Cathy W. Leist,
Mark A. Woolwine,
Cathy L. Bays

The Effects of Using a Critical Thinking Scoring Rubric to Assess Undergraduate Students' Reading Skills

The purpose of this study was to investigate the use of a critical thinking rubric as an assessment of reading achievement for students enrolled in a reading intervention course. A reading prompt and scoring rubric, based on Richard Paul and Linda Elder's critical thinking framework, were created to assess critical reading in an intervention course. The prompt and rubric were used throughout the semester to provide formative reading assessment. The scoring rubric, which is responsive to reading as a cognitive process, was also used for precourse and postcourse assessment to provide a unique measure of reading achievement that incorporates the university's critical thinking initiative. A repeated measures design was used to assess the reading achievement of 164 students on five different reading prompts over the course of a fall semester. Results showed significantly higher postassessment rubric scores ($p < .001$) and a significant change in rubric scores over time ($p < .05$).

The search to measure reading achievement has an extensive history. Numerous standardized tests have been developed through the years to identify and measure specific reading skills, and research has been conducted to examine the validity of these tests. However, no single standardized reading test or informal

reading assessment is currently viewed to be the most effective means to assess college students' reading ability (Flippo & Schumm, 2009). A concern for designers of college reading instruction is that standardized tests used to assess students' reading achievement following college reading instruction may not measure a student's ability to read college-level reading assignments. Researchers who view reading as a strategic process have suggested that a student's mastery of reading subskills and improved performance on a standardized test may not demonstrate development of the college reading skills that contribute to academic success (Nist & Simpson, 2000; Simpson, Stahl, & Francis, 2004).

Designers of college reading instruction also need reading tests to assess the development of reading skill over the course of a semester. To be effective, a reading course needs to provide instruction with authentic reading materials that enable students to reflect on and evaluate their reading, studying, and learning with the goal of controlling their own metacognitive processes (Simpson, Stahl, & Francis, 2004). Reading courses need to explore innovative ways to help students improve their critical reading skills for success in college. Process-based approaches to reading and study strategies based on cognitive theory suggest that students need instruction in particular strategies to assist them to become successful students, including "how to select, transform, organize, elaborate, plan, monitor, and evaluate all critical thinking processes" (Nist & Simpson, 2000, p. 659), and the results of this instruction on learning should be assessed. Metacognitive assessments that promote reflection and critical thinking have been suggested as important to include in the college classroom and to consider for additional research (Flippo, Becker, & Wark, 2009). Reading assessments that help students determine an intentional purpose for reading and include specific criteria for evaluating the reader's written responses may better indicate reading development than standardized tests (Farr, 2003). Individuals charged with choosing appropriate reading assessments should, therefore, consider the use of discipline-specific reading materials to provide authentic reading tasks. Discipline-specific reading materials from linked content courses may be appropriate to use to improve comprehension skills, even for developing college readers (Cox, Friesner, & Khayum, 2003). However, Flippo and Schumm (2009) indicated that little empirical evidence has been documented for the effectiveness of portfolios, authentic assessments, or performance-based assessments.

This study sought to investigate how to assess the reading achievement of undergraduate students in a reading intervention course paired with general education courses. Reading achievement was measured by a required standardized reading test but additionally measured during the

semester and at the end of the semester by an alternative assessment technique, a written response evaluated by a scoring rubric. This study contributes empirical evidence that supports the need to continue investigation of alternative assessment techniques in college reading courses.

Background of the Study

Required Statewide Assessment in Reading

The National Governors Association Center for Best Practices and the Council of Chief State School Officers created a national initiative that resulted in the development of a set of educational standards referred to as the Common Core State Standards. These educational standards are intended to be a framework for preparing students for college and the workforce. These standards define the knowledge and skills that students should learn and develop in high school to be ready to succeed in entry-level, credit-bearing academic college courses and in workforce training programs. Each state that has adopted these standards is expected to create its own process for development and implementation. As a result, what students are expected to learn can vary widely from state to state. Kentucky is one of 45 states that have adopted these standards and is reforming curriculum in secondary schools to better align with entry-level curriculum in colleges and to assess college readiness (National Governors Association, 2010).

To assist the implementation of the Common Core Standards, Kentucky state law defines a standard for college readiness in reading for public colleges and universities. All degree-seeking, first-year students with fewer than 24 credit hours admitted with ACT reading scores less than 20 are required to participate in reading placement testing and course registration as determined by their reading placement results (Kentucky Developmental Education Task Force, 2007). To comply with state law, the university where this study was conducted selected the COMPASS Reading Test as a placement exam. An intervention course, Special Topics in Supplemented College Reading (GEN 105), was created in 2009. This first-year course is offered for one credit hour. The institution's centralized academic support unit instructs course sections of GEN 105 each academic year.

Students required to enroll in a course section of GEN 105 must concurrently enroll in a selected section of a general education content course linked to the specific GEN 105 section. The content courses offered are in biology, psychology, and history. Students enrolled in GEN 105 are required to attend both the general education lecture course each week and two additional classroom sessions of GEN 105. In the GEN 105 class

sessions, students receive instruction in college reading, critical thinking, and study strategies focused on the paired course. The textbook and supplemental readings assigned in each of the content courses serve as the basis for reading assignments and instruction in GEN 105. Grading for GEN 105 is pass or fail and appears as a separate course grade on the transcript. Reading competencies addressed in GEN 105 include the following: identify accurately the major purpose or concepts presented in a content reading; demonstrate a clear understanding of facts, data, or examples that support the purpose or concepts presented in a content reading; define precisely content-specific vocabulary explicit in a content reading; develop well-reasoned and relevant conclusions based on content-specific information; and apply concepts from content reading broadly to other readings or non-reading contexts.

To satisfy the university's need to participate in statewide assessment of students' reading achievement, the COMPASS Reading Test is administered at the beginning and end of the course. Placement (precourse) scores and end-of-semester (postcourse) scores are recorded and maintained as a standardized measure of the reading achievement of each student. These reading scores are used for institutional compliance reports.

Scoring Rubric for Reading Assessment

A different form of assessment, a reading prompt and scoring rubric, was constructed to assess students' critical reading in GEN 105, using the Paul and Elder framework of critical thinking applied to reading (Paul & Elder, 2009). Critical thinking is a key component of this institution's quality enhancement project for its next accreditation review. The university created the project and the task group to design instruction in critical thinking for faculty and staff and guide the curriculum revision process over the next ten years in all undergraduate units. The task group initially reviewed 13 different models of critical thinking. The Paul and Elder (2009) framework was unanimously adopted by the task group because of its comprehensiveness, applicability across disciplines, and high-quality resources. The designer of this reading intervention course chose to create a form of reading assessment that applied the university-selected critical thinking framework from a metacognitive perspective.

Graduate student assistants are trained to instruct GEN 105. These instructors assign enrolled students to read and respond in writing to five discipline-specific readings selected from the content course textbook and supplemental course reading materials during the semester. The instructors assign a specific reading prompt to elicit students' written responses and evaluate each student's written response using a scoring

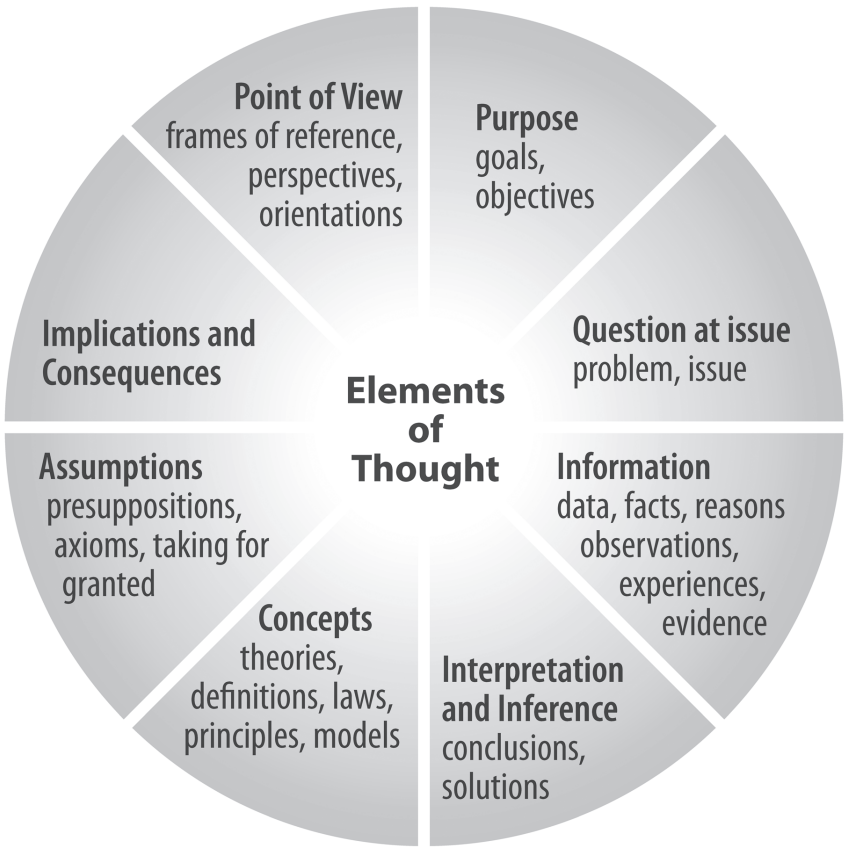
rubric based on the critical thinking framework. The reading prompt and scoring rubric focus students' attention on the development of their critical thinking skills. The scoring rubric is used as a formative assessment, providing detailed feedback from the instructor related to students' levels of critical reading achievement at specified intervals during the semester (Stevens & Levi, 2005). These readings are assigned and scored about every two weeks throughout the semester, excluding the first and last weeks of the 14-week semester. Each time the instructor returns a scoring rubric with comments, students can evaluate their ability to comprehend discipline-specific reading material and note their ongoing improvement or lack of improvement toward the highest level of critical reading achievement. The written responses and scoring rubrics are collected in an ungraded student portfolio to illustrate each student's efforts at the end of the semester (Stevens & Levi, 2005). The initial reading prompt and scoring rubric were reviewed and revised following the first year of the intervention. Data reported result from the revised prompt and rubric used in the second year of the intervention course.

The scoring rubric also serves as a program assessment of students' reading achievement at the end of the semester. During the first week of the semester, each student produces a written response to a reading prompt related to a discipline-specific reading excerpt from the required content course textbook. The student's written response is scored by the rubric and serves as a precourse assessment for students and a diagnostic test for instructors. During the last week of the semester, each student produces a written response to a reading prompt related to the same discipline-specific reading excerpt. The written response is then scored using the rubric and serves as a postcourse test. Comparisons of precourse and postcourse writings and of precourse and postcourse COMPASS Reading Test scores measure reading development. The COMPASS scores also provide standardized assessment data required to satisfy statewide assessment in reading, while the writing-task scores provide assessment data more representative of the reading and critical thinking tasks necessary for college courses ("Assessing Learning Outcomes," 2010).

Paul and Elder Critical Thinking Framework

The Paul and Elder (2009) framework uses a specific method of analysis and evaluation to instruct students in how to improve their critical thinking skills. This methodology provides a common framework identifying the *Elements of Thought*, which can be applied to a set of *Universal Intellectual Standards* with the goal of developing the *Intellectual Traits*. Paul and Elder illustrated the Elements of Thought using a circle diagram as shown in Figure 1 (Paul & Elder, 2009, p. 3).

Figure 1
The Elements of Thought



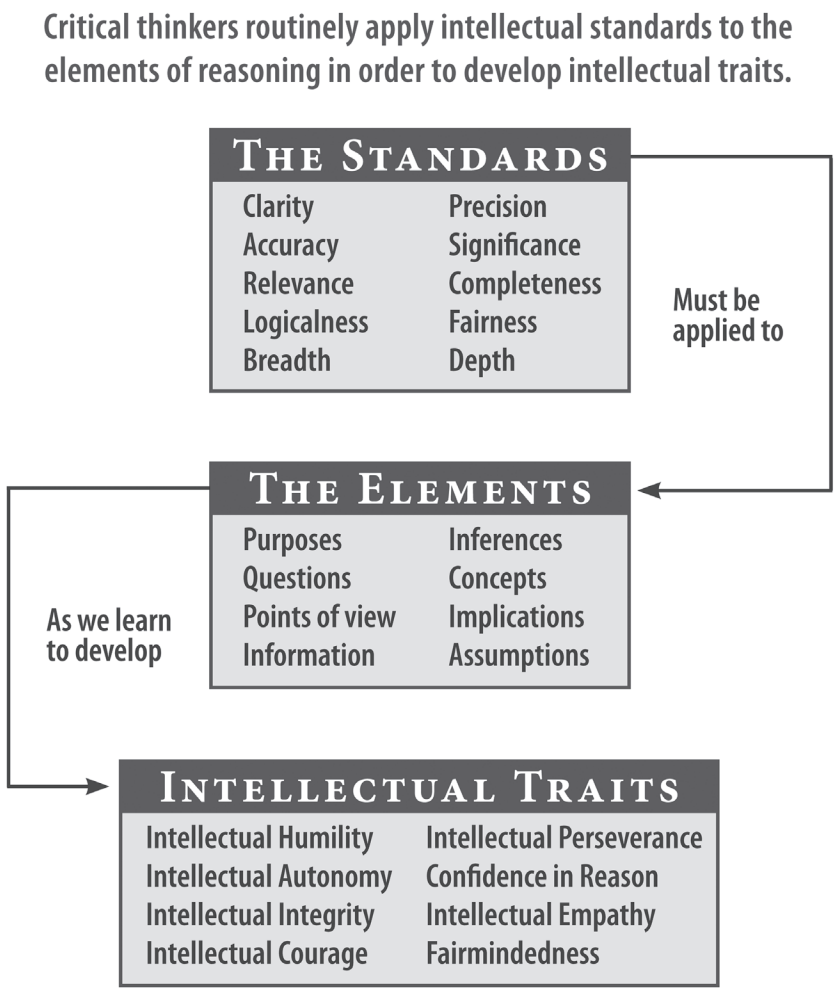
The circle diagram presents components of all reasoning: purpose, question at issue, information, interpretation and inference, concepts, assumptions, and implications and consequences. Students can improve their critical thinking by analyzing their thinking using these elements. Students can create questions from each of the elements to better direct and focus their thinking, for example, "What is the purpose of what I am doing? What is the main question that I need to answer? What information do I need to answer that question?" Using this method, students learn to use a common vocabulary to focus their thinking, monitor their thinking, and correct faulty thinking.

Students can apply the Universal Intellectual Standards to the Elements of Thought to evaluate the quality of their critical thinking. This application is illustrated in Figure 2 (Paul & Elder, 2009, p. 19). When students learn to apply the standards of clarity, accuracy, relevance, logic, breadth, precision, significance, completeness, fairness, and depth to the Elements of Thought, they make judgments about the quality of their own thinking. According to Paul and Elder (2009), the goal is to develop particular Intellectual Traits: intellectual humility, intellectual autonomy, intellectual integrity, intellectual courage, intellectual perseverance, confidence in reasoning, intellectual empathy, and fair-mindedness. The development of these Intellectual Traits is essential to intellectual inquiry and warrants intentional instruction in critical thinking skills (Paul & Elder, 2009).

Paul and Elder Critical Thinking Framework Applied to Reading

Paul and Elder (2008) applied their framework of critical thinking to the reading process. Students who use good critical reading skills approach reading assignments based on the type of material read (Paul & Elder, 2008). For example, a good critical reader reads a biology textbook differently than a novel or magazine article. However, a good critical reader also recognizes the need to develop and apply general critical reading skills that are applicable to any type of reading because many different types of reading tasks may be assigned in college. Paul and Elder (2008) suggested that students need to have a purpose for reading that considers the writer's purpose for writing. Students who are good critical readers read from paragraph to paragraph monitoring the author's thinking. During the reading process, students need to distinguish between their own thinking and the thinking of the author. For example, while students may read for their own purposes, using their own concepts, ideas, assumptions, inferences within their own personal points of view, they must also read "to recognize that embedded in the text is the author's purpose, the author's question, assumptions, concepts and so forth" (Paul & Elder, 2008, p. 8).

Figure 2
The Paul-Elder Critical Thinking Framework



Students can reconstruct the author's thinking by reflecting on their own personal experience of reading and recognize the interaction between reader and author. The Elements of Thought can serve as tools to help students reconstruct the author's thinking. Students who are good critical readers can reason through each of these elements for their own purposes, using their own concepts, ideas, assumptions, inferences, and personal points of view, but can also seek to recognize and understand the author's purpose, ideas, assumptions, inferences, and point of view. This method provides a means for helping students improve their critical reading by reflecting on and evaluating their understanding, and by recognizing their own metacognitive processes (Paul & Elder, 2008). This critical reasoning can be demonstrated by asking students to write about what they understand from reading.

Prompt for Reading Assessment

The reading prompt is constructed to guide students through the reading and writing tasks. The reading prompt gives detailed directions to students and serves as the task description essential to evaluation by rubric (Stevens & Levi, 2005). Students are instructed to apply the prompt to the content reading from the linked content course (either an assigned section of the textbook or a supplemental reading directly related to a textbook topic). The reading prompt asks students to mark and annotate the text and then to write about their understanding of the content as presented by the author. Students are instructed to write an extended summary that includes very specific information:

- the main purpose of the article (in 1 or 2 sentences);
- the key concepts (facts and data) used to support the main purpose;
- the article's application to history, biology, or psychology; and
- the article's personal relevance to the student.

The scoring rubric (Figure 3) is used to evaluate the student's written response. The rubric is based on a grid format with four parts: (a) the task description (reading prompt), (b) a scale indicating levels of achievement, (c) the categories or dimensions of the desired skills, and (d) descriptions for each level of performance (Stevens & Levi, 2005). The first vertical column in this scoring rubric provides the dimensions, which have been constructed using a combination of the Elements of Thought and the Universal Intellectual Standards from the Paul and Elder (2009) framework for critical thinking and the application of this method to reading. Each dimension represents a specific reading skill to be demonstrated and identifies each Intellectual Standard to be applied to the student's written performance reflecting the use of this reading skill.

Figure 3
Critical Thinking Rubric for Critical Reading

RUBRIC for Critical Reading

Student Name: Title of Reading: Date:

Standards and Elements	1	2	3	4	Points
ACCURACY: Identifies main purpose and/or concepts in reading	Highly inaccurate, with wrong or no purpose or concepts stated	Low accuracy, or either the purpose or the concepts stated inaccurately	Some accuracy with the purpose and concepts of the reading, but subtle inaccuracies	Complete accuracy with correct purpose and concepts clearly stated	
CLARITY: Understands the facts, data, or examples used to support	No use of facts, data, or examples from the reading	Incorrect or minimal use of the facts, data, or examples from the reading	Some correct use of the facts, data, or examples from the reading	Frequent correct use of the facts, data, or examples from the reading	
PRECISION: Identifies and uses the content-specific vocabulary from the reading or lecture	Includes no content-specific vocabulary	Low precision, an attempt to use the content-specific vocabulary, but uses incorrectly or minimally	Some precision, does incorporate content-specific vocabulary, may paraphrase correctly	Complete precision with frequent use of content-specific vocabulary, may often paraphrase correctly	
DEPTH: Demonstrates complexity of understanding	No understanding of the connections among purpose, concepts, and/or support in the reading	Limited understanding of the connections among purpose, concepts, and/or support in the reading	Generally understands the connections among the purpose, concepts, and/or support in the reading	Complex understanding of the connections among the purpose, concepts, and support in the reading	

RELEVANCE: Identifies or generates conclusion(s) and personal significance based on content	No relevance or conclusions stated	Low relevance, with basic conclusions stated	Some relevance with basic conclusions, but does not personally connect to the concepts	Complete relevance to the passage, explains several conclusions, may include personal connections to these ideas	
LOGIC: Applies concepts and content to other broad contexts	No application of concepts	Low application of concepts, or incorrect application of concepts	Some application of concepts, but uses generic ideas	Complete application of concepts and content to other contexts by using pertinent examples and supplying details	
Notes:					**TOTAL**
Based on the Richard Paul and Linda Elder critical thinking conceptualization. For more information visit http://www.criticalthinking.org/					
GEN 105 Special Topics in College Reading					

Figure 4

Student A's Written Response Assigned Low Ratings on the Critical Thinking Rubric

Student A – Response

According to "Genetic Contributions to Your Life Story," the main idea of this article is to explain how human beings work scientifically. This article also explains how the environment affects our genes—a unit of DNA on a chromosome that encodes instructions for making a particular protein molecule; the basic unit of heredity. "Packed in that tiny cell was the unique set of genetic instructions that you inherited from your biological parents. Today, that same set of genetic information is found in the nucleus of nearly every cell of your body." Genotypes have different responses to environmental factors…think of people who have a genetic predisposition to sunburn easily, such as redheads or people with very fair skin. For such people, their genotype is especially sensitive to the effects of ultraviolet rays (Last paragraph).

This article and the facts from the previous question pertain to psychology because psychology not only deals with the way people think, or deal with situations, but also why people react in situations. You cannot know why unless you know how the body/mind works.

RUBRIC for Critical Reading

Student Name PSYC Student A Title of Reading "Genetic Contributions To Your Life Story" Date 10/10/10

Standards and Elements	1	2	3	4	Points
ACCURACY: Identifies main purpose and/or concepts in reading	Highly inaccurate, with wrong or no purpose or concepts stated	Low accuracy, or either the purpose or the concepts stated inaccurately	Some accuracy with the purpose and concepts of the reading, but subtle inaccuracies	Complete accuracy with correct purpose and concepts clearly stated	**2**
CLARITY: Understands the facts, data, or examples used to support	No use of facts, data, or examples from the reading	Incorrect or minimal use of the facts, data, or examples from the reading	Some correct use of the facts, data, or examples from the reading	Frequent correct use of the facts, data, or examples from the reading	**2**
PRECISION: Identifies and uses the content-specific vocabulary from the reading or lecture	Includes no content-specific vocabulary	Low precision, an attempt to use the content-specific vocabulary, but uses incorrectly or minimally	Some precision, does incorporate content-specific vocabulary, may paraphrase correctly	Complete precision with frequent use of content-specific vocabulary, may often paraphrase correctly	**2**
DEPTH: Demonstrates complexity of understanding	No understanding of the connections among purpose, concepts, and/or support in the reading	Limited understanding of the connections among purpose, concepts, and/or support in the reading	Generally understands the connections among the purpose, concepts, and/or support in the reading	Complex understanding of the connections among the purpose, concepts, and support in the reading	**2**

RELEVANCE: Identifies or generates conclusion(s) and personal significance based on content	No relevance or conclusions stated	Low relevance, with basic conclusions stated	Some relevance with basic conclusions, but does not personally connect to the concepts	Complete relevance to the passage, explains several conclusions, may include personal connections to these ideas	**3**
LOGIC: Applies concepts and content to other broad contexts	No application of concepts	Low application of concepts, or incorrect application of concepts	Some application of concepts, but uses generic ideas	Complete application of concepts and content to other contexts by using pertinent examples and supplying details.	**3**
The purpose & key concepts are centered on a discussion of the differences between genotypes and phenotypes and how individuals are shaped by their inherited genes as well as their environment. The purpose needs to be stated clearly in the first sentence and then followed by an identification of the key concepts. When explaining the concepts, you should support them with data that explains each one. You identified genes, chromosomes, and DNA (a concept) but did not fully describe how these relate to genotypes and phenotypes and in what ways that environment influences their expression (another concept). Work on paraphrasing rather than using direct quotes. When using examples (i.e., red heads and sun burns) you should explain how their genetics interact with their environment (the sun) to cause their expression (the sun burn). By tying these concepts and examples together you demonstrate depth and understanding. When showing relevance and logic, remember to connect your reasoning back to the purpose and key concepts. I know that you read the article and have a solid grasp of psychology, so be encouraged that you can articulate this in your future summaries! Use the rubric and these comments to guide you as you read and summarize future articles. Thanks!					**TOTAL 14**
Based on the Richard Paul and Linda Elder critical thinking conceptualization. For more information visit http://www.criticalthinking.org/					
GEN 105 Special Topics in College Reading					

Figure 5

Student B's Written Response Assigned High Ratings on the Critical Thinking Rubric

Student B – Response

The main purpose of this article is to discuss the difference of genotypes and phenotypes. A genotype is the chromosomes genetically given to one from their biological parents. Phenotypes, however, are one's displayed characteristics, and are a combination of environmental factors and genes. In this article, there were a few key concepts.

The first key concept involves DNA, chromosomes, and genes. DNA, simply, holds our chemical makeup. Chromosomes are the parallel strands that encode one's chemical structure, and genes are DNA segments in chromosomes. All of these relate to each other because they all work together within each other. Every chromosome, found in a cell nucleus, has DNA, which are composed of genes. The DNA segments are called genes, which are units of DNA code. The second key concept is that the developments of living organisms are guided by genes because genes give us characteristics, which define who we are. Also, genes can mutate, which means rapidly change for no specific reason. Finally, the third key concept is that environmental factors play a role in phenotypes and genotypes. The environmental factors can make certain genotypes react certain ways. If you share the same genotype with a person, but live in separate conditions, their phenotypes may be affected differently. Environmental factors also can damage DNA, resulting in deformities such as birth defects.

This article can be applied to Psychology because it is part of the nature/nurture concept. Our chemical makeup or genotype is our nature, but how we are raised, the nurture part of the concept, is the environmental factors that affect one's phenotypes. Therefore, these two definitions are strongly related to Psychology. This article is clearly relevant to me because it discusses the genetic makeup that makes humans who they are, and more personally, as a human, who I am. It shows how you start as a person, and how your genotype and phenotype affect who you become as well as the environmental factors that surround you

RUBRIC for Critical Reading

Student Name PSYC Student B Title of Readin "Genetic Contributions To Your Life Story" Date 10/10/10

Standards and Elements	1	2	3	4	Points
ACCURACY: Identifies main purpose and/or concepts in reading	Highly inaccurate, with wrong or no purpose or concepts stated	Low accuracy, or either the purpose or the concepts stated inaccurately	Some accuracy with the purpose and concepts of the reading, but subtle inaccuracies	Complete accuracy with correct purpose and concepts clearly stated	**4**
CLARITY: Understands the facts, data, or examples used to support	No use of facts, data, or examples from the reading	Incorrect or minimal use of the facts, data, or examples from the reading	Some correct use of the facts, data, or examples from the reading	Frequent correct use of the facts, data, or examples from the reading	**4**

PRECISION: Identifies and uses the content-specific vocabulary from the reading or lecture	Includes no content-specific vocabulary	Low precision, an attempt to use the content-specific vocabulary, but uses incorrectly or minimally	Some precision, does incorporate content-specific vocabulary, may paraphrase correctly	Complete precision with frequent use of content-specific vocabulary, may often paraphrase correctly	**4**
DEPTH: Demonstrates complexity of understanding	No understanding of the connections among purpose, concepts, and/or support in the reading	Limited understanding of the connections among purpose, concepts, and/or support in the reading	Generally understands the connections among the purpose, concepts, and/or support in the reading	Complex understanding of the connections among the purpose, concepts, and support in the reading	**4**
RELEVANCE: Identifies or generates conclusion(s) and personal significance based on content	No relevance or conclusions stated	Low relevance, with basic conclusions stated	Some relevance with basic conclusions, but does not personally connect to the concepts	Complete relevance to the passage, explains several conclusions, may include personal connections to these ideas	**3**
LOGIC: Applies concepts and content to other broad contexts	No application of concepts	Low application of concepts, or incorrect application of concepts	Some application of concepts, but uses generic ideas	Complete application of concepts and content to other contexts by using pertinent examples and supplying details	**4**
Notes: Great job at clearly stating the purpose and the key concepts! You did well supporting each of the concepts with pertinent data and relevant examples (i.e. "If you share the same genotype with a person but live in separate conditions…") from the text. You also did well demonstrating how the vocabulary words (i.e. DNA, chromosomes, and genes) relate to one another. You had excellent logic with the nature vs. nurture application to psychology. You should be more personal (give an example that relates to you personally) when explaining how the text is relevant to you. You demonstrated depth by connecting the material cohesively and providing important details in a logical manner. Awesome work!!					**TOTAL 23**
Based on the Richard Paul and Linda Elder critical thinking conceptualization. For more information visit http://www.criticalthinking.org/					
GEN 105 Special Topics in College Reading					

The Intellectual Standards (accuracy, clarity, precision, depth, relevance, and logic) serve as the six dimensions of the rubric. The horizontal columns of the rubric create a rating scale to be applied to the differing levels of student achievement. Evaluators use numbers 1 (lowest) through 4 (highest) to assess students' written performance. For each of the dimensions applying the six Intellectual Standards, categories provide a description for each potential level of performance.

Training and Scoring

Three graduate assistants were employed and trained in the summer before the intervention to teach the GEN 105 course sections offered in the fall semester. Each instructor was required to score sample student written responses in all three disciplines (biology, psychology, and history) from the portfolios collected and retained from students enrolled in GEN 105 in the previous semester. In group training sessions led by professional trainers, the instructors completed scoring rubrics individually and then compared and discussed ratings to establish benchmarks for each of the rubric scale levels for each of the three disciplines. In addition, each GEN 105 instructor was required to create a discipline-specific written response to serve as a model for each of the selected textbook or supplemental content readings used for the fall semester and for the reading excerpts or supplemental reading articles used for precourse and postcourse reading assessment in each of the disciplines to provide an additional means of improving the reliability of these instructors' evaluative ratings. The instructors were directed to use this same training procedure to introduce, explain, and model the reading prompt and rubric scoring during the first few weeks of instruction following precourse assessment administered in the first week of class.

To illustrate this instructional procedure, two examples of students' written responses about an assigned biology textbook selection are presented with accompanying scoring rubrics in Figures 4 and 5 (figures are found on the preceding pages). Student A illustrates a written response assigned low ratings (Figure 4). Student B illustrates a written response assigned high ratings (Figure 5).

Method

Design of the Study

The Paul and Elder (2009) critical thinking framework provides a theoretical connection between critical thinking and reading skill, but empirical evidence of the effectiveness of using this framework for instruction and assessment is needed. The purpose of this study was to investigate the use of a critical thinking rubric created using the Paul and Elder (2009) framework as an assessment of reading achievement for 164 students who had complete assessment data and were enrolled in a reading intervention course. The hypothesis for this study was that the use of this framework in the GEN 105 course would improve the reading achievement of students enrolled as measured by two different methods of assessment. This study asked three research questions: (1) Did students' scores indicate an increase in reading achievement as measured by the COMPASS Reading Test and the critical thinking rubric, an alternative measure of reading achievement? (2) What was the relationship between students' COMPASS Reading Test scores and critical thinking rubric scores? (3) What patterns and trends in reading achievement were suggested by the seven critical thinking rubric scores earned at spaced intervals during the semester? The study was approved by the university's Institutional Review Board as a retrospective review of existing data.

Participant Characteristics

The subjects in this study were 71 males (43.3%) and 93 females (56.7%) with a mean ACT of 17.5 (SD = 1.5) enrolled in seven sections of GEN 105 and paired content courses in biology, history, and psychology. Table 1 summarizes the descriptive statistics for the subjects' COMPASS and critical thinking rubric scores.

Table 1
Dependent Variable Means, Standard Deviations, and Ranges

Variable	Mean	SD	Range
preCOMPASS	70.34	10.67	76
postCOMPASS	76.30	11.46	60
preRubric	16.69	2.71	12
postRubric	19.05	2.92	13

Results

Reading Achievement Scores

To answer the first research question and to determine the reading achievement indicated by students' mean score differences, paired-samples *t*-tests were computed. There was a statistically significant difference in the mean scores for the preCOMPASS Test and the postCOMPASS Test, $t(141) = 5.587$, $p < .001$. The postCOMPASS Test score mean (76.30) was significantly higher than the preCOMPASS Test score mean (70.34). Results also indicated a statistically significant difference in the mean scores for the precritical thinking rubric and the postcritical thinking rubric, $t(151) = 10.51$, $p < .001$. The critical thinking postrubric score mean (19.05) was significantly higher than the critical thinking prerubric score mean (16.69).

Relationship between COMPASS Reading Test and Critical Thinking Rubric Scores

To answer the second research question and to determine the strength of the relationship between the two reading achievement measures (COMPASS Reading Test scores and critical thinking rubric scores), a Pearson correlation was computed. No statistically significant relationship was found between the two measures ($p > .05$). This finding suggests that these two measures may be assessing reading achievement differently.

Patterns and Trends in Reading Achievement over Time

To answer the third research question regarding patterns and trends indicated by the seven critical thinking rubric scores that were earned at intervals during the semester, several repeated-measure ANOVAs (RM-ANOVA) were computed. One of the core underlying assumptions of the univariate RM-ANOVA is sphericity, meaning that the variances of the differences between data taken from the same participant are equal. Mauchly's Test of Sphericity was computed and found to be highly significant, $W = .76$, $\chi^2(20) = 30.451$, $p = .063$, suggesting that the observed matrix does have approximately equal variances and equal covariances, so no corrections were applied to the F-ratio computations. Table 2 summarizes the results of the RM-ANOVA analysis. A significant change in the students' critical thinking rubric scores did occur across time, $F(6, 678) = 30.86$, $p < .05$.

Table 2
Repeated Measures Analysis of Variance for Critical Thinking Rubric Scores

Sources	*SS*	*df*	*MS*	*F*	*p*
RM (Rubric over Time)	936.208	6	156.035	30.86	.000
Subjects	3901.826	113	34.529		
Subjects x Rubric Over Time	3428.078	678	5.056		
Total	8266.112	791			

In order to investigate these relationships further, follow-up contrasts compared all time points against the initial score (baseline). Post-hoc comparisons were performed using the Bonferroni adjustment for multiple comparisons. As indicated in Table 2, the impact of the repeated measures, the use of the rubric over time, is significant.

Table 3
Bonferroni Comparison for Critical Thinking Rubric Scores

				95% CI	
Comparisons	Mean Score Difference	Std. Error	Sig.	Lower Bound	Upper Bound
Pre vs. Rubric 1	.228	.303	1.000	1.169	.713
Pre vs. Rubric 2	1.202	.323	.006	2.205	.198
Pre vs. Rubric 3	2.632	.322	.000	3.632	1.631
Pre vs. Rubric 4	2.368	.302	.000	3.308	1.429
Pre vs. Rubric 5	2.772	.304	.000	3.718	1.826
Pre vs. Post	2.377	.255	.000	3.169	1.585

The Bonferroni comparisons can be seen in Table 3; in this section, the mean scores that underpin the data in that table are discussed. The critical thinking rubric scores did not increase significantly from a mean of 16.78 (SD = 2.77) for the precourse rubric score to a mean of 17.00 (SD = 2.9) for rubric 1 (R1). However, significant improvement was found at the following points across time: rubric 2 (R2) (M = 17.98, SD = 3.26); rubric 3 (R3) (M = 19.41, SD = 3.39); rubric 4 (R4) (M = 19.14, SD = 3.17); rubric 5 (R5) (M = 19.55, SD = 2.96); postcourse rubric score (M = 19.15, SD = 2.80). Figure 6 shows the marginal means for the critical thinking rubric scores at each point of measurement.

To examine the critical thinking rubric scores across time for each of the different paired courses in biology, psychology, and history, a RM-ANOVA was again conducted. Figure 7 shows the marginal means for each discipline at each point of observation. Students did show statistically significant improvements as measured by the means of the precourse, postcourse, and rubric scores 1–5 for each discipline; however, for each discipline, improvements occurred among different readings and at different data points in the semester.

Biology trend. Mauchly's test indicated that the assumption of sphericity had been violated (W = .15, df = 20, $p < .05$) for the biology GEN 105 sections; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .565). The results indicated a significant effect on the rubric scores across time in the biology GEN 105 course sections, $F(3.39, 94.95) = 4.92$, $p < .05$. The critical thinking rubric scores did not increase significantly from a mean of 14.41 (SD = 2.69) for the precourse rubric score to a mean of 15.51 (SD = 2.5, p = 1.000) for rubric 1 (R1). In addition, no statistically significant improvement was found across time for rubric 2 (M = 16.14, SD = 2.34, p = .085) nor for rubric 3 (M = 16.72, SD = 2.34, p = .482). However, significant improvement was found across time for rubric 4 (M = 16.89, SD = 2.52, $p < .05$), rubric 5 (M = 17.17, SD = 2.59, $p < .05$), and the postcourse rubric scores (M = 17.00, SD = 2.26, $p < .05$).

History trend. Mauchly's test indicated that the assumption of sphericity had not been violated (W = .506, df = 20, p = .116) for the history GEN 105 course sections; therefore, no corrections were applied to the F-ratio computations. The results indicated a significant effect on the critical thinking rubric scores across time in the history GEN 105 course sections, $F(6, 258) = 18.02$, $p < .05$. The critical thinking rubric scores did not increase significantly from a mean of 16.61 (SD = 2.27) for the precourse rubric score to a mean of 17.52 (SD = 3.25, p = 1.000) for rubric 1 (R1). However, significant improvement was found across time for rubric 2 (M = 18.61, SD = 3.84, $p < .05$), rubric 3 (M = 19.61, SD = 2.44, $p < .05$), rubric 4 (M = 19.25, SD = 3.32, $p < .05$), rubric 5 (M = 20.84, SD = 2.50, $p < .05$), and the postcourse rubric scores (M = 19.25, SD = 2.87, $p < .05$).

Figure 6

Estimated Marginal Means of Rubric Scores Over Time for All GEN 105 Courses

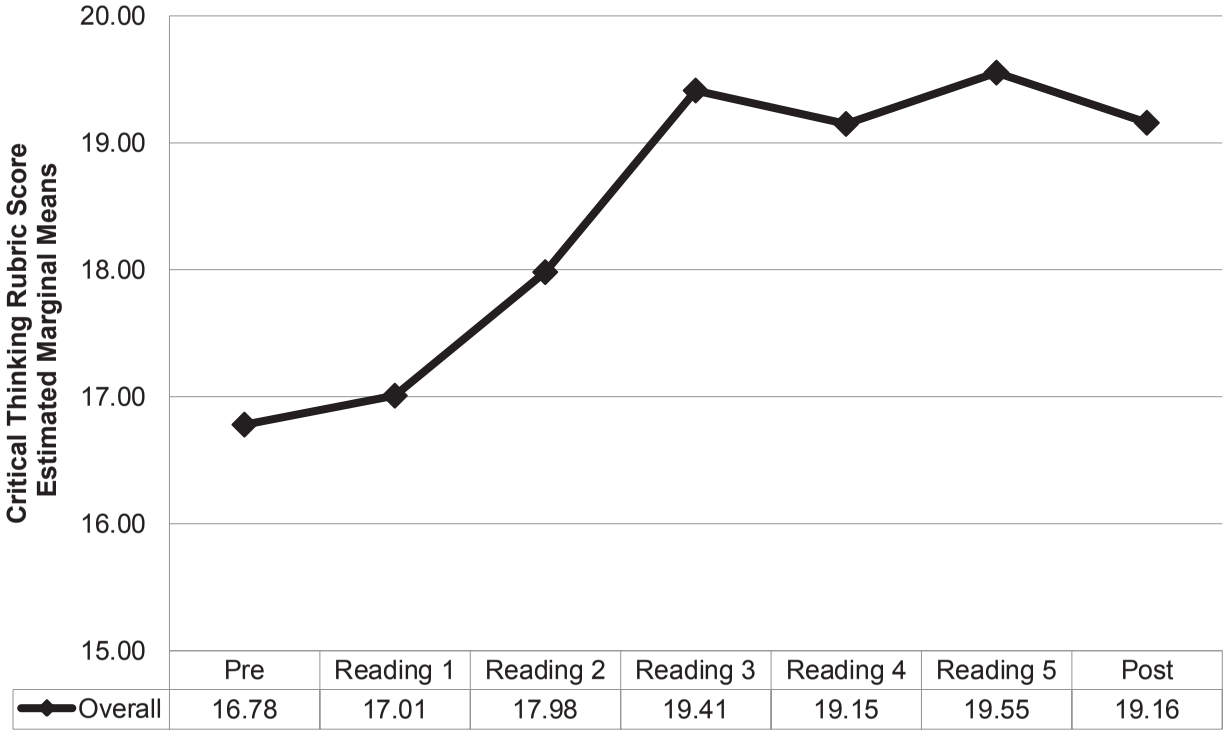
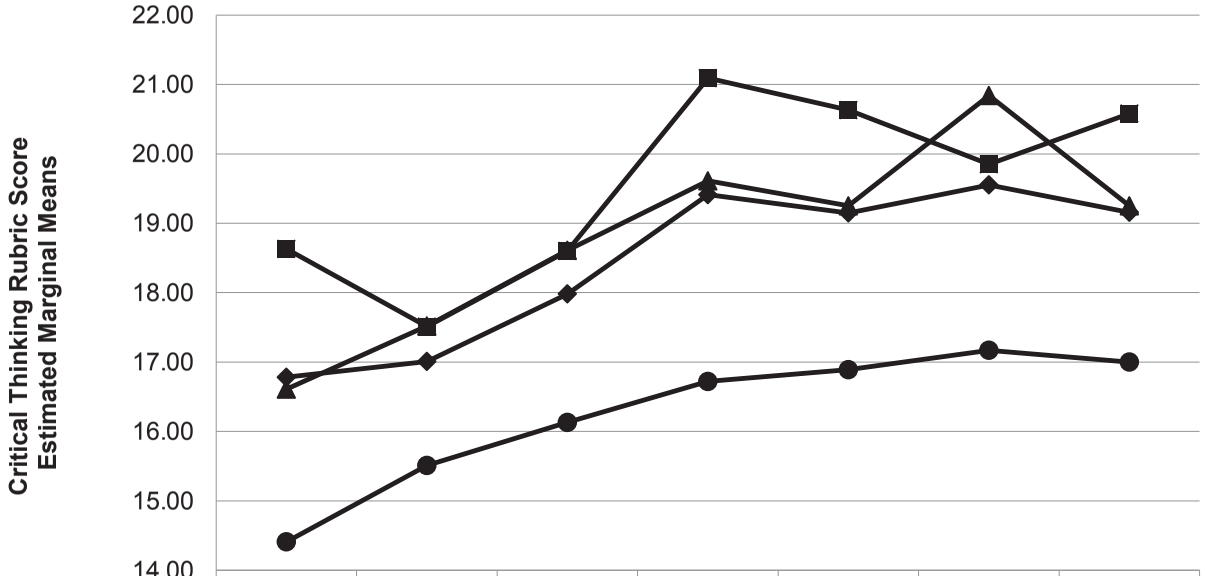


Figure 7

Estimated Marginal Means of Rubric Scores Over Time for Each GEN 105 Discipline



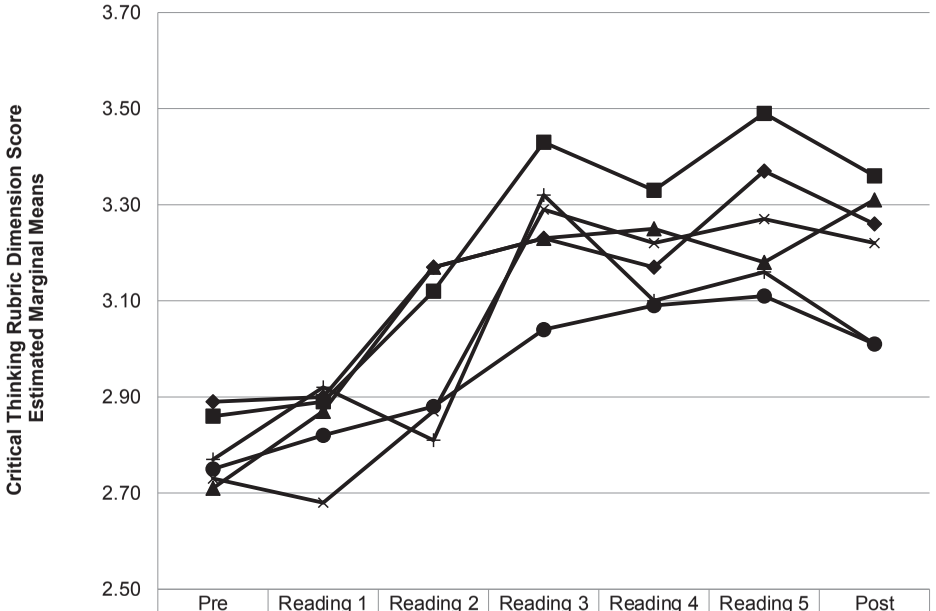
	Pre	Reading 1	Reading 2	Reading 3	Reading 4	Reading 5	Post
Overall	16.78	17.01	17.98	19.41	19.15	19.55	19.16
Biology	14.41	15.51	16.13	16.72	16.89	17.17	17.00
History	16.61	17.52	18.61	19.61	19.25	20.84	19.25
Psychology	18.63	17.51	18.60	21.09	20.63	19.85	20.58

Psychology trend. Mauchly's test indicated that the assumption of sphericity had not been violated (W = .505, df = 20, p = .173) for the psychology GEN 105 course sections; therefore, no corrections were applied to the F-ratio computations. The results indicated a significant effect on the critical thinking rubric scores across time in the psychology GEN 105 course sections, $F(6, 240) = 16.88$, $p < .05$. The critical thinking rubric scores did not increase significantly from a mean of 18.63 (SD = 1.86) for the precourse rubric score to a mean of 17.51 (SD = 2.41, p = .369) for rubric 1 (R1), nor was statistically significant improvement found across time for rubric 2 (M = 18.60, SD = 2.63, p = 1.000). However, significant improvement was found across time for rubric 3 (M = 21.09, SD = 1.97, $p < .05$) and rubric 4 (M = 20.63, SD = 2.46, $p < .05$). No statistical improvement of scores was found for rubric 5 (M = 19.85, SD = 2.70, p = .274), although the precourse to postcourse rubric scores indicated significant differences, (M = 20.58, SD = 2.03, $p < .05$).

Examining the critical thinking standards over time. To examine the relationships of the individual dimensions comprising each critical thinking rubric score further, an analysis was conducted for the following: accuracy, clarity, precision, depth, relevance, and logic across the semester time span. Figure 8 shows the marginal means for each dimension at each time of assessment.

No statistically significant difference was found for any of the mean scores for each dimension of the total precourse rubric score and the rubric 1 score (p = 1.00). However, statistically significant differences were found for some mean scores for different dimensions and for different readings across the remaining time of the semester. Students' scores on accuracy, clarity, and precision showed significant mean differences when comparing the precourse rubric score with scores earned on rubrics 2, 3, 4, 5, and the postcourse rubric ($p < .05$), and students' scores on clarity showed significant mean differences when comparing the precourse rubric score with scores earned on rubrics 3, 4, 5, and the postcourse rubric ($p < .05$). Students' scores on precision showed significant mean differences when comparing the precourse rubric score with scores earned on rubrics 2, 3, 4, 5, and the postcourse rubric ($p < .05$). Students' scores on depth showed significant mean differences when comparing the precourse rubric score with scores earned on rubrics 3, 4, 5, and the postcourse rubric ($p < .05$). The dimensions of relevance and depth showed the least improvement over time. Students' scores on relevance showed significant mean differences only when comparing the precourse rubric to rubrics 3, 4, 5 ($p < .05$), and students' scores on logic showed a significant mean difference only between the precourse rubric score and rubric 4 ($p < .05$).

Figure 8.
Marginal Means for Rubric Dimensions Over Time



	Pre	Reading 1	Reading 2	Reading 3	Reading 4	Reading 5	Post
Accuracy	2.89	2.90	3.17	3.23	3.17	3.37	3.26
Clarity	2.86	2.89	3.12	3.43	3.33	3.49	3.36
Precision	2.71	2.87	3.17	3.23	3.25	3.18	3.31
Depth	2.73	2.68	2.87	3.29	3.22	3.27	3.22
Relevance	2.77	2.92	2.81	3.32	3.10	3.16	3.01
Logic	2.75	2.82	2.88	3.04	3.09	3.11	3.01

Discussion

The significant difference found between the students' mean placement (precourse) and end of semester (postcourse) COMPASS Reading Test scores and the significant difference found between the mean scores for precourse and postcourse rubric scores are encouraging for these designers of college reading instruction. These findings support that students enrolled in GEN 105 improved performance on two different measures of reading achievement following the instruction received in the GEN 105 course sections across all three disciplines. The lack of a statistically significant relationship between the students' rubric scores and COMPASS Reading Test scores may indicate that the standardized reading test and the written summary are measuring different subskills related to reading achievement and cognitive reasoning. The products measured (correct multiple-choice responses vs. a written summary) are very different methods of assessment.

For all three disciplines, students did show statistically significant improvements in rubric scores. However, rubric scores did not increase significantly from the precourse score to the completion of rubric 1. This finding may be the result of the GEN 105 curriculum as structured during the fall semester. Most instruction in the Paul and Elder (2009) framework and in-class explanation of the rubric's dimensions occurred between rubric 2 and rubric 3. The late timing of this instruction may have resulted in less improvement in the critical thinking rubric scores very early in the semester. In addition, differences in the assigned course readings among the three disciplines may also have had a significant effect on this study's results. A review of the number of pages assigned in the textbook and the number of outside readings assigned found the history course to be the most reading-intensive of these three courses, psychology to be the second most reading-intensive course, and biology to be the least reading-intensive course. Differences in the quantity of reading assigned do correspond to this study's results, which indicate statistically significant improvement in critical thinking rubric scores earned following rubrics 2, 3, 4, and 5 in history. Psychology showed similar results only after rubrics 3 and 4, while biology showed similar results only after rubrics 4 and 5. Lastly, biology and history results show (in Figure 7) a consistent positive slope for improving rubric scores across time to rubric 3, but this same result was not found for psychology. Scheduling the due date for rubric 3 in GEN 105 on the same day as the first psychology exam may have contributed to this result by decreasing student attention to the critical reading task.

Additional training has already been implemented for the graduate student assistants selected to instruct course sections of GEN 105 for

future semesters. This training includes generating an inter- and intra-reader reliability score during a longer and more extensive training process in the Paul and Elder (2009) critical thinking framework to strengthen the consistency of rubric scoring among these instructors. Group scoring sessions were also implemented following the scoring of rubric 2 across all three disciplines as an additional training experience to improve scoring reliability and validity. Textbook excerpts and articles selected and assigned as the critical readings are scrutinized carefully for readability, length, and relation to topic. Attention was given to providing similar although supplemental reading experiences across the three disciplines.

Limitations of the Study

One limitation of this study is the lack of a control group; it is not possible to know the strength of factors besides reading instruction on reading achievement gains. The instruction received and the college-level reading assignments completed concurrently in the content courses are assumed to be contributing factors to students' improved reading achievement, and maturation may be another factor; therefore the intentional instruction of the Paul and Elder (2009) critical thinking framework to improve reading skills was not the sole contribution to students' reading achievement. In addition, this study is limited because it presents only one semester's data. Additional semesters of data are needed for additional analysis. A larger number of scores creating a larger data set over time may produce different results.

Future Research

Additional research is needed to explore the use of the critical thinking rubric to assess college reading achievement. If students are expected to read complex and diverse content-related text critically for success in college, then assessment of college programs and courses that address college readiness in reading must continue to evolve and to provide data that best demonstrate readiness for entry-level college courses. Accountability for exploring multiple measures of reading achievement that more accurately reflect the challenges of the college classroom may best be assigned to instructors who create curriculum and prepare students for college readiness. College administrators who are accountable for measures of college readiness for reading for state-wide reporting will want to encourage this exploration by instructors of developmental or intervention courses. At this institution, additional procedures are planned to further explore and improve the use of the critical thinking rubric for reading assessment in GEN 105. Future research is planned

to compare data from multiple years to review the possible effects from additional instructor training and from improved focus on selecting readings for the content areas and may provide more evidence for better understanding the learning outcomes of using this method of assessment for college reading.

References

Assessing learning outcomes at the University of Cincinnati: Comparing rubric assessments to standardized tests. (2010, April). *AAC&U News*. Retrieved from http://www.aacu.org/aacu_news/aacunews10/april10/feature.cfm

Cox, S. R., Friesner, D. L., & Khayum, M. (2003). Do reading skills courses help underprepared readers achieve academic success in college? *Journal of College Reading & Learning, 33*(2), 170–196.

Farr, R. (2003). Building useful instructional reading assessments. *The New England Reading Association Journal, 39*(1), 1–8.

Flippo, R. F., Becker, M. J., & Wark, D. M. (2009). Test taking. In R. F. Flippo & D. C. Caverly (Eds.), *Handbook of college reading and study strategy research* (pp. 249–286). Mahwah, NJ: Lawrence Erlbaum Associates.

Flippo, R. F., & Schumm, J. S. (2009). Reading tests. In R. F. Flippo & D. C. Caverly (Eds.), *Handbook of college reading and study strategy research* (2nd ed., pp. 408–464). New York, NY: Routledge.

Kentucky Developmental Education Task Force. (2007). *Securing Kentucky's future: A plan for improving college readiness and success.* Frankfort, KY: Commonwealth of Kentucky.

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common Core State Standards Initiative.* Retrieved from http://www.corestandards.org/

Nist, S. L., & Simpson, M. L. (2000). College studying. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 645–666). Mahwah, NJ: Erlbaum.

Paul, R., & Elder, L. (2008). *How to read a paragraph: The art of close reading* (2nd ed.). Dillon Beach, CA: Foundation for Critical Thinking.

Paul, R., & Elder, L. (2009). *The miniature guide to critical thinking: Concepts and tools* (6th ed.). Dillon Beach, CA: Foundation for Critical Thinking.

Simpson, M. L., Stahl, N. A., & Francis, M. A. (2004). Reading and learning strategies: Recommendations for the 21st century. *Journal of Developmental Education, 28*(2), 2–32.

Stevens, D., & Levi, A. (2005). *Introduction to rubrics*. Sterling, VA: Stylus.

Cathy W. Leist, *MA & MA, Reading Specialist, is the Executive Director for Resources for Academic Achievement (REACH) at the University of Louisville. She directs the university's centralized academic support unit for all undergraduate students, offering comprehensive services, peer mentoring, and retention programs that serve about 8,000 students each academic year.*

Mark A. Woolwine, *MEd, is a Coordinator for Resources for Academic Achievement (REACH) at the University of Louisville. He is currently coordinating the GEN 105 Supplemented College Reading Program and the Student Success Seminars for REACH and earning a second master's degree focused on the educational applications of new technology.*

Cathy L. Bays, *PhD, RN, is the Ideas to Action (i2a) Specialist for Assessment at the University of Louisville. She served as a faculty member in the School of Nursing for 15 years and as director of the undergraduate program for five years. She is currently coordinating unit and university assessments for the university's regional reaccreditation initiative, Ideas to Action.*