a comparison for each pair of the five tiers in the blog and copy rubrics. The results, presented in Table 13, show which differences between tiers in correlating comment values are statistically significant. The blog table shows that all pass

and fail pairs (Tiers 1, 2, and 3 compared to Tiers 4 and 5) are statistically different ($p < 0.0001$), while differences between passing scores (Tiers 4 and 5) and differences between failing scores (Tiers 1, 2, and 3) are not statistically significant. This finding is promising because it shows that the comment values tend to correlate with the scores, even when the scores are compared by tier. The copy table, however, has a few different results. It shows that the difference between passing scores (Tier 4 and 5) is statistically significant ($p < 0.0001$). However, the difference between Tier 4, a passing score, and Tier 1, the lowest failing score, is not statistically significant ($p = 0.0945$). This finding might be explained by the low number of Tier 1 scores for copy (only two samples). More Tier 1 samples would be needed to resolve this finding.

Table 13. Ordered Differences of Rubric Tiers (Copy and Blogs)

Ordered Differences of Copy Rubric Tiers			
Rubric Tier	Rubric Tier	Difference	p-Value
5	3	32.60	< 0.0001*
5	2	26.27	< 0.0001*
5	1	24.00	< 0.0001*
5	4	16.46	< 0.0001*
4	3	16.14	< 0.0001*
4	2	9.81	< 0.0001*
1	3	8.60	0.1048
4	1	7.54	0.0945
2	3	6.33	0.0442*
1	2	2.27	0.6263

Ordered Differences of Blog Rubric Tiers			
Rubric Tier	Rubric Tier	Difference	p-Value
5	2	20.19	< 0.0001*
5	1	18.06	< 0.0001*
5	3	16.95	< 0.0001*
4	2	14.02	< 0.0001*
4	1	11.89	< 0.0001*
4	3	10.78	< 0.0001*
5	4	6.17	0.0022*
3	2	3.24	0.2362
1	2	2.13	0.4112
3	1	1.11	0.6019

*p-Values are statistically significant

The difference in the number of samples in each rubric tier may also reveal insight into the methods used to measure the writing samples. There were far more Tier 4 scores assigned than any other score. Tier 4 scores accounted for 83% of copy scores and 79% of blog scores. Since that group of scores was so much larger than the other tier scores, the range of comments that correlate with the group is also much larger. Table 14 illustrates the variance of comment

values within each score. The most surprising results from this analysis was the number of overall negative comment values associated with Tier 4 scores. While the mean comment values associated with Tier 4 scores was positive for both copy and blogs (6.54 and 7.17, respectively), there were more negative comments given to Tier 4 writing samples than I would have expected for passing work. Tier 4 copy scores had 36 instances of negative comments (21% of all Tier 4 copy scores) while Tier 4 blog scores had 36 instances of negative comments (20% of all Tier 4 blog scores). While these negative comments may be a small percentage of all Tier 4 scores, one would expect a rubric that assigns passing scores to writing samples to inspire mostly positive and neutral comments from the raters reading those samples.

Table 14. Analysis of Variance of Comment Values by Score (Copy and Blogs)

Copy					Blogs				
Tier	Number of Samples	Mean	Standard Deviation	p-Value	Tier	Number of Samples	Mean	Standard Deviation	p-Value
1	2	-1	5.66	< 0.0001*		18	-4.72	2.95	< 0.0001*
2	22	-3.27	4.92		2	7	-6.86	3.89	
3	5	-9.6	4.51		3	13	-3.62	4.74	
4	185	6.54	6.46		4	174	7.17	6.04	
5	8	23	6.97		5	9	13.33	7.98	

*p < 0.0001 is statistically significant

These results may reveal some issue with current rubric that could be resolved with further study. Or they may reveal that perfect agreement is not the end goal of assessment, and that some complexity of data is required to get a fuller picture of writer performance.

Applications to Workplace Writing Assessment

Although this case study failed to prove that a strict rubric reading of writing tests is predictive of rubric readings of writing samples produced on the job, it did reveal that the current practice (experienced hiring managers using an exemplar reading method) fared comparatively well. Even though only 25% of the copywriters in this study received passing rubric scores on

their writing tests, 85% of the writing samples they produced on the job received passing rubric scores. An experience-based reading of writing tests seems to be more predictive that copywriters can reach a threshold of performance, rather than predicting their actual rate of consistent performance. However, stakeholders in the workplace (mainly managers, like the ones I spoke to in my focus group) often feel nervous about relying on unstandardized and unquantifiable hiring practices, which is one of the reasons I sought to test an assessment method that would be easily standardized and quantifiable. However, in this case at least, relying on a strict rubric reading of the writing tests would sacrifice validity for agreement, as Elbow theorized. While stakeholders might like to be able to point to a rubric that ensured that all applicants were measured by the same standard, that standard does not appear to adequately measure the copywriter behind the text, nor does it assess the copywriter's abilities beyond her performance on that particular test.

However, although experience-based writing test assessment may have proven to be fairly predictive in this case, I recognize that abandoning the holistic value of agreement is not feasible in a workplace setting, especially after the hiring process has ended and copywriters must perform on a daily basis. Agreement is essential to stakeholders who have less experience with writing as a discipline of study. Managers require consistent standards to hold employees accountable for their performance. These standards ensure quality control for clients who pay for the texts. Agreement also helps the very copyeditors who complain about inconsistent writers. When both parties, copyeditors and copywriters, agree to a standard by which they are judged, there is far less conflict when a copyeditor sends a writing task back to a copywriter for revision. For freelance writers who are paid by completed tasks rather than by hour, consistency is even more important. Vague institutional standards could result in inconsistent pay. Agreement and

consistency are therefore values that must be upheld in the workplace, even if they do not perfectly measure writing ability.

Current hiring practices in many workplaces do not solely rely on writing tests to make hiring decisions. They gather resumes and writing samples and conduct interviews in addition to administering writing tests. Most companies already recognize that hiring is a high-stakes situation and that they must gather as much information about candidates as possible before making a hiring decision. “High Hits” was no different: hiring managers invested significant time and energy into evaluating potential employees. For this study, I evaluated the writing tests outside of the context of the other information hiring managers gathered to evaluate copywriter candidates. It is likely true that hiring, like writing, is a highly complex process that is difficult to assess.

Limitations of the Present Study

This study is limited by its small scope. I was only able to secure ten participants out of twenty potential participants. These ten participants also have self-selection bias. The copywriters knew me personally from my time working at the company and trusted me not to reveal personal information to the company. It is probable that my personality and behavior as an employee and manager at the company affected the participants’ decision to give consent to my study.

In addition, I only had access to twelve months of data from most of the participants. Some participants left the company before they could create twelve months of data and another participant was rehired as a freelancer in the middle of the study period, which resulted in uneven numbers of writing samples per copywriter. The inconsistencies in number of samples made the statistical analysis more complicated to perform and less certain in its conclusions.

Also, due to inconsistencies in hiring practices, two copywriters did not have writing test data available. Two other copywriters had taken the writing test twice for some reason, so I had twice as many data sets for them. Not having the writing test data for two participants likely significantly affected the statistical analysis of the variables.

If I could begin the research again, I would make a few changes to my methods. I would ask more questions about writing experience in the demographic survey. For example, in addition to asking about level of education, I would ask the participants to list their majors, minors, and past work experience with copywriting. Since I found limited evidence that training and practice improve writing proficiency, past copywriting experience could be a predicting factor in writing performance after hire.

Future Research

Based on the findings of this study, I propose that future study ought to be undertaken to test underlying theories of agreement, validity, and reliability in writing assessment. Since I found that under the circumstances in this case study, a strict rubric reading of writing tests during the hiring process was not predictive of a strict rubric reading of writing samples produced on the job, I propose a future study in which copywriter applicants have access to the company rubric when they complete the writing tests. Along with clear instructional writing prompts, access to institutional standards reveals more information to applicants about which elements of “good writing” must be present in their writing tests. It is possible that this extra information may function the same way grade norming works for raters. If a candidate has a background in writing, the rubric may serve to focus the writer on demonstrating certain aspects of writing important to the potential employer rather than simply relying on past writing experiences. In fact, it is possible that the lower scores observed for writing tests occurred

largely because of the applicants' ignorance of the institutional standards. However, in order to feasibly test the rubric as a writing test assessment tool, the rubric needs to be revised so it can be understood without additional training. The language currently used in the rubric at "High Hits" is intended for use by employees with company training in SEO copywriting and online marketing. With this change, a future study may find more useful conclusions about using a rubric reading as a method of evaluating writing tests as predictive measures of writing on the job.

I also propose changes to the rubric itself. The range in comment values within a Tier 4 rating, including a quantity of negative comment values, seems to indicate that the rubric has not clearly defined Tier 4. The difference between a passing text and a failing one appears to be incompletely defined. Copyeditors may be assigning passing scores to texts that require significant edits because the rubric forces them to or because the rubric does not provide enough justification to send a text back to the copywriter for revision. I cannot with confidence propose specific changes that ought to be made to the rubric, but a future project could involve looking more closely at the comments given to Tier 4 texts that also have negative comment values and finding patterns among the keywords. The negative keywords most frequently mentioned may indicate a place to start revising the rubric.

I also posit that this process that I have outlined could be used to assess any rubric designed for writing assessment, in the workplace or in the classroom. By comparing the numeric scores that the rubric yields with specific qualitative feedback from qualified raters, I found that the rubric involved in this project is largely successful in assigning scores. I also identified places the rubric could be improved to relieve stress on the copywriters and maintain agreement for the company stakeholders. This method, based in grounded theory, has yielded

promising results in this study and could yield similarly promising results to similar studies.

While this project failed to answer the central question of whether writing tests can ever adequately predict writing proficiency, there is hope that the question can be answered in the affirmative or negative in some future study. The methods described in this study could be replicated with a larger group of participants to yield better statistical data. To my knowledge, no other researchers have applied grounded theory to examine practical application of Elbow's empirical reading theory in writing assessment or Haswell's methods of classification. The methods used in this study ought to be repeated to see if the insights gained by this project apply to other writing situations or in other contexts, professional and academic. Since so little work is being done in workplace writing assessment, there is ample opportunity for future researchers to do groundbreaking work that also has practical application.

Works Cited

- Ackerman, John M. "Reading, Writing, and Knowing: The Role of Disciplinary Knowledge in Comprehension and Composing." *Research in the Teaching of English* 25 (1991): 133-78.
- Beaman, Ronda. "Alternative Student Assessments in Advertising Copywriting." *Annual Meeting of the Association for Education in Journalism and Mass Communication*. Minneapolis. 1-4 Aug. 1990. Presentation. ERIC.
- Bean, John C. *Engaging Ideas: The Professor's Guide to Integrating Writing, Critical Thinking, and Active Learning in the Classroom*. 2nd ed. San Francisco: Jossey-Bass, 2011.
- Berkenkotter, Carol and Thomas N. Huckin. *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*. Hillsdale, NJ: Erlbaum, 1995.
- Boettger, Ryan K. "The Technical Communication Editing Test: Three Studies on This Assessment Type." *Technical Communication*. 61.4 (Nov 2014): 215-31. *IngentaConnect*.
- CCCC Committee on Assessment. "Writing Assessment: A Position Statement." *College Composition and Communication* 46.3 (Oct. 1995): 430-37. JSTOR.
- Dias, Patrick, Aviva Freedman, Peter Medway, and Anthony Pare. *Worlds Apart: Acting and Writing in Academic and Workplace Contexts*. Ed. Charles Bazerman. Mahwah: Lawrence Erlbaum Association, Inc., 1999.
- Diederich, Paul. *Measuring Growth in English*. NCTE. ERIC. 1974.
- Downs, Douglas, and Elizabeth Wardle. "Teaching about Writing, Righting Misconceptions: (Re) Envisioning 'First-Year Composition' as 'Introduction to Writing Studies'." *College Composition and Communication* 58.4 (2007): 552-84. JSTOR.

- Elbow, Peter. "Do We Need a Single Standard of Value for Institutional Assessment? An Essay Response to Asao Inoue's "Community-Based Assessment Pedagogy." *Assessing Writing* 11.2 (2006): 81-99. *ScienceDirect*.
- Elbow, Peter and Kathleen Blake Yancey. "On the Nature of Holistic Scoring: An Inquiry Composed on Email." *Assessing Writing* 1.1 (1994): 91-107. *ScienceDirect*.
- Haswell, Richard. "Rubrics, Prototypes, and Exemplars: Categorization Theory and Systems of Writing Placement." *Assessing Writing* 5.2 (1998): 231-68. *ScienceDirect*.
- Hayes, John R., Jill A. Hatch, and Christine M. Silk. "Does Holistic Assessment Predict Writing Performance? Estimating the Consistency of Student Performance on Holistically Scored Writing Assignments." *Written Communication* 17.2 (Jan. 2000): 3-26. *SAGE*.
- Huot, Brian A. "The Influence of Holistic Scoring Procedures on Reading and Raring Student Essays." *Validating Holistic Scoring for Writing Assessment*. Ed. Michael M. Williamson and Brian A. Huot. Cresskill: Hampton P, 1993. 206-36.
- Kaufers, David S., and Richard Young. "Writing in the Content Areas: Some Theoretical Complexities." *Theory and Practice in the Teaching of Writing: Rethinking the Discipline*. Ed. Lee Odell. Carbondale, IL: Southern Illinois UP, 1993.
- Lee, Yong-Won, Claudia Gentile, and Robert Kantor. "Toward Automated Multi-trait Scoring of Essays: Investigating Links among Holistic, Analytic, and Text Feature Scores." *Applied Linguistics* 31.3 (2009): 391-417. *ERIC*.
- MacDonald, Susan Peck. "Problem Definition in Academic Writing." *College English* 49 (1987): 315-31.
- Moss, Pamela A. "Validity in High Stakes Writing Assessment: Problems and Possibilities." *Assessing Writing* 1(1) (1994): 109-128. *ScienceDirect*.

- Neff, Joyce Magnotto. "Grounded Theory: A Critical Research Methodology." *Under Construction: Working at the Intersections of Composition Theory, Research, and Practice*. Ed. Christine Farris and Chris M. Anson. Logan: Utah State UP, 1998. 125-35.
- Penny, Jim, Robert L. Johnson, and Belita Gordon. "The Effect of Rating Augmentation on Inter-Rater Reliability: An Empirical Study of a Holistic Rubric." *Assessing Writing* 7 (2000): 143-64. *ScienceDirect*.
- Petraglia, Joseph. "Introduction: General Writing Skills Instruction and Its Discontents." *Reconceiving Writing, Rethinking Writing Instruction*. Mahwah, NJ: Laurence Erlbaum, 1995. xi-xvii.
- Rezai, Ali Reza, and Michael Lovorn. "Reliability and validity of rubrics for assessment through writing." *Writing Assessment* 15.1 (2010): 18-39. *ScienceDirect*.
- White, Edward M. "The Scoring of Writing Portfolios: Phase 2." *College Composition and Communication* 56.4 (Jun. 2005) 581-600. *JSTOR*.
- Wiggins, Grant. "The Constant Danger of Sacrificing Validity to Reliability: Making Writing Assessment Serve Writers." *Assessing Writing* 1.1 (1994): 129-39. *ScienceDirect*.

Appendix One: Blog Rubric

☆ Tier 1 Rejection	Originality	Is the piece original in form and content? Does it repeat a topic already written for the client?
	Audience	Does this module address the intended audience?
	Accuracy	Does the blog contain accurate information?
	Positivity	Does the blog focus on positives rather than negatives?
☆☆ Tier 2 Rejection	Use Value	Is the blog topic relevant to the client's industry and useful to the reader?
	Controlling idea	Does the blog have a clear, singular focus? Is it cohesive?
	Tone	Does the tone reflect the client's tone requests?
☆☆☆ Tier 3 Rejection	Details	Does the blog include specific and interesting information rather than redundant, generic, mundane, or filler information?
	Organization	Is the blog organized logically and clearly? Do ideas flow well within and between paragraphs?
	Scannability	Is the blog easy to scan?
	Sentence Structures	Does the blog contain sentences that are simple and easy to understand?
	Style/Grammar/Spelling	Do sentences follow the guidelines outlined in the style guide? [Note to editor: This section should be used when issues affect readability.]
☆☆☆☆ Tier 4 Acceptable	Spelling/Idioms	Does the blog use regional spelling and terms?
	Style/Grammar/Spelling II	Is the blog grammatically perfect? Does the blog follow all style guidelines from the Style Guide? [Note to editor: This section should be used for smaller issues that occur occasionally.]
	Expertise	Does the blog sound like it was written by an expert in the field?
	Repetition	Does the blog avoid unnecessary repetition of words, phrases, or sentence structures?
	Unique Angle	Does the blog take a new or unique angle?
☆☆☆☆☆ Tier 5 Perfect	Wow factor	The blog is exceptionally memorable, creative, and interesting to the intended audience.

Appendix Two: Copy Rubric

☆ Tier 1 Rejection	Originality	Is the piece original in form and content?
	Audience	Does this module address the intended audience?
	Accuracy	Does the copy contain accurate information?
	Positivity	Does the copy focus on positives rather than negatives?
	Missing information	Does the copy contain all relevant information provided by the client?
☆☆ Tier 2 Rejection	Advertising restrictions	Does the copy follow relevant advertising restrictions?
	Controlling idea	Does the copy have a clear, singular focus? Is it cohesive?
	Tone	Does the tone reflect the client's tone requests?
	Keywords	Does the copy use best keyword practices?
☆☆☆ Tier 3 Rejection	Details	Does the copy include specific and interesting information rather than redundant, generic, mundane, or filler information?
	Organization	Is the copy organized logically and clearly? Do ideas flow well within and between paragraphs?
	Scannability	Is the copy easy to scan?
	Sentence Structures	Does the copy contain sentences that are simple and easy to understand?
	Style/Grammar	Do sentences follow the guidelines outlined in the style guide? [Note to editor: This section should be used when issues affect readability.]
	Features/Benefits	Does the piece strike a good balance between the features the client offers and the benefits experienced by the audience?
	Conversion Funnel	Does the copy lead the reader to the next step of the conversion funnel?
☆☆☆☆ Tier 4 Acceptable	Spelling/Idioms	Does the copy use regional spelling and terms?
	Style/Grammar/Spelling II	Is the copy grammatically perfect? Does the copy follow all style guidelines from the Style Guide? [Note to editor: This section should be used for smaller issues that occur occasionally.]
	Expertise	Does the copy embody the persona described by the client? Does it sound like it was written by an expert in the field?
	Repetition	Does the copy avoid unnecessary repetition of words, phrases, or sentence structures?
☆☆☆☆☆ Tier 5 Perfect	Wow factor	The copy is exceptionally memorable, creative, and interesting to the intended audience.

Appendix Three: Keyword Codes

Keyword	Tier
accuracy	1
audience	1
audience multiple	1
audience wrong	1
client request	1
follow instructions	1
links	1
originality	1
positivity	1
relevance	1
advertising restrictions	2
changed meaning	2
controlling idea	2
didn't fix problem	2
keywords	2
keywords inappropriately combined	2
keywords incorrect	2
keywords missing	2
keywords pluralsingular	2
keywords stuffing	2
keywords unnatural	2
tone	2
use value	2
wordcount	2
wordcount bloated	2
wordcount insufficient	2
acceptable but not great	3
benefits	3
clarity	3
client branding	3
cohesion	3
conversion funnel	3
CTA	3
details	3
details obvious	3
differentiation	3
headings	3
logical	3

organization	3
page requirements	3
readability	3
scannability	3
enjoyable	4
expertise	4
good sentences	4
grammar	4
interesting	4
punctuation	4
specific	4
spelling	4
style	4
style awkward wording	4
style clarity	4
style cliches	4
style concision	4
style confusing	4
style fluff	4
style gender neutrality	4
style generalization	4
style get to the point	4
style inappropriate absolutes	4
style offensive	4
style parallelism	4
style passive voice	4
style redundancy	4
style run ons	4
style salesy	4
style subjectverb agreement	4
style transitions	4
style vague pronouns	4
title	4
typos	4
unique angle	4
wrong comma placement	4
creative	5
memorable	5
wow factor	5

Appendix Four: Copywriter Writing Test

BLOG PROMPT

Our content writers often write for blogs and websites that target a specific audience. Here are some examples:

<http://www.daveywaveyfitness.com/> — targets bodybuilders, mainly men
<http://livingprettynaturally.com/> — targets organic shoppers and health-conscious women
<http://www.livingwellspendingless.com/> — targets homemakers and moms on a budget

Choose one of the sites above and write an article for their blog. You can choose your topic, but remember to keep your audience in mind—the best blog posts include action items and are inherently interesting or directly relevant to your audience. You are welcome to format your answer with italics, bullets, bold text, and other formatting markup.

Make sure to:

Indicate which website you have chosen to write for. Include the keyword “neck pillow” into the content once — make it relevant. Write between 250–350 words to complete the post. Give your post a compelling title. Avoid writing in first person (this should be an informative, not a personal, blog post)

Please do not submit content previously written for another project or website. We are looking for original work.

COPY PROMPT

The copy for Transcend Dental is not as effective as it could be. Pediatric dentists Dr. Fronk and Dr. Madsen have asked that you revise the content to be more friendly, useful and SEO-optimized. They believe that if the site makes a better impression on their website visitors, the dentistry will attract more patients.

Here is the current text:

“Welcome to Transcend Dental. Our goal is to exceed your expectations. Specialists in Pediatric Dentistry: our doctors provide comprehensive care for infants, children and teens.

We are dedicated to our patients and dedicated to our community. Our brand new state of the art facility is specifically designed for pediatric dentistry with an emphasis on comfort and efficiency for patients, parents and staff. The office is equipped with the latest technologies including digital X-ray radiographs and electronic records which allow for safer procedures and convenient access. We are a prevention focused practice with an emphasis on patient education. We will strive to build a long lasting positive relationship with your child if given the opportunity and privilege to be your pediatric dentistry provider.”

Other information you learn after visiting the office: The dentists most regularly treat children, but they also see teenagers, adults, and special needs patients. There are television sets in the

ceiling that can be controlled by the patient in the dental chair. The office accepts most insurance plans. The lobby has a large play area with toys. The office phone number is 867-5309.

Make sure to:

Create a compelling headline. Rewrite the main body copy. You can repurpose any information currently in the content and employ any other rhetoric (emotional or logical) that will persuade readers to choose Transcend Dental. Remember to speak as the company, using a kind and professional tone. Include the keyword “family dentists in Townville” twice. Include a strong “call to action.” Write between 150–250 words to complete the copy.

Appendix Five: Copywriter Demographic Survey

1. What is the highest level of school you have completed? If currently enrolled, please choose the highest degree received.

- Less than high school
- High school graduate
- Some college
- 2-year degree
- 4-year degree
- Professional degree
- Doctorate

2. Which of the following categories best describes your employment status? Choose all that apply.

- Employed full time
- Employed part time
- Unemployed looking for work
- Unemployed not looking for work
- Retired
- Student
- Disabled

3. How much total combined money did all members of your HOUSEHOLD earn in 2015? This includes money from jobs; net income from business, farm, or rent; pensions; dividends; interest; social security payments; and any other money income received by members of your HOUSEHOLD that are EIGHTEEN (18) years of age or older. Please report the total amount of money earned--do not subtract the amount you paid in taxes or any deductions listed on your tax return.

- Less than \$10,000
- \$10,000 - \$19,999
- \$20,000 - \$29,999
- \$30,000 - \$39,999
- \$40,000 - \$49,999
- \$50,000 - \$59,999
- \$60,000 - \$69,999
- \$70,00 - \$79,999
- \$80,000 - \$89,999
- \$90,00 - \$99,999
- \$100,000 - \$149,999
- More than \$150,000

4. Please specify your ethnicity.

- White

- Black or African American
- American Indian or Alaska Native
- Asian
- Native Hawaiian or Pacific Islander
- Other

5. What is your marital status?

- Married
- Widowed
- Divorced
- Separated
- Never married

6. How did you find out about this copywriting position? Choose all that apply.

- Heard about it from a “High Hits” employee
- Found it through the “High Hits” website
- Job fair
- Recruited through university connections
- Online job boards

7. If you are not a current employee, what were your reasons for leaving this company? Choose all that apply.

- Skills not effectively put to use
- Difficulty accessing the resources needed to do the job well
- Conflict with supervisors
- Pay was insufficient for the work performed
- Contributions were not recognized
- Expectations were not clear
- Expectations were not realistic
- Work-related stress
- Discomfort in work environment
- Let go
- Still a current employee

Appendix Six: Participant Recruitment Email

Dear Present and Former “High Hits” Employees,

I'm emailing you to let you know about an opportunity to participate in an academic study by one of our former employees, Lindsay Tanner. Lindsay is in grad school now at BYU pursuing a master's degree in Rhetoric and Composition. She's chosen to do her thesis project on writing tests and she wants us to participate!

Attached is a link to a simple survey (it should take 5 minutes) and a consent form that states you are willing to let independent raters look at (1) your writing test and (2) a random sampling of your work from September 2015-September 2016.

No one but Lindsay will have access to any identifying information about who wrote what. And she won't be releasing any information she finds to “High Hits” (except for her published thesis, which you can all read if you want). Read through the whole consent form at the beginning of the survey to find out more about her study and your role in it.

Lindsay will be providing a \$10 gift card as a thank you for participating in the study. The survey portion should only take about 5 minutes. At the end of the survey, you will be asked for your email address. Your thank you gift card will be emailed to you within 3 business days.

<link to survey>

Use this password to access the survey: xRjG847

Thank you for all your hard work!

Shannon Williams

Appendix Seven: Informed Consent for Focus Group Participants

Introduction

This research study is being conducted by Lindsay Tanner, second year masters' student at Brigham Young University. The project is being completed under the mentorship of Professor Jon Balzotti at Brigham Young University. The purpose of the study is to determine whether there is a connection between the scores of writing tests provided before hiring and the scores of work produced after hiring. You were invited to participate because you are a stakeholder in the professional copywriting field.

Procedures

If you agree to participate in this research study, the following will occur:

You will participate in a focus group with other stakeholders in professional copywriting. The focus group should take approximately one (1) hour.

The focus group will be audio recorded to ensure accuracy in reporting your statements.

The focus group will take place in "High Hits"'s offices at a time convenient for you or it will take place at a time and location convenient for you.

This study involves no time commitment.

Risks/Discomforts

There are virtually no potential risks involved in this study. You may feel embarrassment about your work being qualitatively and quantitatively evaluated by scorers whom you do not know.

However, the researcher aims to minimize discomfort by anonymizing the writing samples. Your name will never appear on any writing you have produced; instead, your work will be given a numerical code. No quotations from your work will appear in the final study. The study will only publish metadata derived from your samples.

Benefits

The results of the study will provide no direct benefit to you. It is hoped, however, that through your participation researchers will learn more about professional literacy. This study may help universities determine curricula to help students who aim to work in professional content creation.

Confidentiality

The research data will be kept on password-protected computer and only the researcher will have access to the data. At the conclusion of the study, all identifying information will be removed and the data will be kept in the researcher's locked cabinet.

Compensation

Participants will receive a \$10 gift card as a thank you for participating in the study.

Participation

Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your employment status, performance, or standing with the company.

Questions about the Research

If you have questions regarding this study, you may contact Lindsay Tanner at lindsay_tanner@byu.edu for further information.

Questions about Your Rights as Research Participants

If you have questions regarding your rights as a research participant contact IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent

I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Name: _____ Signature: _____ Date: _____

Appendix Eight: Informed Consent for Copywriter Participants

Consent to Be a Research Subject

Introduction

This research study is being conducted by Lindsay Tanner, second year masters' student at Brigham Young University. The project is being completed under the mentorship of Professor Jon Balzotti, from the Department of English at Brigham Young University. You are being invited to participate in this research study of professional writing tests. The purpose of the study is to determine whether there is a correlation between the scores of writing tests provided before hiring and the scores of work produced after hiring. You were invited to participate because you are a professional copywriter who completed a writing test before being hired and produced many more samples during your employment.

Procedures

- If you agree to participate in this research study, the following will occur:
- You will fill out a short survey that should take no more than 5 minutes to complete.
- The writing test you submitted to be hired by "High Hits" will be graded by independent raters.
- A random sampling of your written work (copy and blogs from August 2015–August 2016) will be graded by independent raters.

No identifying information will be released to the independent scorers. All names of authors will be removed from every blog and piece of copy before they are scored. All identifying information will be placed in a password-protected computer only the primary researcher has access to.

Risks/Discomforts

There are very few risks involved in this study. You may feel embarrassment about your work being qualitatively and quantitatively evaluated by raters whom you do not know.

However, the researcher aims to minimize discomfort by de-identifying the writing samples. Your name will never appear on any writing you have produced; instead, your work will be given a numerical code. No quotations from your work will appear in the final study. No identifying information of any kind will be published. The study will only publish general conclusions drawn from all participants' samples.

Benefits

The results of the study will provide no direct benefit to you. It is hoped, however, that through your participation researchers will learn more about professional literacy. This study may help universities determine curricula to help students who aim to work in professional content creation.

Confidentiality

The research data will be kept on password-protected computer and only the researcher will have access to the data. At the conclusion of the study, all identifying information will be removed and the data will be kept on an external storage device in the researcher's locked cabinet. The data will be kept secure for a period of 3 years, after which it will be destroyed.

No information from this study will be released to “High Hits”. The company will have access to the published results of the study, but no other information will be shared.

Compensation

Participants will receive a \$10 gift card as a thank you for participating in the study.

Participation

Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your employment status, performance, or standing with the company.

Questions about the Research

If you have questions regarding this study, you may contact Lindsay Tanner at lindsay_tanner@byu.edu for further information. You can also contact Lindsay's faculty mentor, Professor Jon Balzotti, at (801) 422-2440; 4127 JFSB, Brigham Young University, Provo, UT 84602; jonathon_balzotti@byu.edu.

Questions about Your Rights as Research Participants

If you have questions regarding your rights as a research participant contact IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent

I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Type your signature here:

Appendix Nine: Student Rater Job Requirements

Job title: Research Assistant/Editor

Category: Academic

Openings: 3

Start Date: 09/05/16

Shift: 5 hours/week, flexible schedule

Hourly Wage: \$15/hour

Description: Research assistants/editors will read, score, and provide substantive edit commentary for professional writing samples. The editors' scores and commentary will be analyzed as part of a graduate student's thesis project. The job provides excellent opportunity for editing in a real-world scenario. Exceptional work could lead to a recommendation for full-time employment following graduation. This job is limited to the Fall 2016 semester, but there may be an opportunity for more work Winter 2017 semester.

Skills Needed:

- Experience meeting deadlines
- Ability to self-direct and self-monitor progress
- Excellent substantive and copyediting skills
- Solid understanding of grammatical principles
- Ability to adjust editing style to different style guides
- Ability to accept and give constructive criticism
- Knowledge of web-writing principles
- Knowledge of SEO and internet marketing
- Understanding of buyer psychology and best copywriting practices

Qualifications: Editing minor, preferably completed or taking at least ELang 350 or ELang 410.

Application Instructions: Apply online. Please attach a cover letter, resume, and writing sample to your application. Applications without attached documents will not be considered.

Appendix Ten: Student Rater Grammar Test

Total Questions: 30

This test will help us assess your basic grasp of the English language. The quiz is timed and you have 15 minutes to complete the quiz.

Name: _____

1. Choose the correct sentence
 - a. The street vendor has many oranges.
 - b. The street vendor has much oranges.

2. Punctuate this sentence correctly: We wanted to go to the game however we couldn't because it was snowing.
 - a. We wanted to go to the game however, we couldn't because it was snowing.
 - b. No change
 - c. We wanted to go to the game; however, we couldn't because it was snowing.
 - d. We wanted to go to the game, however, we couldn't because it was snowing.

3. Which is incorrect?
 - a. The rickety old bridge has collapsed, so we won't be able to cross it.
 - b. The rickety old bridge has collapsed, we won't be able to cross it.
 - c. The rickety old bridge has collapsed; we won't be able to cross it.
 - d. The rickety old bridge has collapsed—we won't be able to cross it.

4. Which sentence is correct?
 - a. She is married to him.
 - b. She is married with him.

5. In the following sentence, which word is spelled incorrectly? I feel like your not telling me the truth about what happened to their toys.
 - a. their
 - b. to
 - c. your

6. Select the answer that is the best version of the underlined part of the sentence. If the sentence should not be revised, select "No change needed." Working as a receptionist for my uncle's business taught me to handle customer complaints, answering the telephone, and how to maintain a Rolodex.
 - a. How to handle customer complaints, to answer the telephone, and maintaining a Rolodex.

- b. No change needed
 - c. Handling customer complaints, answering the telephone, and maintenance of a Rolodex.
 - d. How to handle customer complaints, answer the telephone, and maintain a Rolodex.
7. Choose the correct sentence.
- a. Dave invited Bob and I to the party.
 - b. Dave invited Bob and me to the party.
8. Choose the best revisions of the following sentence. If the sentences should not be revised, select "No change needed." We could watch the stars sitting on the balcony.
- a. We could watch, sitting on the balcony, the stars.
 - b. Sitting on the balcony, we could watch the stars.
 - c. No change needed.
 - d. On the balcony, we could watch the stars sitting.
9. Which best completes the following sentence? _____ car is parked outside?
- a. Whos'
 - b. Whose
 - c. Whos
 - d. Who's
10. Which sentence is correct?
- a. He was born 2001.
 - b. He was born in 2001.
 - c. None of these is correct.
 - d. He was born on 2001.
11. Choose the best sentences from the options below.
- a. For the young and aspiring climbers, the steep silhouette of Everest was aspiring.
 - b. For the young and aspiring climbers, the steep silhouette of Everest was respiring.
 - c. The steep silhouette of Everest inspired the young and aspiring climbers.
 - d. The steep silhouette of Everest respired the young and aspiring climbers.
12. Fill in the blank. I enjoy this flavor of ice cream _____.
- a. a lot
 - b. a lot
 - c. much
 - d. very
13. Which best completes the following sentence? Children should not be _____ to say mean

things _____.

- a. allowed; aloud
- b. aloud; allowed
- c. aloud; aloud
- d. allowed; allowed

14. Which punctuates the following sentence correctly? On November 11 1918 the armistice ending World War I was signed.

- a. On November 11 1918, the armistice ending World War I was signed.
- b. On November 11, 1918, the armistice ending World War I was signed.
- c. On November, 11, 1918, the armistice ending World War I was signed.
- d. On November 11, 1918 the armistice ending World War I was signed.

15. Which best completes the following sentence? Whistler's mother has been _____ in bed all morning.

- a. laying
- b. lieing
- c. lying

16. Which sentence is correct?

- a. I liked her, she was kind and honest.
- b. I liked her because she was kind, and honest.
- c. I liked her because she was kind and honest.

17. Which best completes the following sentence? It rained a lot more _____ it did yesterday, stopping _____ starting again.

- a. then; than
- b. then; then
- c. than; than
- d. than; then

18. Select the best version of the underlined part of the sentence. If the sentence should not be revised, select "No change needed." Coming in from Fourth Street, the Student Union is seen by students who are arriving on campus.

- a. the students having seen the Student Union.
- b. Student Union is being seen by students who are arriving on campus.
- c. No change needed.
- d. students arriving on campus see the Student Union.

19. Choose the best form of the underlined part of the sentence: After it was repaired, it ran perfect again.

- a. could of run perfectly again
 - b. ran perfectly
 - c. could have run perfect
 - d. No change needed
20. The cat has ____ the canary.
- a. eated
 - b. ate
 - c. eat
 - d. eaten
21. Pick the word that best completes the sentence: Lying cheating, and stealing often have a negative ____ on election results.
- a. affect
 - b. effect
22. Microsoft announced ____ releasing a new product each week.
- a. they are
 - b. its
 - c. it is
 - d. she is
23. Which is the correct use of the semicolon?
- a. Mary is bringing sandwiches; and chips, I think.
 - b. Don't worry about bringing sandwiches; Mary is bringing them.
 - c. I hope Mary is bringing sandwiches; because I'm not bringing any.
 - d. Don't worry, Mary is bringing sandwiches; chips; and drinks.
24. Fill in the blank: Wow she writes really ____!
- a. great
 - b. well
 - c. excellent
 - d. superb
25. Which best completes the following sentence? The committee ____ these questions carefully.
- a. debate
 - b. debates
26. Chairs ____ don't have cushions are uncomfortable to sit on.
- a. that

- b. what
- c. which
- d. whose

27. Which best completes the following sentence? She was not in the basement, ____ was she upstairs.

- a. but
- b. or
- c. so
- d. nor

28. Fill in the blank with the correct word. The story of Little Red Riding Hood is ____ my favorite.

- a. definitely
- b. definately
- c. definatley

29. Which best completes the following sentence? Come with us ____ I can introduce you to my friends!

- a. thus
- b. so that

30. Choose the best option to fill in the blanks. ____ hard to describe the painting. ____ colors are striking, though.

- a. It's, Its
- b. It's, It's
- c. Its, Its
- d. Its, It's