Assessments of Evidence-Based Argument in Three Disciplines: History, Science and Literature

Project READI Technical Report #10

Susan R. Goldman, M. Anne Britt,
Carol D. Lee, Patricia Wallace
and Project READI

PROJECT READi

inquirium

Northern Illinois University

NORTHWESTERN UNIVERSITY

UIC UNIVERSITY OF ILLINOIS AT CHICAGO

WestEd

Citation for this Report: Goldman, S. R., Britt, M. A., Lee, C. D., Wallace, P. & Project READI. (2016). *Assessments of evidence-based argument in three disciplines: History, science and literature.* READI Technical Report #10. Retrieved from URL: projectreadi.org

Please send us comments, questions, etc.: info.projectreadi@gmail.com

Project READI was supported by the *Reading for Understanding (RFU)* initiative of the Institute for Education Sciences, U. S. Department of Education through Grant R305F100007 to the University of Illinois at Chicago from July 1, 2010 – June 30, 2016. The opinions expressed are those of the authors and do not represent views of the Institute or the U. S. Department of Education.

Project READI operated as a multi-institution collaboration among the Learning Sciences Research Institute, University of Illinois at Chicago; Northern Illinois University; Northwestern University; WestEd's Strategic Literacy Initiative; and Inquirium, LLC. Project READI developed and researched interventions in collaboration with classroom teachers that were designed to improve reading comprehension through argumentation from multiple sources in literature, history, and the sciences appropriate for adolescent learners. Curriculum module materials were developed based on enacted instruction and are intended as case examples of the READI approach to deep and meaningful disciplinary literacy and learning.

Assessing Evidence-Based Argument from Multiple Sources in History, Science and Literature

Susan R. Goldman, Ann M. Britt, Carol D. Lee, Patricia Wallace,
and Project READI

A primary purpose of the Project READI efforts is to increase adolescents' skills at closely reading multiple texts for purposes of comprehending and constructing arguments that respect disciplinary differences in criteria for sound argumentation and their functions (Goldman, et al., 2016). For example, historical narratives may convey causal accounts of events from the past whereas scientists create causal accounts of natural phenomena. Literary argumentation strives to establish interpretations of the human condition that are revealed through authors' choice of language, plot structure, character types, and so forth (see Lee, Goldman, Levine, & Magliano, 2016). Critical to intervention development was the development of assessments that were aligned with the learning goals and instructional processes. These were intended largely for formative use in the classroom and in the design of the interventions. But in the context of evaluation of the impact and effect of READI interventions, they were also viewed as useful as summatively. Accordingly, basic studies and the design teams devoted some of their efforts to the design, development and testing of assessments of evidence-based argument (EBA) assessments. These were often used in conjunction with intervention implementations as pre and post assessments. As well, the development work encompassed students not engaged in interventions to provide "baseline" information on performance levels and range in performance across as well as within grade bands.

In each discipline the assessments include both open-ended essays and more structured responses that are designed to provide a more detailed profile of students' skills at applying disciplinary literacy and inquiry practices to different content in the discipline. It was decided to include essay tasks because essays provide an authentic disciplinary goal and represent the high level intellectual work we want students to achieve. At the same time, the interventions were focused on reading. Over the time course of the research and development work, it became clear that support for structuring sentences, paragraphs and essays to conform to and convey disciplinary thinking was needed. Such supports were incorporated into the interventions. The more structured responses provided assessments of comprehension under conditions of reduced production demands.

The EBA assessments are designed to assess the READI Learning Goals as a group (See Table 1) and are aligned with our discipline-specific definitions of argumentation (Goldman et al., 2016). In developing these assessments, we endeavored to make the EBA assessments for the three disciplines as parallel or comparable as possible while respecting important differences among the disciplines.

Table 1. Learning goals

Label	Description
1. Close reading	Engage in close reading of texts as appropriate to the disciplinary task and text.
2. Synthesize	Synthesize within and across aspects of texts important to the disciplinary tasks and texts.
3. Construct	Construct written arguments with claims, evidence and warrants,

explanations	organized logically and expressed clearly and that reflect disciplinary norms for argument.
4. Establish criteria	Establish criteria for judging interpretive claims and arguments that are appropriate to the discipline.
5. Provide and connect support	5. Construct arguments explaining the logic of how the claims are supported by evidence using appropriate disciplinary criteria for claims, evidence and logic.
6. Epistemology	6. Demonstrate understanding of the nature of knowledge and how that knowledge is constructed as appropriate to the discipline.

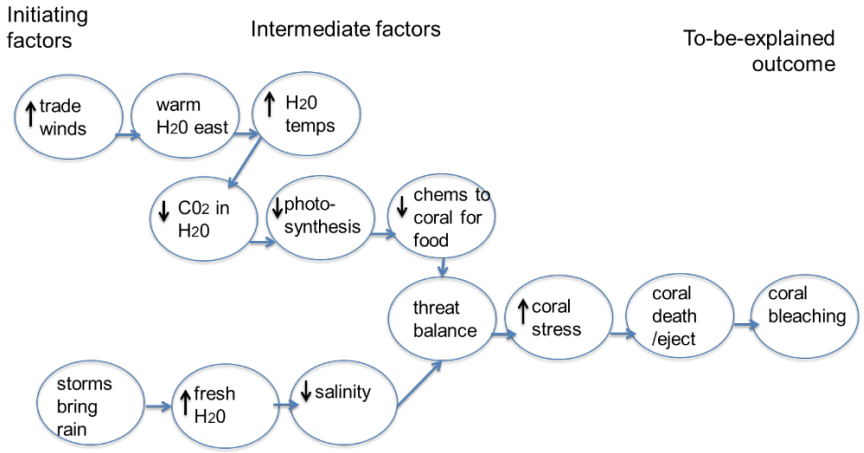
History and Science EBA Assessment Development

M. Anne Britt, Patricia Wallace, Dylan Blaum, Brent Steffens, Michael Manderino, Monica Ko, Susan R. Goldman *

Phase 1. Initial development of the history and science EBA assessments

Because the history and science versions are so similar, they are described together. The history and science assessments were created using the same principles as those we have used in the basic processes studies over course of the grant (Blaum, Britt, Griffin, & Wiley, 2014; Griffin, Wiley, Britt, & Salas, 2012; Jaeger, Griffin, Britt, & Wiley, 2015; Shanahan, et al., 2013; Steffens, Britt, & Manderino, 2014; Wallace, Blaum, Rupp, & Britt, 2014; Wiley, Steffens, Britt, & Griffin, 2014). Students were asked to read a set of documents to answer a how or why question about the causes of a historical event or a scientific phenomenon by integrating across multiple sources. The document sets were built around a causal model (see Figure 1), and each document contributes only part of the overall causal model with some overlap between the sources (Wiley & Voss, 1999). Consistent with our learning objectives, a variety of document types are included in the set (Goldman et al., 2016). The documents are presented one per page with source information at the bottom of each page. The pages are not stapled or numbered so information can be compared across sources. Students are asked to compose their essays and complete all post-essay items with the documents available (Goldman et al., 2016).

Figure 1. Causal model for the coral bleaching document set.



* With acknowledgment to Kathryn Rupp, Karina Perez, Thomas Griffin, Jennifer Wiley, Cynthia Greenleaf and Willard Brown.

Prompt and document sets

History. We created documents sets supporting a controversy similar to those investigated in previous research on reading like a historian (Britt, Rouet, Georgi, & Perfetti, 1994; Sanchez & Wiley, 2006; Wineburg, 1991). The two historical events selected were the Birmingham Children's Crusade and the Scopes trial. These topics are related to but not typically covered in adolescent history curricula. It was also possible to create document sets for these topics that (1) were within acceptable readability ranges for adolescents; (2) supported construction of a causal model, (3) were of appropriate length and number, (4) and included multiple types of historical documents. Each document set began with a short (184 and 229 words) background document to establish the to-be-explained outcome and create an interesting hook (the hook was "why were children protesting and getting arrested?" or "why did a teacher get arrested for teaching evolution?"). This sheet was then followed by 6 documents (M=237 words for Birmingham and M=199 words for Scopes) of a variety of types (e.g., an interview, a graph, textbook excerpt, a letter, article in newspaper). Five features of source information were provided and labeled (e.g., "Author:", "Title:", "Published:", "Date:", and "About Source"). Because source information is critical in historical inquiry, we flagged the source features in this way so that students would be able to correctly parse the source elements.

Science. Coral Bleaching and Skin Cancer were the two topics selected for EBA development based on their appropriateness to assess effects of a READI intervention in either earth sciences or life sciences. Each of these science courses was under consideration as potential sites of the READI efficacy trial at the time the EBA development began. As in the case of history, each of these topics is related to but is not specifically covered in earth science or life science curricula for adolescents. As well, neither was the intended topic of the READI intervention. Also as in history, the prompts for each topic were the similar: "explain how and why coral bleaching rates vary at different times," and "explain how and why rates of skin cancer differ around the globe". The science document sets begin with a short (255 (CB) and 273 (SC) word) background document to establish the to-be-explained outcome, followed by 4 documents (M=288 (CB) words and M=250 (SC) words) of a variety of types (e.g., descriptive texts, images, graphs and maps).

The EBAs were developed for high school grade bands. Middle school versions modified the readability by simplifying vocabulary and shortening sentences. The causal models and the number of documents for the two grade bands remained constant.

Tasks and procedure

Table 2 shows the assessment tasks and the learning goals targeted to be assessed by each type of task/item type for science; for history there was a parallel chart. To assess prior topic knowledge, students completed a 3-5 minute prior knowledge assessment in which they estimated their knowledge of related concepts and listed the causes for the event or phenomena that they were aware of prior to reading. This was intended as a gross indicator of general familiarity with each topic. The materials provided as part of the assessment contained all the information needed to address the prompt.

Table 2. Assessment tasks/item types and the learning goals targeted for the science assessment.

	Learning objective	
Measure	First iteration	Final assessment
Annotations on texts	Close reading*	Close reading*
Main essay	Synthesize*, Construct explanations*, Establish criteria*	Synthesize*, Construct explanations*, Establish criteria*
Alternative-cause essay	Synthesize*, Construct explanations*	N/A
Graphical model comparison	Synthesize, Construct explanations	Synthesize, Construct explanations
Peer essay evaluation	Synthesize, Construct explanations, Establish criteria*	Synthesize, Construct explanations, Establish criteria*
Multiple choice	Synthesize, Construct explanations	Synthesize, Construct explanations

*Note. An asterisk indicates it requires production rather than recognition.

After completing the prior knowledge items, students were given a task sheet explaining their close reading and essay task. They were told that this task may be challenging and they should try their best. Based on findings from the basic studies (Wiley et al., 2012), students were provided with explicit directions to "use specific information from the sources to support your conclusions and ideas" and that they would "have to piece together important information across the sources to construct a good understanding. No one source will provide the answer. You are the one making connections across sources and coming up with an explanation." These explicit task instructions are important to make sure all students understand what the task entails (Jaeger et al., 2015; Steffens et al., 2014; Wallace et al., 2014; Wiley et al., 2014). Only minor wording changes occurred between the task sheet for history and science except for tense (e.g., happened/ happen and historic events/scientific phenomena). Students were then asked to read and annotate the texts ("While reading, it is important to show your thinking by making notes in the margins or on the texts"). They were given approximately 40 to 45 minutes on the first day and 15 minutes at the beginning of the second day to read and write. Students were told that after writing their essay they would be "asked questions and need to use specific information from the sources to support your conclusions and ideas". The students were told they had the document set available when doing the tasks. There were given approximately 35 minutes for these post-reading tasks.

The post-essay tasks were created to be as similar as possible for each topic and discipline. Two of these tasks were identical for science and history. The first post-essay task, *alternative-cause essay*, asks students to write an explanation for the less salient alternative initiating factor connectable to the to-be-explained outcome. In Figure 1, this corresponds to the bottom chain connecting storms to coral bleaching. This alternative-causes question was based on findings from the basic studies showing students often mention only one initiating factor (Hughes et al., 2015; Wiley et al., 2012). The other identical task was *peer essay evaluation* items modified from those used in our prior basic studies (Wiley et al., 2012; Wiley et al., 2014; Wiley et al.,

2009). Students were given a "student" essay and asked "What did the student do well in the essay?" and "What advice would you give the student for improving this essay?"

The other post-essay questions were slightly different for science and history so they will be described separately. For the history assessment, a short answer question asked students to identify an important disagreement among participants that could lead to *alternative explanations* of the to-be-explained event. The next question asked students to *classify the perspective* of authors and involved participants. Finally, three *very short answer questions* that targeted single document comprehension of information important the causal model (e.g., understand major motivation and actions that underlie the causal model, and interpreting the graph) were included. These final questions were designed to assess basic comprehension of the text (textbase questions) as "covariates" of multiple source comprehension needed to answer the essay and other task questions.

For the science assessment, students were given a *graphical model comparison* item and *multiple choice* items. Because of the importance of graphical representations in science (Lemke, 1998), a graphical model comparison was included in which 2 graphical models of potential explanations were shown and students had to select the best explanation and explain why it was best. We selected the most common problem with scientific explanations that we identified from the essays in the basic studies (Wiley et al., 2012), that of directly connecting each cause to the outcome rather than creating a coherent causal chain. For example, in Figure 1, the events of increased trade winds, increased water temperatures, and coral ejection were either presented as a chain to coral bleaching or as 3 distinct chains (increased trade winds → coral bleaching, increased water temperatures→ coral bleaching, coral ejection→ coral bleaching). We also created two multiple choice items: one that required an understanding of part of the causal explanation and another that required an interpretation of the data in a graph.

Pilot Testing of the EBA assessments in history and science.

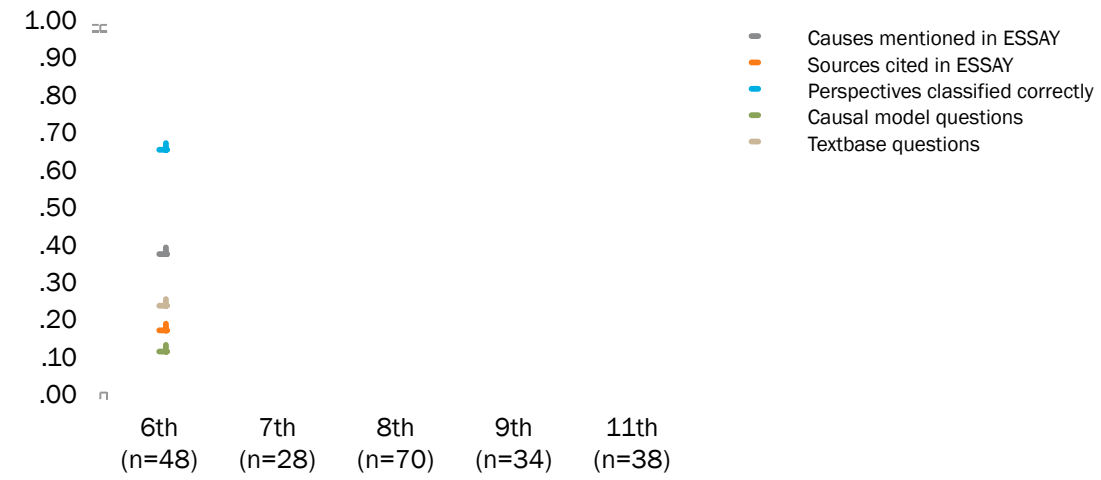
In the Fall, 2013, pilot testing of the history and science EBA assessments were conducted. We recruited science and history teachers participating in the Chicago Teacher Network to conduct these studies. For the EBA science, participants were drawn from 5 sixth, 5 seventh, 3 ninth, and 2 eleventh grade classes. For the history EBA, participants were drawn from 4 sixth, 2 seventh, 4 eighth, 2 ninth and 2 eleventh grade classes. For each assessment, approximately 400 students consented and participated. Participants were randomly assigned to topics within each classroom. Essays were coded for mentions of key causes (max was 13).

Detecting change over time. We were also interested in whether the EBA assessment was sensitive to change over time so re-administered the history assessment several months later in a subset of the classrooms of the teachers in the network. However, since the instruction in these classrooms was only minimally including elements of the READI approach, we did not expect to see much difference in performance. For the pretest, 146 of the students were in middle school, and 72 students were in high school. After half of the students in each grade and had Birmingham at pre and Scopes at post and vice versa. Students completed the prior knowledge inventory by marking their level of knowledge on a 5-point scale across several topics, and answered an open-ended question about the causes of the historical event. As in the prior sample, on the pretest high school students (M =2.69, SD =.58) reported having more prior knowledge than middle school students (M =2.33, SD =.72). Overall, students rated their knowledge of

Birmingham to be higher (*M* =2.68, *SD* =.72) than Scopes (*M* =2.22, *SD* =.59), but only 7 students were able to list a single cause of the events and these were all for Birmingham.

Figure 2 shows the average proportion correct for each dependent measure by grade on the pretest. Essays were coded for the number of causes mentioned for the event. For each measure, there was a significant linear trend by grade. In only two cases, *causes mentioned in essays* and *textbase questions*, were there differences in document set, with students performing better with the Birmingham set. This is not unexpected given the prior knowledge scores. For no measures was the interaction of grade and document set significant, so we collapsed the document set for presentation here.

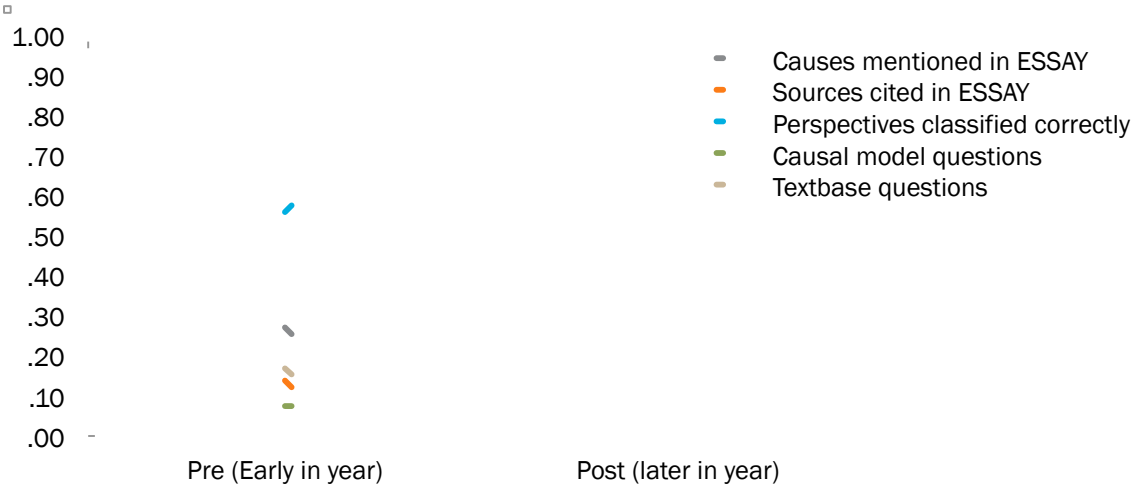
Figure 2.



On average, students mentioned 3.2 of the 16 Scopes causes and 4.1 of the 13 Birmingham causes. Thus, in general, students identified multiple causes, but there were many other causes that could have been integrated into their answer. Although most of the students (66%) mentioned at least one source (external or embedded), the actual proportion of sources cited was rather low. The post essay measure attempted to understand the essay performance better. Performance on the task of classifying the *perspective of 4 authors* and involved participants as pro, con or neutral found that even the 6th grade students were better than chance (.33). Thus, they seemed to be sensitive to perspective or bias. The two open ended *causal model questions* (14 for Scopes and 18 for Birmingham) tapped single document comprehension of important information to support understanding of the causal model to examine whether students knew more about the event than included in the essay. The performance overall was very low, and this may be because students often will not include information that was already included as an answer for a prior question (Wiley, Steffens, Britt, & Griffin, 2014). Performance on the three *textbase questions* (max of 14) was higher than on the causal model question. These post essay measures suggest that the essay is a good indication of what they learned. The results also show a general pattern of increased learning by grade. However, the point of increase may not be the same for all measures. For example, for the causal model questions, high school seems to be the point where the increase occurs, whereas for the classify-the-perspectives task, an increase is already being seen in 7th grade.

Although we intended to re-administer the assessment to all of the students whose data are shown in Figure 2, we were only able to obtain data for a subset of the students (41 6th, 67 8th, and 35 11th). There was an approximately equal number of students within each grade receiving each order (Birmingham at pre and Scopes at post vs. Scopes at pre and Birmingham at post), order was collapsed in Figure 3. As expected, the results suggest that integrating information from multiple texts to explain a historical event is a skill that will not improve without a serious intervention.

Figure 3.



After initial development, it was determined that the efficacy study would be in the discipline of science, specifically biology at the 9th grade level. Therefore, we suspended development of these history assessments.

Results: Science

We had scorable data from 45 sixth, 34 seventh, and 26 high school students (grades 9 – 11). The essays were scored against the causal model for structure and concepts mentioned. The essay question ("explain how and why coral bleaching rates vary at different times") affords a discussion of both normal processes (when the rates are low to none) and change in normal processes (what caused the rates to increase). The texts provide information about 12 normal elements (e.g., trade winds travel west) and 13 change elements (e.g., a decrease or reverse in direction of trade winds) that can explain changes in coral bleaching by creating two causal chains (one beginning with a decrease in trade winds and the other beginning with storms) resulting in two ideal causal models to use for essay and short answer comparisons. For instance, a student could state that a decrease in trade winds leads to increased water temperatures which then leads to coral bleaching. Thus, there were 25 elements that could be linked with 23 unique connections.

As shown in Table 3, students did construct reasonably long essays. While about a third of it was a near verbatim copy of information from the text set (as determined by an LSA cosine of about .75; see Britt, Wiemer-Hasting, Larson, & Perfetti, 2004), students were able to identify several elements important to the prompt. In Table 3, the "included in essay" columns show the average amount of information from the causal model mentioned anywhere in the essay

regardless of how it was mentioned. On average, students mentioned 6.74 elements (27% of the 25 possible, range from 0 to 7), but they mentioned far fewer connections between these elements, 1.74 links (8% of the 23 possible, range from 0 to 6) suggesting that they are able to identify several of the important elements but have difficulty connecting them to one another in a coherent way.

Table 3. Descriptive statistics for the information in students' coral bleaching essays. Note, Standard deviations in parentheses.

			Included in essay		Explicitly connected to to-be-explained outcome	
Grade	Average number of sentences	Proportion of sentence copied or "nearly" copied	Average number of links mentioned	Average number of elements mentioned	Average number of links in explanation	Average number of elements in explanation
6th	12.33 (5.49)	.24 (.19)	1.44 (2.08)	6.87 (4.43)	.78 (1.08)	1.22 (1.52)
7th	6.21 (3.32)	.24 (.23)	1.29 (1.70)	4.82 (3.01)	.91 (1.14)	1.47 (1.71)
9th, 11th	10.88 (5.45)	.42 (.29)	2.85 (2.66)	9.04 (5.54)	1.50 (1.68)	2.23 (2.20)
Overall	9.99 (5.54)	.29 (.24)	1.74 (2.21)	6.74 (4.59)	1.00 (1.29)	1.55 (1.80)

Although students mentioned important and relevant information from the texts, they did not explicitly transform this content from the way it was mentioned in the texts so that it directly answered the question by explaining how the elements led to changes in rates of coral bleaching. In Table 3, the "explicitly connected to to-be-explained outcome" columns show the average number of links and elements in the students' explanations that specifically connected back to the to-be-explained outcome of coral bleaching (i.e., length of a causal chain). In fact, most of the relevant information in the essays was never explicitly transformed in a way to connect it to the to-be-explained outcome.

Table 4 provides a more detailed account of the nature of the causal chains produced in the essays. As shown in Table 4, overall, half of the students never answered the question. That is, they never explicitly connected the elements to the to-be-explained outcome (i.e., production of a causal chain), and on average, the essays included less than one causal chain (M = .92). When they did directly say that an element affected coral bleaching rates, the average length of the explanation was only 1.89 elements long with the average longest chain being 1.32 links long.

Table 4. Causal chains in the coral bleaching essays.

Grade	Percent of no chains	Average number of chains	Average longest chain	Average number of connections	Average length of chains	Number making two distinct paths

6th	56%	0.64	0.56	0.78	1.22	1
7th	50%	0.65	0.79	1.09	1.59	3
9th-11th	38%	0.88	1.27	1.62	1.89	0
Overall	50%	0.92	1.32	1.68	1.89	4

Two characteristics of high quality explanations in science are coherence and completeness. In order for students to produce complete explanations for the prompt given, they needed to integrate information across the texts. Each text contributed portions of the complete causal chain with slight overlap between two texts to allow for representation of the fully connected causal model when all texts were considered. Causal chains from the essays were compared with each text's contribution to the causal model to determine the degree of text use and text integration in creation of the essay. As shown in Table 5, when the students constructed an explanation using information in the text set (although not all of them did so), it was mostly from a single text. This means that their claims were often one or two elements connected with the to-be-explained-outcome. Although some students created multiple causal chains, most of these were from separate document and should not have been treated as distinct chains because they could have been coherently combined into a single, longer, and more accurate causal chain. In the cases of either a single text or multiple texts separately mentions, the student's causal chain was simply connecting a document to the prompt without integrating across documents. Only the students that created causal chains from multiple texts actually attempted a coherent and integrated explanation. In terms of completeness, as shown in Table 3, only 4 students explicitly connected both initiating factors (storms and trade winds) to changes in coral bleaching in their explanations. However, as shown in the analysis of the short answer responses, more students were able to locate this information when explicitly asked to. In general, the analyses of the essays suggest that this assessment has the opportunity to detect improvement in students' levels of coherence and completeness of explanations.

Table 5. Percent of participants constructing causal chains from the texts.

		Causal chains from:			
Grade	No scorable chain	single text	multiple texts (separate)	multiple texts (integrated)	N
6th	56%	36%	9%	0%	45
7th	50%	41%	6%	3%	34
9th-11th	38%	38%	15%	8%	26

Responses to the short answer question were analyzed in a similar way. This question asked the students to "explain how or why changes in salinity affect coral bleaching events." In order to encourage a more complete answer, students were told that their answer might require repeating content that they already provided in their essays. Although the question focused on salinity, the goal was to get students to focus on the less prominent cause of coral bleaching (i.e., storms). Upon initial analysis of coded responses, it was discovered that students interpreted the question is two very different ways. Table 6 presents the variation in response types that were observed. An ideal answer would require students to consider decreases in salinity and the factors leading to this before explaining how this decrease in salinity leads to coral bleaching

(i.e., salinity as both a cause and effect). The data suggest that very few students (4.5%) took this approach when answering the question, and all who did so were in high school. The next best response would have involved presenting the elements in the causal chain connecting salinity to changes in coral bleaching (i.e., salinity as a cause). Again, very few students took this approach (3.6%) and most were in high school. Instead, most students either answered the question in such a way that they simply described what leads to changes in salinity (i.e., salinity as an effect; 29.5%) or they failed to produce a causal chain at all (62.5%). It is interesting to note that the elements needed for considering salinity as an effect rather than a cause can be captured by paraphrasing or copying a single document (an average of 0.6 of 2 possible elements were included across all essays). Considering salinity as a cause by connecting it to changes in coral bleaching requires integration among multiple documents (on average, only 0.1 of 3 possible elements were included across all essays). There appears to be no grade related change in this approach, whereas the likelihood of producing no causal chain at all diminished with the more advanced grade levels. It is possible that the nature of this question did not lend itself well to producing the secondary causal change (storms to coral bleaching); therefore, a modified short answer prompt that focuses on storms is being considered for future use.

Table 6. Percentage of causal chain production type in response to the short answer question across the different grades.

Grade	Salinity as a cause	Salinity as an effect	Salinity as both a cause and effect	No causal chain produced
6th	4.4%	20%	0%	75.6%
7th	0%	38.5%	0%	61.5%
9th-11th	7.7%	46.2%	19.2%	26.9%
Overall	3.6%	29.5%	4.5%	62.5%

In summary, across the essay and short answer responses, most students found a subset of important elements and connected them directly to the outcome without integrating them into a coherent explanation. A second strategy was to copy key sentences but again not connect information to coherently answer the prompt. Finally, very few students connected up the less salient alternative explanation (storms). The omission of intervening elements and the focus on the main causal chain was very consistent with the findings from the basic studies.

The graphical model comparison task was scored for accurate selection (chance was 50%) and for the accuracy of the justification of that selection. As shown in Table 7, students found the graphical model comparison task difficult. Less than a third of the students were able to select the appropriate model and correctly state why that was the correct selection (27% for coral bleaching and 29% for skin cancer). The multiple choice items suggest that the problem we have typically found with constructing explanation-based essays from the basic studies is not simply a production problem. Across these two topics, only about 30% of the students were able to answer the explanation item correctly. Part of the reason for this lower performance may be that the multiple choice item require the students to fill in two blanks in a single question. We have changed this item in our revised version of the assessment. Students were slightly better at correctly interpreting the graph (42% of students) when the graph was available, but there is room to detect improvement of this skill as well.

Table 7. Percentage of students correctly answering graphical model comparison task and multiple choice items. Note. Standard deviations in parentheses.

		Graphical model comparison task			Multiple choice (chance 25%)		
	Grade	Graph selection	Justification for selection	n	Explanation item	Interpreting graph	N
Coral bleaching	6th	70% (46%)	33% (47%)	43	33% (47%)	47% (50%)	43
	7th	42% (50%)	26% (44%)	31	27% (45%)	40% (50%)	30
	9th, 11th	58% (50%)	17% (38%)	24	17% (38%)	46% (51%)	24
	Overall	58% (50%)	27% (44%)	98	27% (45%)	44% (50%)	97
Skin cancer	6th	43% (50%)	23% (42%)	44	35% (48%)	26% (44%)	46
	7th	39% (50%)	23% (43%)	31	32% (48%)	55% (51%)	31
	9th, 11th	62% (49%)	44% (50%)	34	31% (47%)	49% (51%)	35
	Overall	48% (50%)	29% (46%)	109	33% (47%)	41% (49%)	112

Testing new items and procedural changes

Our group decided that it was important to have the documents available during writing (see reasons above). However, we have noticed a high frequency of direct copying (see Table 3). It is important to examine whether students are actually learning even when they are predominately selecting and copying sentences in their essays. We created three additional causal chain multiple choice items that did not require the students to fill in two blanks in a single question and will provide a more stable measure of learning. In addition to testing out these additional multiple choice questions, we also wanted to examine whether students learned better when they wrote the essay from memory or when they wrote with the documents available. A new sample of 60 students from three 9th grade biology classes were randomly assigned to one of three conditions: documents present (wrote essay with the documents available), documents present with warning (wrote essay with the documents available but were also given a warning that they would be asked questions later without the documents available), memory with warning (wrote essay from memory and were given a warning that they would be asked questions later without the documents available). Students read and wrote their essays. Then were given 5 new multiple choice items at the end of the second day and participants had to answer without the documents available.

The means for coral bleaching essays are presented in Table 8. The essays were very similar to those from the pretest study presented above. We conducted a between participants ANOVA on each of the measures in Table 8 to examine whether condition mattered. The only significant difference was for the proportion of copied sentences, F (2, 57) = 3.998, MSE = .037, p = .024. Post hoc analyses found that, as expected, students writing with the documents available copied a larger proportion of sentences than those writing from memory.

Table 8. Descriptive statistics for the information in students' coral bleaching essays. Note, Standard deviations in parentheses.

	Included in essay		Explicitly connected to to-be-explained outcome			
Condition	Average number of sentences	Proportion of sentence copied or "nearly" copied	Average number of links mentioned	Average number of elements mentioned	Average number of links in explanation	Average number of elements in explanation
Docs present	9.00 (3.94)	0.27 (0.23)	2.57 (1.94)	8.43 (4.23)	1.76 (1.3)	2.67 (1.71)
Docs present, with warning	9.24 (3.21)	0.31 (0.19)	2.62 (2.11)	8.52 (4.82)	1.05 (1.12)	1.90 (1.61)
Memory, with warning	10.67 (3.11)	0.14 (0.15)	2.17 (2.09)	8.39 (4.79)	1.00 (0.91)	1.78 (1.4)
Overall	9.58 (3.47)	0.25 (0.2)	2.47 (2.02)	8.45 (4.53)	1.28 (1.17)	2.13(1.61)

As shown in Table 9, overall, students answered the interpreting the graph question at only a slightly lower rate (*M* = 37%) than the high school pretest sample (*M* = 46%) which was answered with the documents available. As expected, the new causal chain multiple choice items were answered at a much higher rate (*M* = 54%) than the high school pretest sample (*M* = 17%). This suggests that students can answer these causal chain items from memory and that asking students to coordinate filling in two blanks in a single question was not a good item. For the high school students, a score of 54% correct from memory is also promising because there is still room to detect growth. In comparing the three conditions, there were no significant differences among conditions. This suggests that even though students "copied" more in their essay, they still learned from the texts supporting our decision to allow students to write with the documents present. However, it may be with inference only items, we may find that copying is detrimental. These results also suggest that the interpretation of what the essay is as a measure may need to be examined (e.g., selection of important information vs. composed explanation vs. argument for a composed explanation).

Table 9. Percentage of students correctly answering multiple choice items (chance .25). Note. Standard deviations in parentheses.

Condition	Explanation items (4 items)	Interpreting graph item (1 item)	N
Docs present	51% (33%)	38% (50%)	21
Docs present, warning	61% (22%)	48% (51%)	21
Memory, warning	52% (34%)	24% (44%)	18
Overall	54% (30%)	37% (49%)	60

Detecting change over time. During the Year 5 reporting period, we have analyzed the complete data set from year 4 for this pretest and posttest assessment. For the pretest, 136 of the students were in middle school (6th and 7th graders), and 56 students were in high school (9th

and 11th graders). Students completed the prior knowledge inventory by marking their level of knowledge on a 5-point scale across several topics, and answered an open-ended question about the causes of the phenomenon. Because there were differences between topics and the middle school students received more of one topic, we will not collapse topic. For coral bleaching, high school students (n = 23, *M* =2.52, *SD* =.80) reported having the same amount of prior knowledge as middle school students (n = 78, *M* =2.62, *SD* =.69) and only 2 of the students correctly mentioned any part of the explanation to the open-ended question. For skin cancer, high school students (n = 26, *M* =2.53, *SD* =.79) reported having less prior knowledge than middle school students (n = 58, *M* =3.09, *SD* =.76). Overall, about 1/3rd of the students were able to mention 1 element of the explanation except the 6th graders (only 15% correctly mentioned any part of the explanation to the open-ended question).

The pretest baseline data and gives us some confidence in the validity of the essay assessment. Performance on the essay task was very promising. Even the middle school students wrote reasonably long essays. Students were able to identify some of the important elements of the explanation (about a quarter of the elements for coral bleaching and just under half for skin cancer). However, much of this information was not connected to the outcome to create a coherent or complete explanation. For example, most of the essays were simple cause to explanation with no intervening elements (e.g., shift in trade winds causes CB) and most were only a single path (see "Mentioned both unique causes" e.g., only trade winds OR salinity).

This lack of integration was likely a result of the high number of sentences in the essays that were untransformed from the documents (on average 69% and 49% of the sentences were borrowed or copied from the coral bleaching and skin cancer documents, respectively). This suggests that students are dealing with the rhetorical problem of writing an essay by not doing it – but are instead listing points that one might include in an essay. It is hard to know whether this is a production problem or an issue of understanding the goals for the task (transform and integrate to form multiple extended explanations). Even though the instructions were worded to deal with some of these issues based on the baseline studies, the instructions certainly could not eliminate the problems. In fact, these goal issues may be deeper issues with epistemic beliefs (e.g., phenomena have a single cause), which is addressed in the individual difference section of this report. We were able to detect some expected differences between grade level as indicated by an * in Table 10. Across both document sets, the high school students included more explanation elements, more connected explanations, and longer connected explanations.

Table 10. Baseline Pretest Performance. Significant grade differences are marked with an *.

	Coral bleaching		**Skin cancer**	
	High school	Middle school	High school	Middle school
Essays				
Number of sentences	11.00 (5.5)	9.34 (5.5)	12.08 (5.9) *	7.95 (5.0)
Proportion of sentences borrowed	.37 (.29)	.25 (.21)	.55 (.23)	.48 (.30)
Proportion of	.32 (.18) *	.17 (.12)	.51 (.20) *	.29 (.17)

explanation elements mentioned					
Average number of distinct explanations connected	1.13 (1.17) *	.65 (.89)	1.19 (.90)	*	.78 (.75)
Longest explanation	1.52 (1.34) *	.73 (.98)	1.65 (.88)	*	.88 (.90)
Mentioning at least one intervening cause	44%	17%	45%		28%
Mentioned both unique causes	0%	1%	8%		7%
Non-essay tasks					
Graph comparison selection (chance .50)	.50 (.51)	.59 (.50)	.68 (.48)		.48 (.50)
Graph comparison justification	.20 (.41)	.32 (.47)	.60 (.50)	*	.34 (.48)
Interpret Figure (chance .25)	.45 (.51)	.45 (.50)	.52 (.51)		.37 (.49)

Because it is difficult to infer understanding from an essay task, we included more targeted non-essay tasks. Performance on these measures provides converging evidence that students did learn but that they have much more to learn. For example, the graph comparison task shows that the single cause explanations typical of the essays was also a very attractive choice, and even when the more complete explanation was selected, students had difficulty explaining why it was preferred. One very promising result is that the main Figure 1 connecting the primary initiating cause (e.g., trade winds) to the outcome was correctly interpreted at a better than chance level.

Final Modification

Following analysis of the year 4 data, there were a number of concerns with the originally created documents and assessment materials that were addressed through revision and pilot testing. The results of the pilot testing informed which revisions provided more informative evidence of students' comprehension and explanatory modeling competencies. Based on the results from the pilot testing of the revised assessment, we finalized the EBA science assessment for use in the RCT in 9th grade biology classes.

Prompt and document set changes

The original essay prompt presented on the instruction page and the top of the essay writing sheet asked the student to either "…explain how and why coral bleaching rates vary at diffl erent times" or "…explain how and why rates of skin cancer differ around the globe." There was concern that asking "how and why" is too confusing for middle and high school students, but

asking just "how" or just "why" seemed too limiting. Students were often able to identify initiating causal factors, but they were lacking in their ability to identify intermediate steps leading to the outcome. The prompts were reworded to be more direct without limiting the meaning. The new coral bleaching prompt is "…explain what leads to differences in the rates of coral bleaching," and the new skin cancer prompt is "…explain what leads to differences in the risk of developing skin cancer." We also made small changes to the document set to encourage integration.

Post-essay assessment changes

There were a number of changes made to the post-essay assessment items. Because it is difficult to infer what students know from an essay, several multiple choice questions were added to the graph interpretation and causal explanation questions on the original assessment. Based on the essays from the pilot we were in a position to make up more targeted questions that reflected issues we noticed in the essays. These additional items were included to tap into students' understanding of proximal and distal connections at the beginning and end of the main causal chain, the secondary causal factor, and the normal process underlying the causal explanation.

A second change in the assessment materials involved the graphical model item. In this item, students choose the better of two graphical representations of the causal chain and then justify their choice. One representation shows an initiating factor leading to an intermediate factor leading to a second intermediate factor leading to the final outcome. The other representation shows three factors each leading directly to the final outcome. It was noted that for both topics the content for the various items matched except for one. It was determined, based on students' justifications for the better representation, that the lack of matching content might explain errors that were made. Therefore, the items were modified so that they matched completely across the two representations.

The final change involved the two peer evaluation items. For these items, students read essays that were supposedly written by other students with the task of writing down what the author did well and what needed improvement. These essays were purposely varied such that one provided both the main causal chain and the secondary causal chain but the language was vague and lacking direct causal information, and the other only provided the main causal chain but did so with direct causal language. Students' responses to these essays were focused on many issues other than explanation quality, so it was decided to switch to a more systematic manipulation that followed five dimensions across the pair of peer essays: relevance, completeness, coherence, citations, and deviation from normal process. These characteristics as well as accuracy were identified as the key components of good essays. Accuracy was not manipulated because the lack of accuracy would likely overshadow identification of any other essay flaws. So, the new peer essays include one essay that contains two of the five components listed above and the other essay contains the other three components.

Pilot Study

Pilot data were collected in May of 2014. Although the study was partly intended to test out the feasibility of including the additional multiple-choice questions, the primary goal was to test the effects of replacing the short answer item (the question about the secondary causal factor) with a pre-writing task. The instructions were as follows:

> "You will have the remainder of your class period to work on this pre-writing think sheet. You can jot down your ideas in an outline, in a bulleted list, by drawing a diagram, or in whatever way helps you to collect and organize your ideas for writing. Tomorrow you will use this think sheet to help you write your essay explaining how and why coral bleaching (skin cancer) rates vary at different times (around the globe). "

This change was motivated by a frequent observation made during essay scoring. Students often wrote their explanations in ways that suggested an understanding of the relevant content but a failure to adequately link the content in a meaningful way. Because of this, their essay scores were fairly low. It was thought that maybe students needed an opportunity to revise their essays so that their final answers reflected an understanding of both the content and the connections among the content. The time allotted to the assessment was not enough to allow for essay revision, but it was possible that students could spend a brief amount of time writing notes to organize their thinking prior to beginning their essays.

Participants included a new group of 214 ninth grade biology students from seven schools in the greater Chicagoland area. These students were not participating in other Project READI activities. The study was set up just like the traditional assessment procedures. Students completed demographics and prior knowledge questions prior to beginning to read the documents. After a set amount of time, they were asked to begin writing their essays. The essays were followed by a series of post-essay items: multiple-choice questions, graphical model selection, and peer essay evaluations. Unlike the standard assessment procedures, half of the students completed the pre-writing task in between the reading task and the essay writing. Data collection took place in the classroom during two consecutive regular class meetings.

The feasibility of including additional multiple-choice items (the initial motivation for the study) was supported by the data. As with the standard assessment, most students were able to complete all items in the allotted time; however, the data suggested less support for the usefulness of including the pre-writing task. Because there was very little time between data collection and decisions about the final assessment materials, it was determined that the multiple-choice scores would be used to measure the effectiveness of the pre-writing task. There was no significant difference in the mean proportion correct on the nine multiple-choice items between the two groups [$t(191)=-0.574$, $p=0.566$]. The mean proportion correct for the pre-writing group was .49 compared to .51 for the no pre-writing group. Based on this finding as well as the concern that students may not feel compelled to write a thorough essay if they already wrote a lot in the pre-writing task, it was determined that the pre-writing task would not be included in the final assessment materials.

The responses to the new peer essay task were much more focused on the features we manipulated. The presence and absence of relevance (e.g., "The explanation was well written in general because this student gave great detail on how coral starts to bleach.", "Need more detail on coral bleaching and it's rates and less on what coral is.") and coherence (e.g., "Explaining about the temperature and the 'cause and effect' of what if the water temperature rose.", "I think the student could have given some more explanation about the main question that you should answer") were detected at a much higher rate. On a few students mentioned completeness (e.g., "I would say the student needed to talk about the importance of salt and the job of the algae."), citations, or deviation from normal process.

(See Appendix 1 for History EBA Assessment; See Appendix 2 for Science EBA assessment)

Development of EBA Assessments for Literature

Susan R. Goldman, Carol D. Lee, Stephen Briner, Sarah Levine,
Joe Magliano and Masha Yukhymenko*

Understanding literary texts involves moving beyond the literal meaning to constructing interpretative meanings that explore the messages or morals of the text with respect to human nature and the world in which we live (Langer, 2010; Lee, 2007; 2011). Empirical research on experts and novice readers of literature, including high school students, indicates that for the most part adolescent readers take a literal stance toward literature and do not tend to adopt an interpretive stance (Goldman, McCarthy, & Burkett, 2015; Graves & Frederiksen, 1996; Peskin, 1998). Accordingly, a fundamental goal of READI literature interventions was to support students in going beyond literal meaning to understand themes about human nature using sophisticated literary reasoning based on analyses of features of literary texts such as rhetorical devices (e.g., symbolism, unreliable narration), plot and character types (Lee, 2011). This is a complex problem solving activity even for a single text. Nevertheless, we wanted students to reason intertextually as well as intratextually about features of texts so that contrasts and comparisons would come into play in the reasoning process. These goals required us to develop an EBA for literary reading.

Development of the literature EBA was informed by the READI core constructs in literature and the learning goals for literature so that it would align with the intervention and thus be sensitive to effects of the READI instructional approach (Goldman, et al., 2016; Lee, Goldman, Levine, & Magliano 2016). We also drew on a hierarchy of literature comprehension skills validated by Hillocks and Ludlow (1984). Their hierarchy distinguished between three fundamental types of skills: comprehension of basic story elements, complex inferences, and generalization in literary reasoning of two types – structural and author. Structural and author generalization reflect reasoning about features of the literary work such as word selection and patterns of use, plot type, and characterization to conjecture about potential author-intended message(s) beyond the story world.

Phase 1: Initial Development

The initial effort to develop an EBA literature assessment examined basic, complex, and generalization processes using a series of tasks that culminated in an essay for which the prompt asked students to compare two stories with respect to the image used by the author of each to convey a message about the world. These studies are reported as baseline study 1 and 2 in the basic studies technical report (Technical Report #8) and are simply summarized here, as background to the eventual form of the literature EBA. Across the two studies, approximately 225 students distributed across grades 6 – 12 participated in the baseline studies. The essays students produced showed little evidence of comparison of the stories, very rudimentary evidence of noticing symbols, and almost no connection of symbols to author message. We also found that students who completed the basic and complex inference tasks prior to the essay often

* With acknowledgement to Teresa Sosa, Allison Hall, Angela Fortune and Candice Burkett.

did not complete an essay, likely due to the two-class period time constraint. Conversely, those students who wrote the essay prior to completing the basic and complex inference questions, typically did not complete these.

Despite the issues with task completion the essays collected during these baseline studies allowed us to compare two forms of scoring. The first used four broad categories (Addressing the task, Synthesizing, Literary reasoning, and Structure of argument) that proved too ambiguous and global to apply to the essays in a reliable manner. We then developed a set of dimensions that we applied to the essays we had collected from the 225 students. The definitions of these dimensions and the rubric score points for them were refined in collaboration with the design team teachers as they used them in their iterative enactments of the literature interventions. The basic dimensions were claim, function of claim, evidence, reasoning, organization of ideas, symbolism, coming of age theme, and synthesis.

Phase 2: Refinement and Revision

We refined and revised the EBA assessment as we used it in the context of the literature module implementations in middle and high school classrooms. We intended the EBA assessment to function as a pre/post assessment of changes in dimensions of literary interpretation. Another goal in the design was to have students work with different stories at post test than they had worked with at pretest so in the Phase 2 development work we attempted to find story pairs that we could "equate" for difficulty. Equated story pairs would allow us to counterbalance across time of assessment and students.

In the context of the initial iteration of the literature modules focused on symbolism and coming of age, we administered a version of the EBA assessment to 9th and 12th grade students. (See Technical Report #8 for description of this work.)

Students in 9th and 12th grade read two sets of texts. One set consisted of "Flowers" by Alice Walker and an excerpt from "We Were the Mulvaneys" by Joyce Carol Oats. The other set consisted of "Eleven" by Sandra Cisneros and "The Butterfly" by James Hanley. Students read and wrote essays responding to both sets of texts. About half of the students read the "Flowers" and "Mulvaneys" set first and the "Eleven" and "Butterfly" set second (i.e., the Flowers First Condition) and the rest read "Eleven" and "Butterfly" first and "Flowers" and "Mulvaneys" second (i.e., the Eleven First Condition). Essays were coded according to the rubrics shown in Table 11. Note that with the exception of Synthesis, each rubric is applied to each text and the scores are averaged to yield one score on each rubric for each student. There is only 1 synthesis score.

Table 11. Dimensions of Essays Assessed and Score Point descriptions

Dimension	Rubric Score Points				
	< -----Lower ----------Higher--------->				
Claims	Description; Story Summary; Restates question (0)	States claim but not supportable by text (.5)	One supportable claim (1) or more than one supportable claim (2)	More than one supportable claim; claims connected (3)	More than one supportable claim; claims nested (4)
Function of	Story		Describes	Language of	

Claims	Summary; Restates question (0)		internal state of character or describes social world of text (1)	text examined to explain internal character state and/or social world of text (2)	
Evidence	No evidence provided (0)	Only inaccurate evidence provided (.5)	One piece of evidence (1) or more than one (2) based on text	More than one piece and belong to same pattern of repetition, prominence, etc. (3)	Comprehensive evidence and competing evidence (counterclaim) (4)
Reasoning	No discussion of why evidence supports claim (0)		Claims and evidence logically connected implicit (1.1) explicit (1.2)	Claims and evidence supported by appeal to cultural norms Implicit (2.1) Explicit (2.2	Claims and evidence supported by appeal to literary norms implicit (2.3) Explicit (2.4)
Organization of Ideas	No clear organization (0)		Some organizational framework (e.g., claims and warrants in proximity) (1)	Well-organized; connections between sections explicit (e.g., *but, in contrast)* (2)	
Language and Structure	Language and syntax unclear, difficult to follow (0)		Language and syntax are mostly clear with some lapses (1)	Language and syntax are very clear (2)	
Symbolism (rhetorical device)	No identification of symbols or attribution of symbolic significance (0)	Identifies symbol; no interpretation (1) Identifies symbol and interprets but interpretation not supported by text (1.1)	At least one symbol identified; interpretation local to text (2)	At least one symbol identified; interpretation local to text but also extended beyond (2.1)	

Coming of Age (theme)	No discussion of development or evolution of character (0)		Discussion of development or evolution of character in terms of criteria of coming of age (1)	Discussion of development or evolution of character in terms of criteria of coming of age AND connects to symbolism in text (2)	
Synthesis	No connection; stories discussed or analyzed separately (0)	Surface connection between stories (1)	Explicit compare or contrast of claims about characters or worlds across stories (2.1) About symbols across stories (2.2)	Explicit compare or contrast of claims with evidence or reasoning about character or social worlds (3.1) About symbols (3.2)	

Analyses of the rubric scores from the pre/post intervention assessments revealed the following.

- In general, 9th graders who read Flowers/Mulvaneys first tended to score better at posttest than at pretest, suggesting improvement. Students scored significantly better at posttest than at pretest for Symbolism, t (10) = 2.89, p = .02. Students were also marginally better at posttest than at pretest for Evidence, t (10) = 1.92, p = .08.
- 9th graders who read Eleven/Butterfly first tended to actually perform worse at posttest than at pretest for most of the rubrics. Scores were significantly worse at posttest than pretest for Evidence, t (13) = 2.59, p = .02; Function of the Claims, t (13) = 2.51, p = .03; Coming of Age, t (13) = 2.86, p = .01, and Synthesis, t (13) = 3.04, p < .01. However, students performed significantly better at posttest than at pretest for Organization, t (13) = 2.11, p = .05; and Symbolism, t (13) = 2.83, p = .03.
- 12th graders who read Flowers/Mulvaneys first tended to score better at posttest than at pretest, suggesting improvement. Students were significantly better at posttest than at pretest for Claims, t(18) = 3.88, p < .01; Evidence, t(18) = 3.11, p < .01; Reasoning, t(18) = 4.43, p < .01; Function of the Claims, t(18) = 3.36, p < .01; Symbolism, t(18) = 3.57, p < .01; and Coming of Age, t(18) = 3.59, p < .01.
- 12th Graders who read Eleven/Butterfly first tended to score better at posttest than at pretest, suggesting improvement. Students were significantly better at posttest than pretest for Reasoning, t(13) = 2.21, p = .04; Function of the Claims, t(13) = 4.32, p < .01;

and Symbolism, $t(13) = 4.20$, $p < .01$. Students were marginally significantly better at posttest than pretest for Evidence, $t(13) = 2.01$, $p = .07$.

The Flowers/Mulvaneys pair was more difficult for 9th graders than for 12th graders. 12th graders scored better on both the pre and post-tests on this pair. These findings suggest that the Flowers/Mulvaneys text set was more challenging than the Eleven/Butterfly set. Yet, even so, students reading Flowers and Mulvaneys improved on one of the main targets of the intervention, namely identifying and interpreting symbolism. We still find developmental differences as 12th graders on the whole performed better on both pre and post-tests for the Flowers/Mulvaneys set. Nevertheless, these pairs of texts failed to meet the criterion of equivalent difficulty across the grade span we wished the assessment to serve. Our analysis of the texts suggested that the excerpt from Mulvaneys was difficult for younger adolescents because of the indirect manner in which loss of a family member was portrayed. Content analyses of the essays indicated that students understood that there was a dead body in Flowers but not that the protagonist in Mulvaneys was reacting to death. We decided to find an alternative story to pair with Flowers in the next iteration of the assessment design.

Assessment Design, Phase 2, Iteration 2: We administered the pre and post version of the coming of age/symbolism assessment to one class of 9th grade and one class of 8th grade students. We again used the story set "Eleven" and "Butterfly" but paired "Flowers" with an adapted version of the Leslie Norris story "Shaving." The adaptations were needed to bring it closer to the length of the other three stories. Although we had samples of under 20 in each of the classes, we again counterbalanced story set across pre/post so that students worked on one story set at pre and the other at post. Essays were scored for the same dimensions and with the same rubrics as shown in Table 11. Results of analyses of the pre/post scores indicated that the assessment was sensitive to change for several of the dimensions in both the 8th and the 9th graders. Table 12 shows the performance changes for the 9th grade students along with the results of a Wilcoxon matched pairs signed rank test on each coding dimension.

Table 12
Wilcoxon Results for Changes in Dimension Scores for Pre- and Post- Assessments

	Negative Ranks	Positive Ranks	Ties	*Z*	*p* value
Claims	0	6	5	-2.21	0.03*
Function of Claims	0	3	8	-1.73	0.08
Evidence	1	8	2	-2.48	0.01*
Reasoning	2	3	6	0.00	1.00
Symbolism	1	7	3	-2.43	0.02*
Coming of Age	0	2	9	-1.41	0.16
Synthesis	1	6	4	-2.16	0.03*
Organization	0	4	7	-2.00	0.05*
Language and Syntax	0	4	7	-2.00	0.05*

*Significant at $p \leq .05$

The low frequencies of negative ranks shown in Table 12 indicate that on all dimensions students improved or maintained their pretest levels of performance. Improvements were significant for several dimensions, including claims, evidence, symbolism, synthesis, organization and language and syntax. However, reasoning and theme did not improve significantly nor as much as we had targeted. The 8th grade students were significantly more likely to achieve higher scores at posttest than at pretest for Evidence ($Z = 2.23$, $p = .03$), Symbolism, ($Z = 2.76$, $p = .01$) and Coming of Age ($Z = 3.16$, $p = .002$). Nevertheless, closer analyses of the essay content revealed that we had not achieved equivalent text sets. This was partly due to the impact on coherence of the adaptations we had made to shorten the story "Shaving." In conjunction with the fact that the reasoning dimension reflected little change from pre to post, we decided to alter the form of the EBA in literature to more explicitly provide evidence relevant to inferences about symbol identification and reasoning about symbols in each story. The changes are described in Phase 3.

Phase 3: Literature EBA: Final Form

The final form of the EBA for literary reading captures both basic comprehension of the story events and interpretive meaning, as originally intended. However, repeated efforts to develop story sets of "equal" complexity fell short of achieving this goal. There are a variety of reasons related to characteristics of the texts as well as individual differences in psychosocial development within grade bands, and experiential differences within and across grade bands. Lee and Goldman (2015) discuss these issues in detail. Because of the significant problems of equating text complexity we decided to use the same story set at both testing times. Performance at time 1 could then be used in analyses as a covariate or predictor variable in examining performance at time 2. Although not an ideal approach, we felt it was better than using text sets that we knew were not equated for complexity of literary interpretation. We were also concerned that the production demands of the essay were masking analytic and interpretive processes related to symbolic and thematic interpretation. In the DBR enactments it often appeared to be the case that students' oral discussions of literary works reflected more reasoning from textual evidence to interpretive claims about the story world and life more generally than appeared in written essays. These considerations led us to organize the EBA assessment so that we could develop estimates of basic comprehension as well as processes associated with taking an interpretive stance toward the literary work, as we describe in the remainder of this report.

Design of the Final EBA of literary reading

The EBA literature assessment requires two days to administer. The design of the literature EBA assessment task differs from the EBAs in history and science in that students write an essay after doing two tasks that assess components of the literary analysis and reasoning skills that are involved in writing the essay. We organized these tasks in this order intentionally and based on our experiences with prior versions of the assessment. There are several components of writing the essay, each of which might impact the content and quality of the written essay response. These components are (1) quality of close reading of the text at the story grammar level; (2) ability to use rules of notice in order to decide what details of the story hold symbolic significance; (3) the ability to organize salient details in a comprehensive manner to

structure claims and evidence for both themes and symbolic import; and (4) the rhetorical problems of translating one's ideas into a coherent argument, including the ability to structure points of comparison and contrast across stories. As noted earlier, in our experiences administering prior versions of the EBA, it had been difficult to disentangle these different skill sets in evaluating the quality of students' responses to the open-ended essay writing prompts.

The final design was structured as four tasks sequenced to allow us to better sort out these components: story reading and annotating, event recognition and sequencing, heuristics for interpretation structured as a graphic organize chart, and writing one essay about the two stories. For example, performance on the event recognition and sequencing task provided an opportunity to assess memory for explicit content and an index that could be correlated with the quality of the essays produced. Similarly, the graphic organizer allows us to interpret the essay content in terms of qualities of interpretative comprehension (e.g., what did students pick out as important symbols and what do they think they mean for each story and in comparison to one another) separate from quality of rhetorical structure of the argument as reflected in the essay.

Task Sequence

On the first day, students are asked to read and annotate each story and complete the event recognition and sequencing task. Note that when administering the task, story order is intended to be counterbalanced across participants.

Stories. We selected two stories that we had used in several previous versions of the EBA: *Eleven* (Cisneros, 1991) and an adapted version of *Butterfly* (Hanley, 1961). These two stories had also been used in basic studies with convenience samples of undergraduates as well as with adolescents across the grade 6 to 12 range. These data suggested that the two stories are accessible in terms of social maturation for both middle and high school students; yet afforded opportunities for more mature social insights with older students that we might expect to see in the types of claims and the reasoning supporting them. Both stories embody coming of age themes and symbolism is a prominent rhetorical device in reasoning about possible interpretations of each story.

Event Recognition and Sequencing. A list 10 events for each story consisting of five that occurred in the story and five that did not were presented in a fixed randomly determined order. The events that did not occur in the story use the same characters as the story and are plausible but actually contradict the story events. (See Appendix 3 for list of events.) Students are asked to first decide which events were in and which were not in the story; they are then asked to indicate the order of occurrence of those events they decided were in the story. Students complete this task after they have read and annotated each story. Thus, day 1 of the assessment involves reading and annotating each story and completing the basic comprehension assessment for each.

On day two, students complete the heuristic graphic organizer and write the essay. The instructions on day two are the following.

> *Today, you will write an essay about the two stories that you read yesterday. When writing stories, authors often create symbols that help reveal ideas about the way people are and the way the world is. Symbols are words or phrases in stories that stand for more than what they seem to be. Symbols can be images,*

actions, objects, or characters (what they do, how they think, how they look, their names.)
To help you write your essay, please begin with this chart (refer to chart in student packet). It will help you organize your thinking so that you can write your essay. You may refer back to the stories as much as you like. After 15 minutes, we will let you know that it's time for you to move from the chart to writing the essay if you have yet to do so.

Heuristic task prompts chart. Three questions about each story and three questions comparing/contrasting the two stories are presented in a graphic organizer chart as illustrated in Table 13. The questions in the chart are intended to "walk" students through the thought processes that would be required to identify important symbols in each story and then compare them (see Table 13). After working on the chart for 15 minutes, students are directed to go on to the essay, using the chart in their writing process. Students have the texts including whatever annotations they made on day 1 available to them on day 2.

Table 13. Heuristic task prompts in the graphic organizer chart

Eleven	**Butterfly**	**Compare and contrast**
What do you think are one or two of the symbols in the story "Eleven"?	What do you think are one or two of the symbols in the story "The Butterfly"?	Compare and contrast one symbol from each story.
Explain why you think they are symbols.	Explain why you think they are symbols.	Compare and contrast the reasons you think they are symbols.
In what ways does each symbol show us something important about the main character and their world? Be sure to explain why.	In what ways does each symbol show us something important about the main character and their world? Be sure to explain why.	Compare and contrast how the symbols might show something important about the main characters and their worlds.

<u>Essay task.</u> The essay prompt is designed to integrate author generalization (i.e. theme) and structural generalization (i.e. how the author uses language and structure to convey ideas). Students are asked to write an extended essay in which they discuss how the theme or themes of the story are conveyed through the use of symbolism.

Deriving Results for the Different Tasks

The event, graphic organizer, and essay tasks were each scored using different approaches for capturing performance due to the diverse nature of the tasks and types of responses that indicate more sophisticated versus less sophisticated task performance. As well, data on each task were available for different subsets of students. Accordingly, in this section we discuss each task from the perspective of how it was coded and scored and the results that that process generated.

Event recognition and sequencing task

The event-sequencing task was scored in two ways. The first, the *event detection score,* indexed the correct recognition of events that were in the story and discrimination of them from events that were not in the story. The event detection score was computed using a measure of sensitivity in signal detection called A'. Conceptually A' involves subtracting false alarms (incorrectly indicating that an event on the list was in the story, but it was not) from hits (correctly identifying that a event was in the story and it was). A' scores range from 0 - 1.0, with .50 indicating that the student is performing at chance. The formula for A' is presented below based on the work of Snodgrass and Corwin (1988) and Stanislaw & Todorov (1999).

Hit = correctly identifying that event was in story
FA – incorrectly indicating than an event that was not in the story was in the story

When Hits >= to FA:

$$A' = 0.5 + \frac{(Hits - FA)(1 + Hits - FA)}{4Hits(1 - FA)}$$

When FA > Hits:

$$A' = 0.5 + \frac{(FA - Hits)(1 + FA - Hits)}{4FA(1 - Hits)}$$

The second index, the *longest chain score*, referred to the longest chain of events that was correctly sequenced. The highest possible score was 5 and the lowest possible score was 0. The longest chain captured the number of events that were correctly sequenced, independent of the specific ordinal label used by the participants. For example, if participants correctly labeled events 1-3 with those numerical values, but did not correctly order events 4 and 5, they would get a score of 3. However, they would also receive a three if those same events were ordered 3-5 because the longest chain of correctly sequenced event was still three.

Analyses assessing changes in scores pre and post intervention. 209 participants provided data on the event recognition and sequencing task for both the pre and post intervention sessions. These participants were students in 8th (n = 86), 9th (n = 30) and 11th (n = 93) grades who were participating in enactments in several different design team teachers' classrooms. Preliminary analyses indicated that there were no grade differences on any of the variables. We re-analyzed the data without this variable. Table 14 contains the means and standard errors for the pre-post event task scores. A repeated measures 2 (Story) X 2 (Pre-post) ANOVA on the event detection scores revealed a significant main effect of story ($F(1, 208) = 14.24$, $Mse = .008$, $p < .001$) that was qualified by a significant two-way interaction ($F(1, 208) = 11.61$, $Mse = .008$,

p = .001). This interaction indicated that detection scores were higher pre-post for *Butterfly*, but not *Eleven*. A repeated measures 2 (Story) X 2 (Pre-post) ANOVA on the longest chain scores revealed a main effect of pre-post (F (1, 208) = *4.59, Mse = 1.74, p = .033), such that scores where higher post intervention (M* = 2.79, *Mse* = .09) than pre intervention (*M* = 2.57, *Mse* = .09). Additionally there was a main effect of story (F (1, 208) = 4.59, Mse = 1.74, p < .001), such that scores where higher for Butterfly (*M* = 2.96, *Mse* = .10) than for Eleven (*M* = 2.40, *Mse* = .09).

Table 14
Event detection and longest chain scores

	Butterfly		*Eleven*	
	Pre	Post	Pre	Post
Event detection score	.87 (.01)	.93 (.010	.93 (.01)	.92 (.01)
Longest chain score	2.81 (.12)	3.10 (.12)	2.34 (.11)	2.47 (.12)

These data indicate that students were highly sensitive to which events had occurred in the story and which had not, and where there was room to improve, i.e., on the Butterfly story, students did. At pretest on Eleven, sensitivity was already close to ceiling and it remained at that level on post. Furthermore, there were improvements from pre to post on both stories in correctly ordering events that were detected as being in the story. These data suggest that this measure provides a promising and easy to implement and score assessment of comprehension of the basic plot of these stories.

Analyses assessing the correlation of the scores with essay performance in the pre and post sessions. The event recognition and sequencing task was developed to support essay writing by encouraging close reading of the stories. To evaluate the relationship between performances on each of these tasks, correlations were run between the sensitivity and detection and longest chain scores for claims, function, evidence, reasoning, and symbolism dimensions for each story. (see Tables 15 and 16).

Table 15.
Pre and post intervention correlations for *Butterfly*.

	Pre Intervention		Post Intervention	
	Event Detection	Longest Chain	Event Detection	Longest Chain

Claims	0.12	*0.15*	*0.14*	**0.17**
Function	**0.21**	**0.23**	-0.05	0.05
Evidence	-0.02	**0.17**	-0.01	0.02
Reasoning	0.11	**0.23**	0.06	0.08
Symbolism	**0.20**	0.13	0.07	**0.22**

Note: Bolded coefficients had $p < .05$ and italicized coefficients had $p < .10$

Table 16.
Pre and post intervention correlations for *Eleven*.

	Pre Intervention		Post Intervention	
	Event Detection	Longest Chain	Event Detection	Longest Chain
Claims	*0.15*	**0.23**	0.10	0.08
Function	-0.05	0.12	0.07	0.00
Evidence	0.12	**.22**	0.11	0.06
Reasoning	0.12	*0.14*	0.10	0.05
Symbolism	*0.08*	0.09	0.10	0.10

Note: Bolded coefficients had $p < .05$ and italicized coefficients had $p < .10$

As hypothesized performance on the event recognition and sequencing task was correlated with dimensions in the scoring rubric, but more strongly at pre than at post intervention. This is interesting because it suggests that the intervention may have improved students' attention to the details of the plot sufficiently to reduce the variability among students in basic comprehension, thereby reducing the correlations. Put another way, prior to the intervention, basic comprehension indexed by sequencing events that had occurred in the story predicted several dimensions of the essays. Post intervention this relationship was largely absent. Thus variation in essay performance at posttest could not be attributed to basic comprehension of story events. More research is needed to confirm this potential effect of the intervention and the event ordering task and scoring developed in project READI could be valuable in said research.

Heuristic Graphic Organizer Task

The following rubric-based scoring system was developed to code and score student responses on the graphic organizer. A series of dimensions and ordered score points were developed for each part of the organizer as described below. (See Table 17.) Independent raters achieved reliability of .82 when using this system, indicating acceptable inter-rater reliability.

Table 17. Dimensions and score points for each question on the heuristic task prompts in the graphic organizer chart

Heuristic question prompt	Dimension	Rubric Score points
What do you think are one or two of the symbols in the story?	Frequency	0 = No symbols identified 1 = Uses very frequent symbols (“Butterfly” = Caterpillar; “Eleven” = Sweater) 2 = Uses less frequent symbols (“Butterfly” = Butterfly; “Eleven” = Age) 3 = Uses another symbol (All other symbols account for < 10% of data) 4 = Mixed (e.g., one very frequent and one “other”)
	Type	0 = None identified 1 = Objects (e.g., sweater, box) 2 = Entities (e.g., caterpillar, Brother Timothy) 3 = Situations (e.g., Rachel’s birthday) 4= Abstractions (e.g., growing up, birthdays) 5 = Mixed (e.g., one object + one entity)
Explain why you think they are symbols	Type	1 = Student answers by explaining what s/he thinks symbol represents 2 = Student answers by referencing rule of notice (“The sweater appears frequently throughout the story.”) 3 = Student references rule of notice *and* explains what symbol represents
	Depth of explanation	0 = No explanation given 1 = Symbol given in row 1 is supported by intuition or knowledge about the world (“Growing up is a part of your life”; “I think it stands for how mature or free a person feels.”) 2 = Symbol given in row 1 is supported by details of plot (“Because in the story Cassidy had to be honest about why he missed mass”)
‘In what ways does each symbol show us something important about the main character and their world?	Type of response	0 = No response 1 = Response based on personal opinion or knowledge about world (“This character might be a liar.”) 2 = Response consistent with details of plot (“Friendship is important, and the boy grew fond of it, so it shows how much friendship

		can mean to him.”) 3 = Response draws parallels between text and world (“Because caterpillars represent hope, the crushing of the caterpillar represents the crushing of Cassidy’s hope”), or between elements within the text (“It shows that the author looks at the caterpillar as freedom…and when the brother kills it, he takes his freedom too”).
Compare and contrast one symbol from each story	Type of response	0 = No response 1 = Description of one symbol (“The caterpillar is alive”) 2 = Symbols described separately (“The red sweater was something that can be misunderstood. The caterpillar showed that the main character might be innocent.”) 3 = Symbols compared OR contrasted (“They both make somebody mad.”) 4 = Symbols compared AND contrasted (“The red sweater and the caterpillar are not the same because in one story it was angry and the other happiness. They’re the same because they both symbolize emotion.”)
Compare and contrast the reasons you think they are symbols.	Type of response	0 = No response 1 = Comparison and/or contrast of features of or characters’ response to symbols (“They are both something someone can mentally have.”) 2 = Comparison and/or contrast of reasons (“The role of the color red and of caterpillars in the real world are clues that the red sweater and the caterpillar are symbols.”)
Compare and contrast how the symbols might show something important about the main characters and their worlds.		0 = No response 1 = Statement of knowledge or opinion (“They can be good kids but when they grow up they can change good to bad.”) 2 = Comparison and/or contrast of features of or characters’ response to symbols (“The caterpillar made Cassidy happy but the sweater made Rachel upset.”) 3 = Comparison and/or contrast of functioning of symbols in the story (“In each story, the reader observes the character’s actions, but it is through their interactions with and reactions to the symbol that we understand their feelings.”)
Connecting organizer to essay	Type of response	0 = No comparison or contrasting of stories in essay

		1 = Comparison and/or contrast of stories in essay departs from thinking in G.O. 2 = Comparison and/or contrast of symbols (given in row 1) included in essay (but not reasons or functioning) 3 = Comparison and/or contrast of symbols (given in row 1) and reasons (given in row 2) included in essay (but not functioning) 4 = Comparison and/or contrast of symbols (given in row 1), reasons (given in row 2), and functioning (given in row 3) included in essay

Analyses assessing changes in scores pre and post intervention. 157 participants provided data on both pretest and posttest heuristic organizer charts (70 8th graders, 33 9th graders, 54 11th graders). These students were participating in enactments in design team teachers' classrooms.

8th Grade: There were no significant differences in response distributions for any of the questions. At both pre and post, about half of the sample identified no symbol and between 15% - 30% of the students named the salient symbol in each story.

9th Grade: Ninth graders produced more "frequent" symbols at posttest for both The Butterfly and Eleven. For eleven, only 21.2% of participants identified a "frequent" symbol at pretest, and this rose to 36.3% at posttest ($p = .041$). For The Butterfly, 33.3% of the students identified a frequent symbol at pretest, which rose to 54.5% at posttest ($p = .003$).

For types of symbols identified, we also saw significant movement for 9th graders from pre to post for both Eleven ($p = .04$) and The Butterfly ($p = .043$). For Eleven, this change was driven by a greater percentage of students identifying an object (e.g., the sweater) as a symbol at posttest (39.3%) compared to pretest (18.2%). For the Butterfly, students went from identifying an entity (e.g. Brother Timothy or the Caterpillar) 42.2% of the time at pretest, which rose to 57.7% at posttest.

11th Grade: The differences that emerged for 11th graders suggest that they come to rely on the graphic organizer less at posttest than at pre. For Eleven, more students failed to provide a symbol at posttest (11.1%) than at pretest (2.0%), $p = .043$.

For Depth of Explanation, fewer students provided an explanation at posttest (24.1%) than at pretest (6.0%) for The Butterfly, $p = .033$.

Students in Grade 11 also provided fewer explanations about what the symbol shows about the characters and their worlds. This was true for both Eleven (pretest: 3.7% failed to explain compared to the posttest, where 22.2% failed to explain, $p = .009$) and The Butterfly (pretest: 3.7% versus posttest: 29.6%, $p = .002$, failed to explain).

A significantly greater proportion of students neglected to compare the symbols between the stories at posttest (42.5%) than at pretest (18.5%), $p = .02$.

Overall, these data suggest that younger students may make more use of the graphic organizer than older students, who are substantially more likely to skip sections of the graphic organizer at posttest compared to pretest. However, this pattern suggests that graphic organizers may be especially useful for helping younger students, such as our 8th and 9th grade samples, scaffold their ideas before writing an essay.

Correlating Graphic Organizers to Essays

To explore connections between the graphic organizer and the essay, we compared the graphic organizer score on "What does the symbol tell us about the characters and their worlds," with the Reasoning rubric on the essay scores. The essay rubric scores were recoded for this analyses as follows: 0 (No reasoning) 1 (General reasoning), 2 (reasoning about cultural norms) and 3 (reasoning about literary norms).

At pretest, 8th graders showed no correlation between their graphic organizer and their reasoning scores, $p = .63$. However, there was a moderate positive correlation between GO scores and Reasoning scores for the 9th graders ($r = .395$, $p = 023$) and for 11th graders ($r = .330$, $p = .015$). At posttest, there is a trend *towards* a positive correlation between GO scores and Essay Reasoning scores ($r = .210$, $p = .81$). The correlations are even stronger at posttest for 9th graders ($r = .441$, $p = .01$) and 11th graders ($r = .626$, $p < .001$).

Overall, these correlations show us that at pretest, older students are more likely to show a connection between their reasoning in the graphic organizers and their reasoning on the essays, and these relationships are even stronger at posttest than at pretest.

Synthesizing the Essay Data

For instructional purposes, having detailed dimensions and rubrics that provide descriptive characterizations of students' thinking is very useful. However, with 9 dimensions and multiple score points, it can be cumbersome for taking a more global view of performance across a group or groups of students and for summarizing how the intervention impacted students. A more global summary of performance and changes in performance seems especially beneficial when students "start" at different points and may be expected to benefit in different ways from the intervention. We decided to employ a latent class approach to determining whether we could identify patterns across the scores on the 9 rubrics and whether these patterns constituted interpretable clusters.

During the 2013- 2014 academic year we collected, scored, and analyzed data from 247 students in classes with 7 different teachers distributed over 5 schools. All of these students were participating in some form of the READI approach to literature, with variation in amount of time, literary pieces read, emphasis on argumentation and writing, and grade level (grades 8, 9, and 11 were represented). Essays were scored by two raters who achieved an intraclass correlation of .80 with a subset of essays prior to scoring. During scoring of all of the essays, these researchers scored five of the same essays and achieved an intraclass correlation of .76. Of the 247 students, 156 students provided both a pre- and post-test and analyses were conducted on this subset.

Latent class and discriminant function analyses. Students' essays received scores for their claims, evidence, reasoning, etc. for *The Butterfly* and *Eleven.* To make analysis more straightforward, we took a "best score" for each dimension of the rubric. Previous use of this procedure had indicated that over 90% of the students scored the same way on the 9 dimensions scored for each story.

Pre-test essay scores. Using the rubric scores from the pre-essays, a latent class analysis identified three homogenous patterns of scores. Entropy, which measures classification certainty, was .90; the average latent class probabilities for the most likely class membership ranged from .94 to 1. We then performed discriminant analysis to understand better what characterized the three patterns and differences among the three groups. Students in Group 1 (n = 33, 21%) were characterized by low rubric scores on all of the dimensions. Students in Group 2 (n = 69, 44%) were characterized by high rubric scores on all the dimensions. Finally, students in Group 3 (n = 54, 35%) were characterized by relatively high scores on literary comprehension dimensions (Claims, Reasoning, Evidence) but relatively low scores on writing (Organization, Syntax). Table 18 provides *Standardized Canonical Discriminant Function Coefficients* and Figure 4 shows the discriminant function analysis depicting group profile differences in relation to Function 1 and 2.

Table 18. Standardized Canonical Discriminant Function Coefficients[a]

	Function 1	Function 2
Claim	0.37	0.47
Function	-0.07	-0.09
Evidence	0.24	-0.51
Reasoning	0.70	-0.40
Symbolism	0.12	-0.17
Coming of Age	0.05	0.29
Organization	0.09	0.26
Syntax	0.40	0.39
Synthesis	0.31	0.26
Variance Explained	89.6%	10.4%

[a]Standardized canonical discriminant function coefficients for each essay dimension show which rubrics were important in determining Function 1 and 2 (the larger the absolute number is, the more important the rubric is)

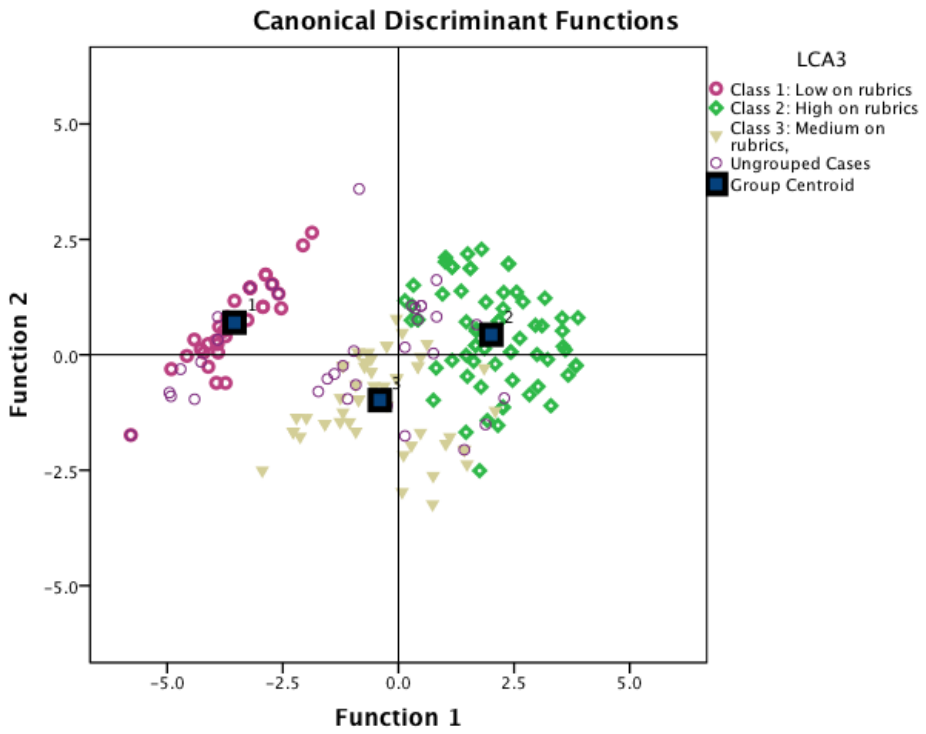


Figure 4. Results of the Pre-test discriminant function analysis depicting class profile differences in relation to two discriminant functions, which represent the level of comprehension and writing (x-axis) and the difference between writing and comprehension (y-axis). Note that Group 3 is labeled medium is a shorthand to reflect its relationship to Group 1 and 2. In fact, students in Group 3 did better on the literary comprehension and argument dimensions than on the writing dimensions.

Postest essay scores. Using the rubric scores from the post-essays, a latent class analysis again identified three homogenous patterns of scores. Entropy, which measures classification certainty, was .88; the average latent class probabilities for the most likely class membership ranged from .94 to 1. We then performed discriminant analysis to understand better what characterized the three patterns and differences among the three groups. Students in Group 1 (n = 24, 15%) were again characterized by low rubric scores on all of the dimensions. Students in Group 2 (n = 81, 52%) were characterized by high rubric scores on all the dimensions. Finally, students in Group 3 (n = 51, 32%) were characterized by relatively high scores on literary comprehension and argument dimensions (Claims, Reasoning, Evidence) but relatively low scores on writing (Organization, Syntax). Table 19 shows the standardized canonical discriminant function coefficients, again indicating that function 1 relates to literary comprehension and argumentation, with special weight on Reasoning and function 2 weights most heavily on writing organization dimensions. Figure 5 depicts group profile differences plotted against the two functions.

Table 19. Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
PostClaimBest0	.031	.456
PostFunctionBest0	.089	-.163
PostEvidenceBest0	.244	-.322
PostReasoningBest0	.762	-.409
PostSymbolismBest0	.222	.103
PostCofABest0	-.060	.147
PostOrgBoth0	.319	.374
PostSyntaxBoth0	.289	.521
PostSynthesisBoth0	.112	.022

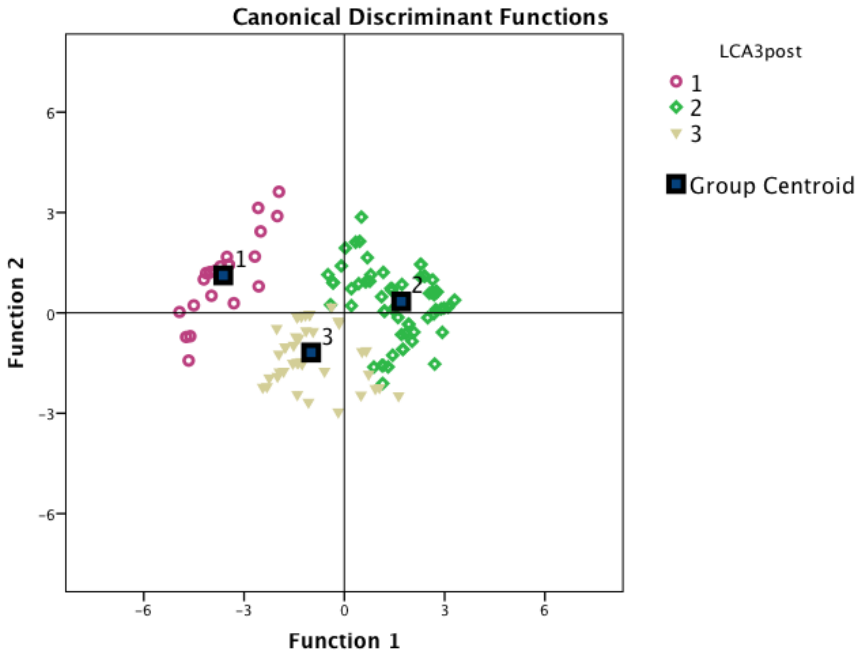


Figure 5. Results of the Post-test discriminant function analysis depicting class profile differences in relation to two discriminant functions, which represent the level of comprehension and writing (x-axis) and the difference between writing and comprehension (y-axis).

Figure 5 in conjunction with the coefficients in Table 19 indicates that Groups 1 and 2 continue to be characterized as low (group 1) or high (group 2) on all dimensions. However, group 3

suggests an even larger discrepancy than at pretest between relatively stronger performance on literary comprehension and argumentation dimensions than on writing dimensions.

Finally we looked at changes in classification group from pre to post interventions, shown in Table 20. Overall, 53% of the students changed group classification. Of the 33 in Group 1 at pre, only 21% remained there at post, with 33% now classified in the high group and 45% in Group 2, having improved their scores on literary comprehension and argument dimensions. Students who began in the high group tended to remain there (67%) although 12% declined across the board, perhaps reflecting a lack of investment in completing the assessment. The other 21% retained high levels on literary comprehension and argumentation but performed less well at post on their writing dimensions, and were classified in Group 3. Similar to what happened with Group 3, 44% of the pre Group 3 students remained in this group with 17% showing lower performance on the literary comprehension and argumentation dimensions but 39% showing retaining high performance on these dimensions and scoring higher on the writing dimension, thereby being classified in Group 2.

Table 20. Pretest group by Posttest group

Pre Test Group	Frequency at PreTest (n)	Post Test Group		
		Group 1	Group 2	Group 3
Group 1 (low on all)	33	7	11	15
Group 2 (high on all)	69	8	46	15
Group 3 (literary-high; writing -low)	54	9	24	21
Frequency at Post Test (n)		24	81	57

Final Remarks

It has been the aim of this technical report to explicate the iterative design of our assessments for argumentation in literature. Through this iterative process, we have developed a system for scoring literary essays that taps into deep-level reasoning. Further, our rubric demonstrates a high degree of inter-rater reliability, which is crucial when scoring relatively subjective disciplines such as literature. These assessments are sensitive to both developmental differences and pre-post changes, making them useful tools for both researchers and educators.

(See Appendix 3 for Literature EBA Assessment)

Pre-reading Task

Circle the number that shows how much you know about the following:

1. **Birmingham Children's Crusade**

I do not know anything				I know a lot
1	2	3	4	5

2. **The history of the United States**

I do not know anything				I know a lot
1	2	3	4	5

3. **The history of Europe**

I do not know anything				I know a lot
1	2	3	4	5

4. **Civil rights**

I do not know anything				I know a lot
1	2	3	4	5

5. **Malcolm X**

I do not know anything				I know a lot
1	2	3	4	5

6. **Racial segregation in the 1960's**

I do not know anything				I know a lot
1	2	3	4	5

Answer the following question. Put down as many different ideas as you can. You can make a list instead of writing in paragraphs.

What were the main reasons for the Birmingham Children's Crusade of 1963?

Task

One purpose of reading in history is to understand the causes of historic events; in other words, we read to understand how and why things happened. To do this, we often need to gather information from multiple sources.

Today you will be reading about **the events leading up to children being arrested in Birmingham, Alabama**. You will have to piece together important information across multiple sources to construct a good understanding of how and why this happened. No single source will provide the answer. You are the one making connections across sources and coming up with an explanation.

> **Your task is to read this set of sources to help you understand and explain what brought about the Birmingham Children's Crusade and why such an event happened.**
>
> **While reading, it is important to show your thinking by making notes in the margins or on the texts.**
>
> **You will be asked to answer questions and use specific information from the sources to support your conclusions and ideas.**

You can read the sources in any order you wish, but you should read the sheet called "Background: Segregation in the U.S." first because it gives general information on the topic.

Background: Segregation in the U.S.

The United States has had a long history of racial tension. At the end of the Civil War in 1865, slavery was declared illegal. However, segregation – separation of people based on race – was still legal. Under segregation, many states made it illegal for African Americans to vote, own land, use public facilities, or attend schools with white people.

Many people tried to fight segregation. In 1954, a famous U.S. Supreme Court case changed the segregation laws. The Court said it was illegal to have segregation in schools. But even with a Supreme Court order to stop segregation, many states in the South were slow to follow the new law or refused to follow it at all. Segregation did not end overnight.

Many people protested segregation in those slow-to-catch-on towns and cities. Many African American protesters were beaten or arrested for marching, picketing, or simply sitting down at a "whites only" restaurant lunch counter. Even children joined the protests and were arrested.

Civil Rights Activism in Birmingham

In the 1960s, protestors continued to fight for Civil Rights across much of the United States. Protestors in Birmingham, Alabama faced the most violent resistance. Birmingham was a large city with a long history of racial violence. Leaders in Birmingham refused to follow the Supreme Court ruling against segregation. Birmingham city officials even closed all of their 38 public playgrounds, 8 swimming pools, and 4 golf courses so they would not have to integrate them. Rev. Dr. Martin Luther King, Jr. called Birmingham "the most thoroughly segregated city in the United States."

In May of 1963, Dr. King and Reverend James Bevel organized mass marches by African-American elementary, high school, and college students in Birmingham. The marches were known as the Birmingham Children's Crusade. Birmingham police turned police dogs and high-pressure water hoses on the student marchers, and the police made more than 2,400 arrests between May 2nd and May 7th.

The use of children and the brutal treatment by police gained national attention. Although many supported Dr. King and Bevel's decision to use children in the protests, Malcolm X, another civil rights leader, argued "real men don't put their children on the firing line."

Author: *Steven Mintz.*
Title: *Bombingham: Civil Rights Activism in Birmingham.*
Published: *www.digitalhistory.uh.edu/disp_textbook.cfm?smtID=2&psid=3326.*
Date: *2013.*
About the Author: *Steven Mintz is the Director of Columbia University Graduate School of Arts and Sciences*

Negro Girls Define 'Freedom' From Cell in Birmingham Jail

Anita is a 12-year old Negro girl, one of the thousand or more juveniles arrested here on Monday for rioting against racial segregation.

She is at the Jefferson County Detention Home, along with about 110 other girls, none older than 13. The youngest person arrested in the riots was a 7-year-old girl.

Anita spoke with a reporter in a locked room she shares with 20 other Negro girls, all arrested for parading without a permit, a violation of Section 1159 of the General City Code.

"Do you want to go home?" the girls were asked.

"Yes!" they chorused.

"But I'd do it again," Anita Woods said. "I'll keep on marching until I get freedom."

"What freedom?" a reporter asked.

"It's equal rights," another girl shouted. "I want to go to any school and any store downtown and sit in the movies." She giggled. "And sit around in a cafeteria."

Dale G. Oltman, the chief probation officer of the juvenile court, a soft-spoken former Nebraskan, said the detention home normally accommodated 62 youngsters. With the present population of 110 or so, it was like having guests drop in unexpectedly, he said. The home is doing its best to feed and bed the children, he said, "but of course, it's not like home."

Author: *Philip Benjamin.*
Title: *Negro Girls Define 'Freedom' From Cell in Birmingham Jail.*
Published: *Special edition of The New York Times.*
Date: *May 9, 1963.*
About the Author: *Philip Benjamin was a journalist for the New York Times, an award winning daily newspaper publishing since 1851.*

Children of Birmingham and the Civil-Rights Movement

In May 1963 Martin Luther King Jr. and Rev. James Bevel launched the Children's Crusade and began a march on Birmingham. By May 7, more than 3,000 black young people were marching on the city.

On May 2, 1963, Raymond Goolsby joined thousands of students who left their classrooms and gathered at the 16th Street Baptist Church in Birmingham. It was there where they spilled out in groups of 50 to march downtown. "My group was the first of 50 to march," says Goolsby. "Our job was to decoy the police. We got arrested about a block and a half from 16th Street."

Gwen Gamble had just been released from jail and didn't want to go back. Shortly before the crusade, the teenager had been arrested for participating in a lunch-counter sit-in and jailed for five days. She and her two sisters were trained by the movement to be recruiters for the Children's Crusade. On the first day of the march, they went to several schools and gave students the cue to leave school and join the march. They then made their way to 16th Street Baptist Church.

"We left the church with our picket signs and our walking shoes," says Gamble. "Some of us even had on our rain coats because we knew that we were going to be hosed down by the water hoses."

Author: *Lottie L. Joiner.*
Title: *Children of Birmingham and the Civil-Rights Movement.*
Published: *The Crisis Magazine.*
Date: *May 2, 2013.*
About the Author: *Lottie Joiner is the senior editor of The Crisis Magazine the official magazine of the National Association for the Advancement of Colored People (NAACP).*

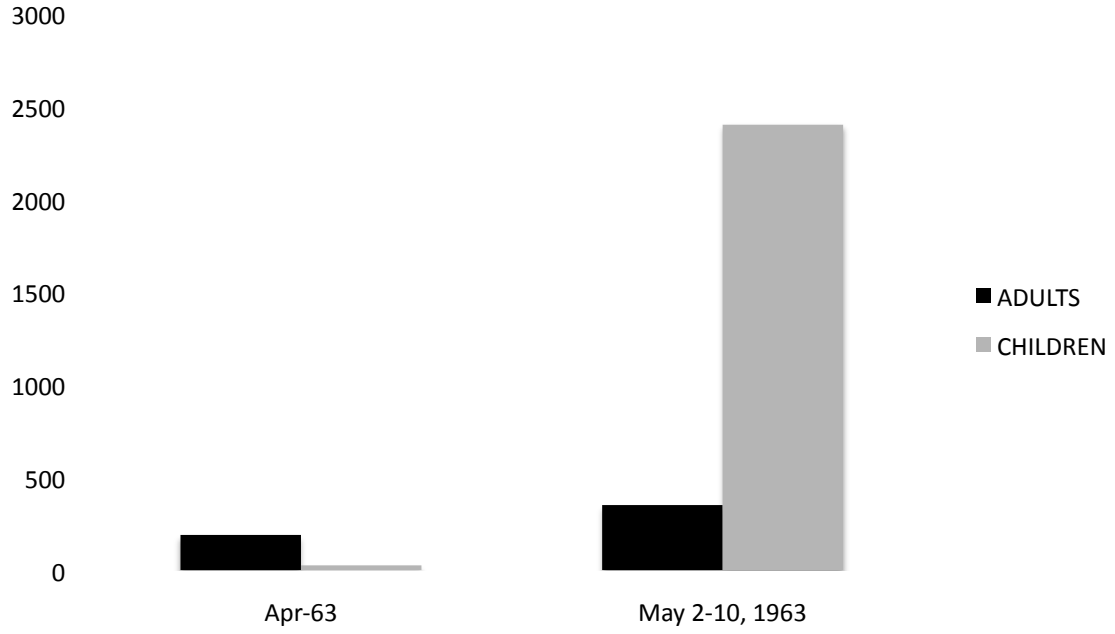
Martin Luther King Jr.'s Project C

On Wednesday, April 3, 1963, Martin Luther King Jr. announced a new plan, called Project C; the "C" stood for Confrontation. Black adults and college students would confront Birmingham city officials again and again, hoping to be arrested. King reasoned that if enough protesters were arrested, they would fill the jails and make it impossible for police to arrest more. Police would not be able to enforce local segregation laws.

Also, volunteers would go to other places where segregation was enforced, and protest in different ways. For example, they would go to segregated lunch counters (restaurants) and order food. They would kneel and pray in segregated parks, and picket stores that maintained segregated dressing rooms.

From the beginning of Project C, the number of adults arrested was discouragingly low.

Number of adults and children arrested for protesting in Birmingham, Alabama



Author: *Author Cynthia Levinson.*
Title: *Martin Luther King Jr.'s Project C.*
Published: *We've Got a Job: The 1963 Birmingham Children's March.*
Date: *February 1st, 2012.*
About the Author: *Cynthia Levinson is an author of non-fiction about children.*

Interview with James Bevel

INTERVIEWER: WHAT MADE BIRMINGHAM A CITY TO FOCUS ON?

"In '63 in Birmingham, most adults felt that segregation was permanent, that it was just that way, that it was a permanent system. The power of the city, the power of the state, the power of the Congress; they saw it all as an impossible situation. And so most of the adults felt that nothing like that could change probably…people didn't think that there was a force or a power within the country strong enough to offset something as entrenched and as reinforced as segregation."

INTERVIEWER: WHY DID YOU DECIDE TO HAVE CHILDREN PROTEST?

"Because of the intense suppression and the conditioning of the adults, it was necessary to use children because children had not been indoctrinated into that kind of violence and suppression. So they could come on the situation with a fresh approach. We wanted to get the black community in Birmingham involved, and the way you get people involved is to get their children involved."

INTERVIEWER: WHAT MADE THE CHILDREN BETTER PROTESTERS?

"Most adults have bills to pay, house payments, rents, car payments, and utility bills. But the young people, while they can think at the same level, do not have all those responsibilities. So a boy from high school, he get the same effect in terms of being in jail because of protesting, in terms of putting pressure on the city as his father. There's no loss of money in the family because the father's still on the job."

Author: *Blackside, Inc.*
Title: *Interview with James Bevel*
Published: *Eyes on the Prize: America's Civil Rights Years (1954-1965). Washington University Libraries, Film and Media Archive, Henry Hampton Co.*
Date: *November 13, 1985.*
About the Author: *Blackside, Inc. was founded by Henry Hampton, and has created several documentaries about social justice.*

King's Letter From Jail After Getting Arrested for Protesting in Birmingham

My Dear Fellow Clergymen:

I am in Birmingham because injustice is here. [...]

There can be no arguing the fact that racial injustice engulfs this community. Birmingham is probably the most thoroughly segregated city in the United States. Its ugly record of brutality is widely known. Negroes have experienced grossly unjust treatment in the courts. There have been more unsolved bombings of Negro homes and churches in Birmingham than in any other city in the nation. These are the hard, brutal facts of the case. On the basis of these conditions, Negro leaders sought to negotiate with the city leaders. But the city leaders consistently refused to engage in good faith negotiation. [...]

Yours for the cause of Peace and Brotherhood, Martin Luther King, Jr.

Author: *Dr. Martin Luther King, Jr.*
Title: *King's Letter From Jail After Getting Arrested for Protesting in Birmingham.*
Published: *American Friends Service Committee*
Date: *May, 1963.*
About Source: Martin Luther King Jr. was a leader of the Civil Rights Movement in the 1950s and 1960s. King wrote this letter after he was arrested for organizing protests in Birmingham, Alabama.

Writing task

Your task is to use this set of documents to write an essay explaining **what brought about the Birmingham Children's Crusade and why such an event happened.** Be sure to use specific information from the documents to support your conclusions and ideas.

Short answer

Using the documents, answer the following short response questions. Some answers may require you to repeat information that you have used to answer other questions.

1. Explain how the actions of the Birmingham city leaders helped lead to the confrontation that resulted in adults being arrested?

2. Rev. Bevel and the girls who were arrested all wanted kids to be arrested in the fight against segregation. How did their reasons differ?

Understanding Sides

Based on the sources provided, indicate each person's or group's side on the Children's protest (put an "X" in the column 1 if you thought the person/group was for the protest, put an "X" in the column 2 if you thought the person/group was against the protest, or use the column 3 if you aren't sure).

Person/Group	1. For the protest	2. Against the protest	3. Not Sure
Malcolm X			
Anita			
African American adults in Birmingham			
Birmingham city officials			

Very short answer

Briefly answer the following questions. You may need to mention information that you have written in prior answer.

1. What was Project C?

2. Using the document titled "Martin Luther King Jr.'s Project C", what happened to the number of arrests in the spring of 1963?

3. What actions did the kids in Birmingham take in May of 1963?

Essay Evaluation Task

Below are essays written by students like you explaining what brought about the Birmingham Children's Crusade and why such an event happened. Read them and answer the questions that follow.

Essay 1

Segregation was legal for a lot of U.S. history. Segregation refers to separating groups based on their races. In the past, U.S. segregation included lots of things, such as separate bathrooms, water fountains, movie theaters, and schools for white and black people. They even had signs to let people know which things they were allowed to use. In 1954, it was decided that public schools shouldn't be segregated. This meant that it was illegal to have separae facilities, even if people thought that they were equal. The states were being required to integrate schools. The change to integration wasn't immediate, and this made a lot of people mad. Many people protested segregation in those places that didn't immediately become integrated. Even children joined the protests. In one case, children were even arrested for protesting. This was called the Birmingham Children's Crusade because it happened in Birmingham, Alabama. Kids as young as 7 years old were arrested.

What did the student do well in the essay? _____

_____

_____

What advice would you give the student for improving this essay? _____

_____

_____

Essay 2

Segregation is the separation of people based on race. This used to be legal in the U.S., but in 1954 it was decided that it was illegal to segregate schools. Not every town and city became desegregated at once. Some places, like Birmingham, Alabama, were slower to do so because they had a lot of money invested in maintaining separate facilities. Dr. Martin Luther King, Jr. went to Birmingham to help organize peaceful protests of segregation. MLK even got arrested and put in jail for doing this. MLK and Reverend Bevel organized something called Project C. The idea was to get people arrested for protesting in order to fill up the jails so that there was no time or room for arresting people for violating the segregation laws in Birmingham. It was hard to get adults to be willing to do this. MLK and Bevel decided it might be easier to have children do this. Children don't have the same responsibilities as adults so it might be easier. Some parents forced their children to protest. So, children protested in what is now called the Birmingham Children's Crusade, and many were arrested.

What did the student do well in the essay? _____

_____

_____

What advice would you give the student for improving this essay? _____

_____

_____

Pre-reading Task

Circle the number that shows how much you know about the following:

1. **The Scopes trial**

I do not know anything				I know a lot
1	2	3	4	5

2. **The history of the United States**

I do not know anything				I know a lot
1	2	3	4	5

3. **The history of Europe**

I do not know anything				I know a lot
1	2	3	4	5

4. **Immigration from Eastern Europe**

I do not know anything				I know a lot
1	2	3	4	5

5. **The theory of Evolution**

I do not know anything				I know a lot
1	2	3	4	5

6. **Social changes in the 1920's**

I do not know anything				I know a lot
1	2	3	4	5

Answer the following question. Put down as many different ideas as you can. You can make a list instead of writing in paragraphs.

What were the main reasons for the Scopes trial of 1925?

Task

One purpose of reading in history is to understand the causes of historic events; in other words, we read to understand how and why things happened. To do this, we often need to gather information from multiple sources.

Today you will be reading about **a famous trial called the Scopes trial. In the Scopes trial, a school teacher was taken to court and found guilty of teaching the theory of evolution**. You will have to piece together important information across multiple sources to construct a good understanding of how and why this happened. No single source will provide the answer. You are the one making connections across sources and coming up with an explanation.

> **Your task is to read this set of sources to help you understand and explain what brought about the Scopes trial and why such an event happened.**
>
> **While reading, it is important to show your thinking by making notes in the margins or on the texts.**
>
> **You will be asked to answer questions and use specific information from the sources to support your conclusions and ideas.**

You can read the sources in any order you wish, but you should read the sheet called "Background: Theory of Evolution" first because it gives general information on the topic.

Background: Theory of Evolution

In the 1800s, scientists began to consider possibilities for how living creatures came to be the way they are. A scientist named Charles Darwin put forward a theory of evolution that suggested several things: First, he said, all living things didn't always look the way they look today. They actually came from the same organism millions of years ago. Second, every group of living things, like lions, fish, or flowers, had so many offspring that some of them just couldn't survive. This means some died off quickly, and some survived longer. The ones that survived did so because they were stronger or more well-adapted to their environments. For example, baby lions who had stronger jaws and sharper claws would survive, while weaker lions would die off. Third, the stronger or better-adapted species would then pass on those strong characteristics to the next generation, and the process would repeat itself over time. By the late 1800s, many had accepted this theory of evolution.

During the summer of 1925, John Scopes, a football coach and substitute high school teacher in Dayton, Tennessee, was arrested for teaching the theory of evolution. A trial was held, and Scopes was found guilty and fined $100.

The Scopes Arrest

In 1925, John Scopes, a substitute teacher, was arrested in Dayton, Tennessee for teaching Darwin's theory of evolution in a high school biology class. Scopes had broken a new law called the 1925 Butler Act. The Butler Act stated, "it shall be unlawful for any public school teacher to teach any theory that denies the story of the Divine Creation of man as taught in the Bible. It shall be illegal to teach that man has descended from a lower order of animals."

A group of businessmen from Dayton called the Dayton "Boosters" were interested in improving, or "boosting," the town's economy. The Boosters had read an advertisement in the local newspaper offering to provide free lawyers to defend any teacher who was arrested for teaching evolution. The Boosters thought a high-profile trial would be a good opportunity to draw business to Dayton. They asked Scopes to admit he taught evolution. The Boosters were able to convince Scopes to admit he taught evolution.

Author: *Dr. Trisha Kennedy*
Title: *The United States and Its People: American Life Changes.*
Published: *Addison-Wesley Textbook Publishers.*
Date: *2003.*
About the Author: Dr. Trisha Kennedy is a professor of American history.

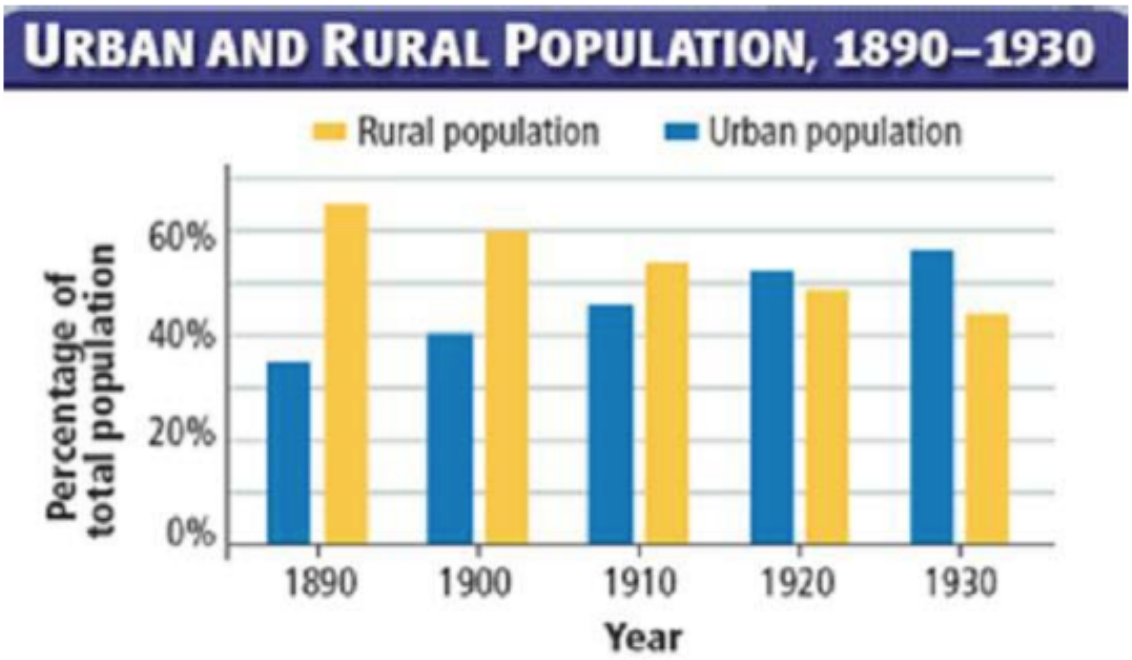
American Federation of Teachers' Statement

The American Federation of Teachers is deeply concerned about the effect of the Tennessee anti-evolution law on the development of learning through teaching in this country.

As teachers we especially fear the effect of the present wave of intolerance in education on the task of providing the schools with well-informed and tolerant teachers. Without freedom in the intellectual life, and without the inspiration of uncensored discovery and discussion, there could ultimately be no scholarship, no schools at all and no education. The minds that now seek an outlet in education would be driven off into other fields, if indeed, they could find a reason for existence anywhere.

Author: *The American Federation of Teachers.*
Title: *Resolution adopted by American Federation of Teachers*
Published: www.aft.org/about/resolutions.
Date: *July 18, 1925*
About the Author: The AFT is a union that was founded in 1916, and included teachers, school workers, and state employees.

Urban (city) and Rural (farmland) Population Chart



Author: *Bureau of the Census.*
Title: Urban (city) and Rural (farmland) Population Chart.
Published: *Historical Statistics of the United States, Colonial Times to 1970, Washington, DC.*
Date: *1975.*
About the Author: The United States Bureau of the Census was created to collect and report data on the American people and economy.

ACLU Ad: Planned Assault On State Law On Evolution

American Civil Liberties Union to File a "Test Case":

"We are looking for a Tennessee teacher who is willing to accept our services in testing this law in the courts," Professor Skinner states. "Our lawyers think a friendly test can be arranged without costing a teacher his or her job. Distinguished attorneys have volunteered their services. All we need now is a willing client. By this test we hope to render a real service to freedom of teaching throughout the country, for we do not believe the law will be sustained."

"The Civil Liberties union is also making efforts against other recent 'restrictions' on teaching. Freedom of teaching, according to a survey by the union's committee on academic freedom, has been more interfered with by law in the past six months than at any time in American history. These laws prohibit the teaching of evolution, require compulsory reading of the Bible and forbid the employment of radical teachers (teachers who have extreme political views) or pacifist teachers (teachers who believe in non-violence)."

Author: American Civil Liberties Union (ACLU)
Title: ACLU Ad**:** Plan Assault on State Law on Evolution.
Published: Chattanooga Daily Times a local newspaper from Chattanooga, Tennessee.
Date: May 4, 1925.
About the Author: The ACLU was created in 1920 in order to protect the freedom of speech of anti-war protesters.

Mrs. Sparks Letter to the Editor

Editor of the Nashville Tennessean:

At the time the bill prohibiting the teaching of evolution in our public schools was passed by our legislature I could not see why mothers in greater numbers were not conveying their appreciation to the members for this act of safeguarding their children from one of the destructive forces which . . . will destroy our civilization. I for one felt grateful for their standing for the right against all criticism. And grateful, too, that we have a Christian man for governor who will defend the Word of God against this so-called science. . . .

Mrs. Jesse Sparks
Pope, Tennessee

Author: *Mrs. Jesse Sparks.*
Title: Mrs. Sparks Letter to the Editor.
Published: Nashville Tennessean.
Date: *July 3, 1925.*
About the Author: Mrs. Jesse Sparks lived in a town called Pope in Tennessee, and worked as a volunteer in the local church.

American Life Changes

Although the start of the 1920s was marked by an economic slowdown, by the mid-1920s the economy was booming. One segment of the economy, however, did not share in the good times. Farming took a hard hit in the post-World War I years, as wartime demand for food dropped off.

As industry grew in the cities, the earnings of American workers also rose. More families could afford to send their children to school instead of sending them to the textile mills and other factories. As a result, high school and college enrollment increased. Likewise, by the 1920s, many states had passed laws requiring young people to attend school. These laws helped force children out of the workplace and into the classroom. Requiring children to attend school was also a way to teach immigrants about American life.

Life in the cities was far different than life in America's small towns. Cities offered museums and art, plays and sporting events, nightclubs and movie houses. City dwellers also tolerated drinking, gambling and casual dating. And, their neighborhoods were filled with people from many cultures.

The Ku Klux Klan grew dramatically in 1920s rural America. Most new members were workers, farmers, and small business owners. They saw their own status declining while the size and cultural influence of urban America were increasing. They believed joining the Klan could help them protect their anti-urban beliefs. Like much of rural America, they supported mandatory schooling in hopes that rural belief systems would be maintained through teaching.

Author: Dr. Benjamin Helt
Title: American Life Changes.
Published: The Americans.
Date: 2009.
About the Author: Dr. Benjamin Helt is a professor of American History.

Writing task

Your task is to use this set of documents to write an essay explaining **what brought about the Scopes Trial and why such an event happened**. Be sure to use specific information from the documents to support your conclusions and ideas.

Short Answer

Using the documents, answer the following short response questions. Some answers may require you to repeat information that you have used to answer other questions.

1. In the 1920's more people lived in cities. Explain how this increase in people living in cities lead to a fight over what was taught in schools.

2. About what do Mrs. Jesse Sparks and the American Federation of Teachers disagree?

Understanding Sides

Based on the sources provided, show each person's or group's side on the Butler Act, which made the teaching of evolution in schools illegal. Put an "X" in column 1 if you thought the person/group was for the Butler Act. Put an "X" in column 2 if you thought the person/group was against the Butler Act. Put an "X" in column 3 if you aren't sure

Person/Group	1. For the Butler Act	2. Against the Butler Act	3. Not Sure
American Federation of Teachers			
Mrs. Jesse Sparks			
The Dayton Boosters			
John Scopes			

Very short answer

Briefly answer the following questions. You may need to mention information that you have written in prior answer.

1. How did the ACLU get a teacher to teach evolution?

2. Based on the sources, how did America change around the time of the Scopes Trial?

3. Why did states require students to attend public schools?

4. From the list below, choose two opposite views in the Scopes Trial debate, and explain how they were opposed to each other about the Scopes Trial.

urban, ACLU, religious, schools, rural, science, military

_____

_____

_____

_____

Essay Evaluation Task

Below are essays written by students like you explaining what brought about the Scopes Trial and why such an event happened. Read them and answer the questions that follow.

Essay 1

In 1925, the Butler Act was passed. This was a law stating that it was illegal to teach evolution in public schools. Evolution is a theory that different living things have evolved from a common ancestor. Charles Darwin came up with this theory, and he said that we have differences in our characteristics because of differences in our environments. The environment changes characteristics over several generations. This means that people evolved from some ape-like creature that was less like a human. A lot of religious people are against the teaching of evolution. They believe in a single Creator who made all living species. John Scopes was a teacher who taught evolution in his science class. He was mainly a football coach, but he also did substitute teaching. After he admitted to teaching evolution, he was arrested and went to court. He was found guilty and given a fine of $100.

What did the student do well in the essay? _____

_____

_____

What advice would you give the student for improving this essay? _____

_____

_____

Essay 2

Life in the U.S. in the 1920s was quickly changing. More people were starting to live in the cities, more people worked in industry rather than agriculture, and children were being required to go to school. At this time, there were more cars around, so people started traveling longer distances. These changes seemed like a threat to people who had strong religious values. They were able to influence the government into passing the Butler Act making it illegal to teach evolution in schools. They thought this was good now that all kids had to go to school. Many people were against this law though. There were a lot of protests. The ACLU put out an ad asking for a teacher to admit to teaching evolution. It was hard to find a teacher to do this. A group of businessmen called the Dayton Boosters wanted to increase business in their town so they found a teacher to do this. They thought a trial in their town would be good for business. John Scopes was the teacher who came forward admitting that he taught evolution. He was arrested, and he went to court. He was found guilty and never taught again.

What did the student do well in the essay? _____

_____

_____

What advice would you give the student for improving this essay? _____

_____

_____

Name: _____

Date: _____

School: _____

Period: _____

Pre-reading Task

Circle the number that shows how much you know about the following:

	I do not know anything				I know a lot
1. Coral bleaching	1	2	3	4	5
2. Life science	1	2	3	4	5
3. Earth science	1	2	3	4	5
4. Plant cell function	1	2	3	4	5
5. Oceanography	1	2	3	4	5
6. The sun	1	2	3	4	5

Answer the following question. Put down as many different ideas as you can. You can make a list instead of writing in paragraphs.

What causes coral bleaching?

Name: _____

Date: _____

School: _____

Period: _____

Task

One purpose of reading in science is to understand the causes of scientific phenomena; in other words, we read to understand how and why things happen. To do this, we often need to gather information from multiple sources.

Today you will be reading about what causes "coral bleaching". Coral, which lives in the ocean, can be many different colors, but sometimes it loses its color and turns white. You will have to piece together important information across multiple sources to construct a good explanation of how and why this happens. No single source will provide all of the important parts of the explanation. Instead, you are the one making connections across sources and coming up with an explanation.

Your task is to read the following set of sources to help you understand and explain what leads to differences in the rates of coral bleaching.

While reading, it is important to show your thinking by making notes in the margins or on the texts.

You will be asked to answer questions and use specific information from the sources to support your ideas and conclusions.

You can read the sources in any order you wish, but you should read the sheet called "Background: What is 'Coral Bleaching?'" first, because it gives general information on the topic.

Background: What is "Coral Bleaching?"

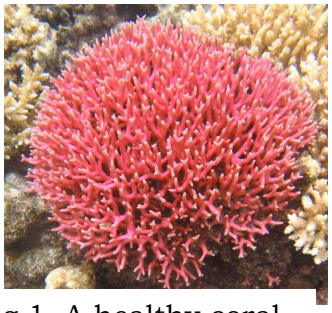


Fig 1. A healthy coral

Marine biologists have discovered many different types of corals living in the oceans. These invertebrate animals live together in colonies and tend to stay in one place. The different types of algae living within the coral polyps give the corals their varying colors (see Figure 1). Some coral have been "bleached" a plain white. Coral bleaching is a phenomenon in which coral loses its color. Events leading to coral bleaching are a serious problem with a serious impact on the world's coral reefs.

As can be seen in Figure 2, coral bleaching is most noticeable in the Pacific Ocean. This ocean covers about 1/3 of the surface of the entire globe, and contains double the amount of water found in the Atlantic Ocean.

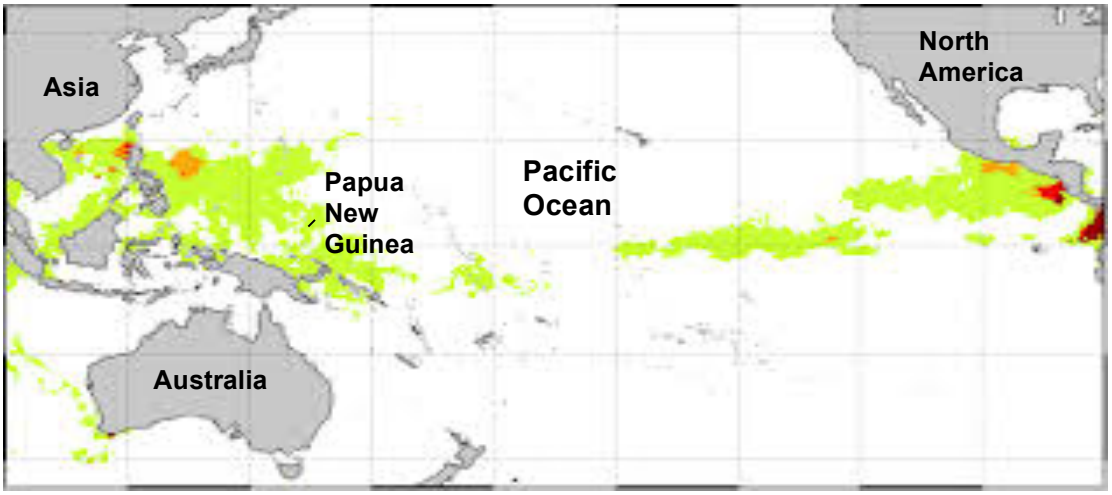


Fig 2. This map of the Pacific Ocean shows coastal regions most affected by coral bleaching (darker areas near land masses).

Shifting Trade Winds

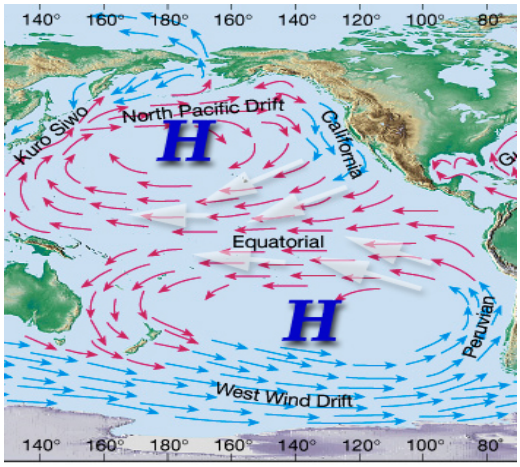


Fig 1. Ordinary Pacific trade winds

Trade winds were originally used to guide boat navigation and were later recognized as being important for merchant trading in the 1800's. Now we know that trade winds have a much more important purpose for earth's oceans. Ordinarily, Pacific trade winds travel from east to west (shown in Figure 1). The winds drag warm surface waters westward. These warm waters pool just north of Australia, near Papua New Guinea. Meanwhile, the deeper, colder waters in the eastern Pacific rise to the surface. This phenomenon is known as upwelling. Upwelling in the eastern Pacific causes surface waters to be colder than those in the western Pacific.

Every few years, trade winds are altered. The trade winds weaken or reverse direction completely to blow from west to east. Under these circumstances, warm surface waters are dragged eastward towards South America.

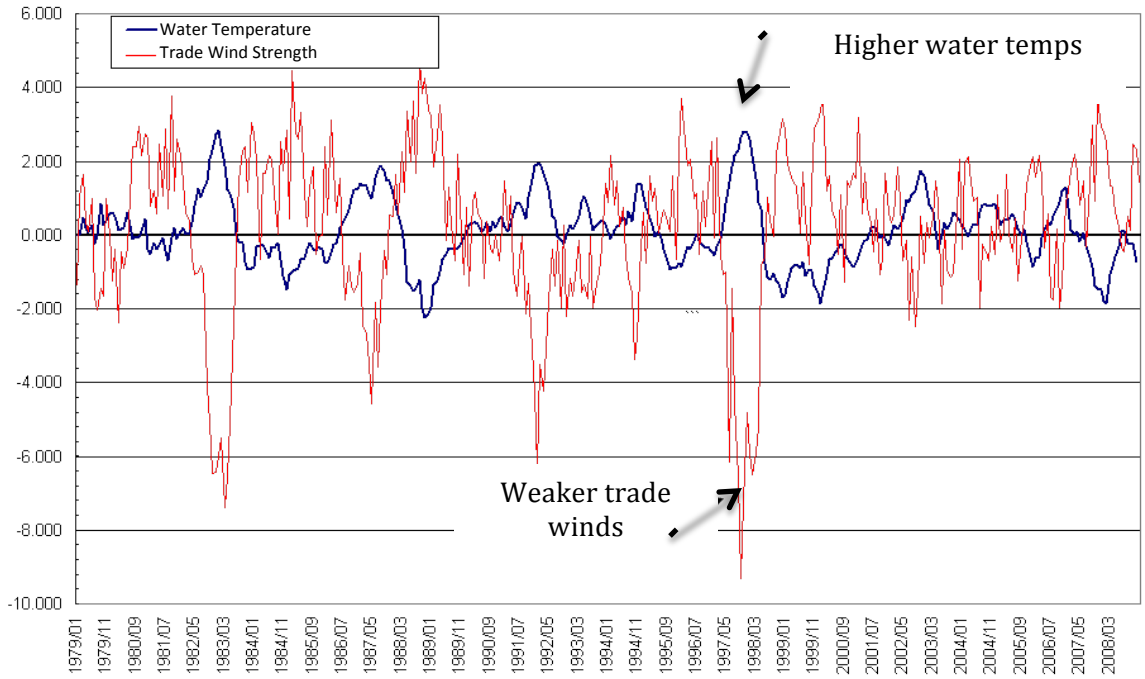
This movement causes the central and eastern Pacific regions to swell. Consequently, sea levels rise anywhere from inches to as much as a foot. In the western Pacific, the sea level declines as the warm surface water increasingly flows towards the east. This movement of warm water eastward causes major shifts in the world's climate. This impact is especially true along equatorial regions. Normal water temperature in these regions is 70-80°F; however, when surface waters shift eastward, ocean water temperatures increase by 3°F to 5°F. In some places, they can increase over 10°F above the customary range.

Source: Adapted from *http://earthobservatory.nasa.gov/Features/ElNino/elnino.php* Retrieved June, 2013

Trade Winds and Coral Bleaching

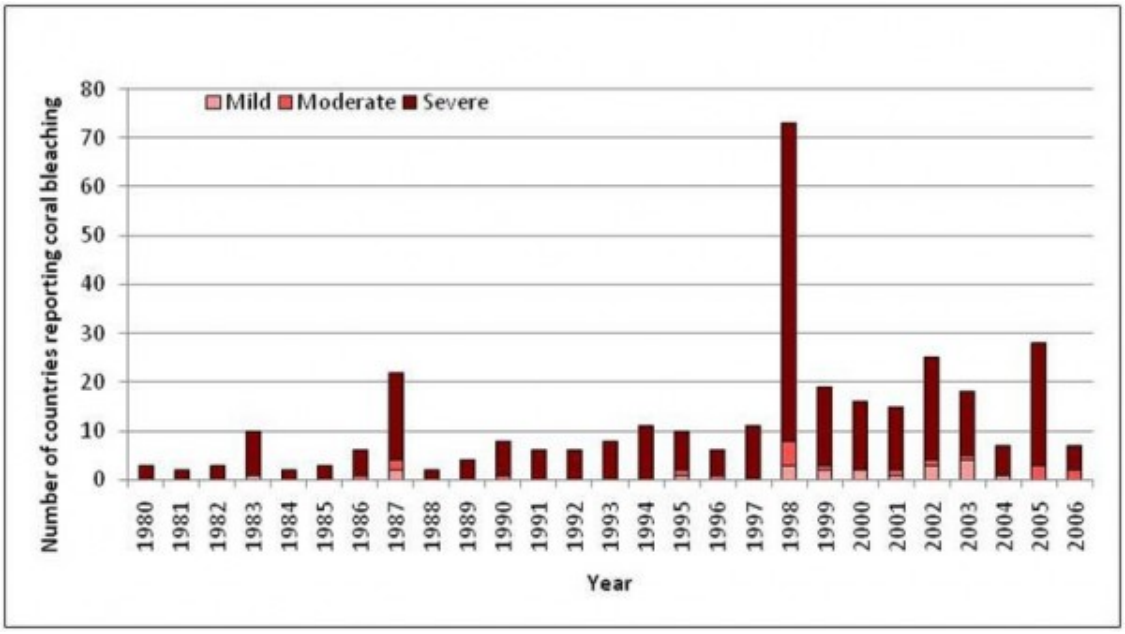
Water Temperature – Change in degrees
Trade Wind Strength – Meters per second (speed)

Water Temperature Versus Trade Winds



Source: http://wattsupwiththat.com/2009/02/17/the-trade-winds-drive-the-enso/ Retrieved June, 2013

Coral Bleaching Reports per Year



Source: Adapted from Reefbase.org Retrieved June, 2013

Coral & Photosynthesis

Coral is often mistaken for a rock or plant. However, it is really made up of tiny animals called polyps. A coral polyp has a saclike body and a mouth surrounded by stinging tentacles. Coral polyps range in size from tiny (the size of a pinhead) up to a foot in diameter. Corals live in colonies that are made up of many polyps. Polyps use limestone from seawater to build a hard skeleton. This hard skeleton protects the delicate body of the polyp. Over time, the hard skeletons build up and form coral reefs.

Coral reefs are only found in clear, shallow, tropical waters. This is partly because algae, called zooxanthellae, live in the coral tissues and need light for the process of photosynthesis. During photosynthesis, zooxanthellae use energy from sunlight to combine Carbon Dioxide (CO_2) in the ocean with water (H_2O). This combination results in the formation of sugars and Oxygen (O_2). As this is occurring, carbon is passed from the algae to the coral in the form of glucose (sugar). As long as there is enough sunlight and a normal amount of CO2, the algae will undergo photosynthesis. That's important because the chemicals that the algae make provide the coral with the energy it needs to survive. If the amount of CO2 is lower than normal, it can be a problem for the coral. Scientists estimate that corals receive 50% to 95% of their energy from this process. This process is very important, but it is also very sensitive to changes in the environment.

Reef building corals need water temperatures between 70-85°F because the process of photosynthesis is sensitive to changes in water temperature. As water temperature increases, the amount of Carbon Dioxide (CO_2) in water decreases. Changes in the amount of CO_2 threaten the delicate balance required to keep corals healthy.

This balance can also be threatened by extreme storms. Corals are sensitive to ocean salinity (how salty the water is). Salinity can decrease in shallow coastal regions during storms that cause excessive rainfall, like hurricanes or tropical storms. These storms increase the amount of fresh water in the ocean, which causes the salinity to drop. This drop in salinity upsets the balance necessary for coral health.

Source: Adapted from *http://water.epa.gov/type/oceb/habitat/coral_index.cfm*
Retrieved June, 2013

Coral and Zooxanthallae

Coral and zooxanthellae algae have a symbiotic relationship. This means that they both benefit from their relationship with each other. In fact, most zooxanthellae cannot live outside of the coral animal's body. This is because there are not enough nutrients in the tropical ocean to allow photosynthesis. The coral provides a protected nutrient-rich environment for the algae. This relationship also benefits the coral. The zooxanthellae pass some of the food they make from the sun's energy to the coral. The algae also provide corals with their colors. However, certain environmental stressors can threaten this relationship.

The algae rely on the coral to stay healthy, but corals can sustain physical damage from some destructive practices. Examples of this include blast fishing and tourists who drop anchors or walk on reefs. However, the most dangerous threats to the health of corals are those that are caused by changes in their environment that put stress on coral. This is because increased stress on corals from the environment forces the coral to eject their food-producing algae

Figure 1. Bleached coral (left) and healthy coral (right)

Coral bleaching (Figure 1 shows bleached and healthy coral) is one example of how environmental stressors can negatively affect the balanced relationship between the coral and zooxanthellae. During bleaching, corals turn white due to the ejection or death of the zooxanthellae algae (see Figure 1). This results in a bleaching event that makes corals vulnerable to disease and starvation. While some dying reefs recover from such events, many do not. For example, a massive coral bleaching event in 1998 is considered one of the worst ever observed. This event resulted in the death of 16% of the world's coral reefs.

Source: Adapted from http://news.nationalgeographic.com/news/2006/05/warming-coral.html Retrieved June, 2013

Name: _____

Date: _____

School: _____

Period: _____

Writing task

Using this set of documents, write an essay **explaining what leads to differences in the rates of coral bleaching**.. Make sure to connect the ideas within your explanation to the differences in the rates of coral bleaching. Be sure to use specific information from the documents to support your ideas and conclusions.

Name: _____
Date: _____
School: _____
Period: _____

Multiple Choice Items

Based on the documents you read, please select the option that best fills in the blanks to answer the question: "**explain what leads to differences in the rates of coral bleaching**."

1. A decrease in the strength of trade winds causes _____, which leads to increased ocean temperatures.
   A. calmer oceans
   B. coral bleaching
   C. a decrease in photosynthesis by zooxanthellae
   D. eastward warm water movement

2. An increase in coral stress leads to _____, which causes coral bleaching.
   A. a decrease in CO_2
   B. the ejection of zooxanthellae
   C. a decrease in trade winds
   D. coral producing bleach

3. An increase in tropical storms over the Pacific Ocean causes _____, which leads to an increase in coral stress.
   A. saltwater fish to die
   B. an increase in water temperatures
   C. a decrease in CO_2 in the ocean
   D. less salt in the ocean

4. A decrease in the amount of CO_2 in the water leads to _____, which causes an increase in coral stress.
   A. death or ejection of zooxanthellae
   B. an increase in oxygen in the water
   C. a decrease in sugars produced by the algae
   D. an increase in water temperatures

5. According to the graphs in the source titled "Trade winds and coral bleaching," it can be concluded that…
   A. coral bleaching occurred around the same time as increasing trade winds.
   B. coral bleaching occurred around the same time as decreasing trade winds.
   C. rates of coral bleaching are the same regardless of trade winds.
   D. there is no relationship between coral bleaching and trade winds.

6. Increases in water temperature in the east lead to
   A. less food for the coral.
   B. decreases in coral stress.
   C. stronger trade winds.
   D. more frequent storms.

7. If there were no changes in the intensity of trade winds, coral bleaching
   A. could never happen.
   B. could be at an all time low.
   C. could happen due to other causes.
   D. could be at an all time high

8. The presence of the right amount of salt in the ocean results in
    A. more intense storms.
    B. greater coral health.
    C. less transfer of food from algae to coral.
    D. changes in water temperatures.

9. Decrease in photosynthesis increases coral stress because of
    A. a decrease in chemicals to the coral for food.
    B. a decrease in water temperatures.
    C. an increase in carbon dioxide.
    D. an increase in upwelling.

Explanation evaluation task

Below are explanations written by students like you who are explaining **what leads to differences in the rates of coral bleaching**. Read the explanations and answer the questions that follow.

Explanation 1

The trade winds usually go in a certain direction across the ocean, but if the trade winds decrease, warm waters shift eastward which leads to higher water temperatures in the east. The increase of water temperatures causes carbon dioxide to decrease in the ocean, which then affects the algae that live among the coral and give them their color. Since there is a drop in carbon dioxide, the process of photosynthesis can no longer occur properly. This causes stress for the coral, and this makes them eject the algae that live inside them. Without the algae, they look white or bleached.

Considering the essay question, what did the student do well in the explanation?

_____

_____

_____

Considering the essay question, what advice would you give the student for improving this explanation?

_____

_____

_____

Another student explanation for **what leads to differences in the rates of coral bleaching**. Read the explanation and answer the questions that follow.

Explanation 2

To understand how and why coral bleaching rates vary at different times, you have to understand that corals are animals that live in the ocean. A lot of people think they are plants or rocks, but they're really animals. This is because they are invertebrates and live in colonies. Corals have algae and small fish that live among them. Sometimes the trade winds slow down or change direction completely, and as it shows in the graph, this makes the water temperature warmer in the eastern part of the Pacific Ocean. This is a problem for the coral and they turn white, and that's why it's called coral bleaching. Coral bleaching can also happen because tropical storms and hurricanes happen over the ocean and this makes the coral turn white.

Considering the essay question, what did the student do well in the explanation?

_____

_____

_____

Considering the essay question, what advice would you give the student for improving this explanation?

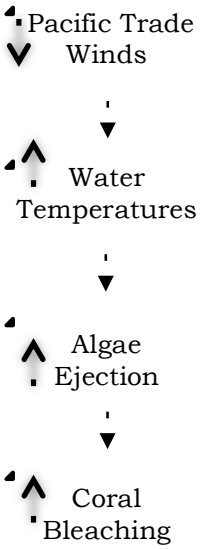
_____

_____

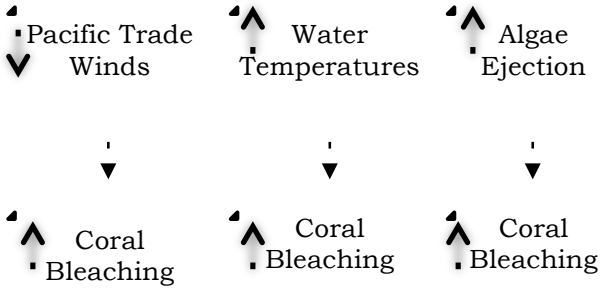
_____

Graphical model comparison

Model 1



Model 2



Above you can see two graphic models. The arrows between the boxes indicate connections between steps in the process. The arrows within the boxes indicate increases and decreases in components of the process. Which graphic model above provides the best explanation for **what leads to differences in the rates of coral bleaching**?

Circle your answer: Model 1 or Model 2

Why do you think the model you selected is better? _____

_____

_____

Name: _____

Date: _____

School: _____

Period: _____

Pre-reading Task

Circle the number that shows how much you know about the following:

	I do not know anything				I know a lot
1. Skin cancer	1	2	3	4	5
2. Life science	1	2	3	4	5
3. Earth science	1	2	3	4	5
4. The earth's coordinates	1	2	3	4	5
5. Cell reproduction	1	2	3	4	5
6. The sun	1	2	3	4	5

Answer the following question. Put down as many different ideas as you can. You can make a list instead of writing in paragraphs.

What causes skin cancer?

Name: _____

Date: _____

School: _____

Period: _____

Task

One purpose of reading in science is to understand the causes of scientific phenomena; in other words, we read to understand how and why things happen. To do this, we often need to gather information from multiple sources.

Today you will be reading about what causes some people to experience abnormal cell growth like skin cancer. You will have to piece together important information across multiple sources to construct a good explanation of how and why this happens. No single source will provide all of the important pieces of the explanation. Instead, you are the one making connections across sources and coming up with an explanation.

Your task is to read the following set of sources to help you understand and explain what leads to differences in the risk of developing skin cancer.

While reading, it is important to show your thinking by making notes in the margins or on the texts.

You will be asked to answer questions and use specific information from the sources to support your ideas and conclusions.

You can read the sources in any order you wish, but you should read the sheet called “Background: Skin Damage” first, because it gives general information on the topic.

Background: Skin Damage

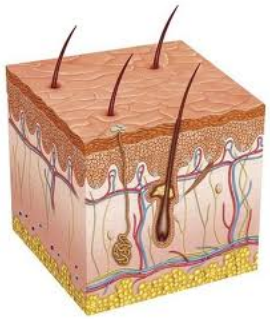


Fig 1. A section of healthy skin.

It may surprise you to learn that the skin on our bodies is our largest organ (see Figure 1). It covers every region of our bodies in order to protect our inner tissue from infection and loss of water. In addition, our skin helps regulate our body temperature. Although we take our skin for granted, there are several ways for things to go wrong with our skin.

There are numerous skin disorders, conditions, and diseases. Of these, skin cancer is among the most feared because everyone is at some risk of developing skin cancer, but some people are at a higher risk than others. Additionally, skin cancer is the most common form of cancer in the United States (Figure 2 shows a patch of skin cancer). Skin cancer is the uncontrolled growth of abnormal skin cells. The variety of skin cancer that develops depends on the type of skin cell that reproduces irregularly.

Fig 2. The dark patch to the right of this person's eye is caused by a basal cell carcinoma.

There are three main varieties of skin cancer: basal cell carcinoma, squamous cell carcinoma, and malignant melanoma. Together, basal and squamous cell carcinomas make up approximately 95 percent of skin cancers. Malignant melanoma only occurs in approximately 5 percent of skin cancer cases. However, malignant melanoma is responsible for the most deaths from skin cancer.

Checking your skin for suspicious changes can help with detecting skin cancer at its earliest stages. Early detection of skin cancer gives you the maximum chance for successful treatment.

Latitude and Direct Sunlight

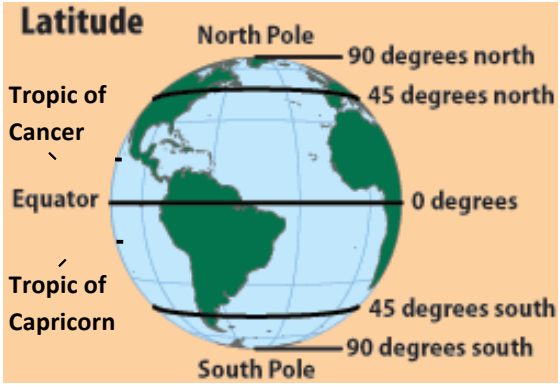


Fig. 1 Important lines of latitude

A common way to locate points on the surface of the earth is by geographic coordinates (see Figure 1). These geographic coordinates are called latitude and longitude. Latitude and Longitude are measured in degrees and represent distances from the center of the Earth. We can imagine the Earth as a sphere, with an axis around which it spins. The ends of this axis are the North and South Poles. The Equator is an imaginary line around the Earth at 0 degrees latitude. Latitude values indicate the distance between the Equator and points north or south of it.

Depending on location, amount of direct sunlight may vary a lot or a little throughout the year. As a rule of thumb, the closer you are to the Equator, the more consistent direct sunlight will be. This means people closer to the equator need to be more aware of the dangers of direct sunlight.Those who live in areas with less direct sunlight may be less concerned about any dangers, although there are still some risks associated with sun exposure wherever you live and at any time of year.

Some locations are more likely to have moderate to extreme levels of year-round direct sunlight. Examples of this include the Northern third of Australia and the Southern parts of the United States. The most year-round direct sunlight occurs between the Tropics of Cancer (23°N) and Capricorn (23°S). Due to the amount of direct sunlight in these areas, the amount of UVb radiation is also high. Generally speaking, the more direct sunlight there is, the more UVb radiation there is.

Source: Adapted from http://www.uvawareness.com/uv-info/uv-strength.php#uv_strength Retrieved June, 2013.

Skin Cancer and Latitude

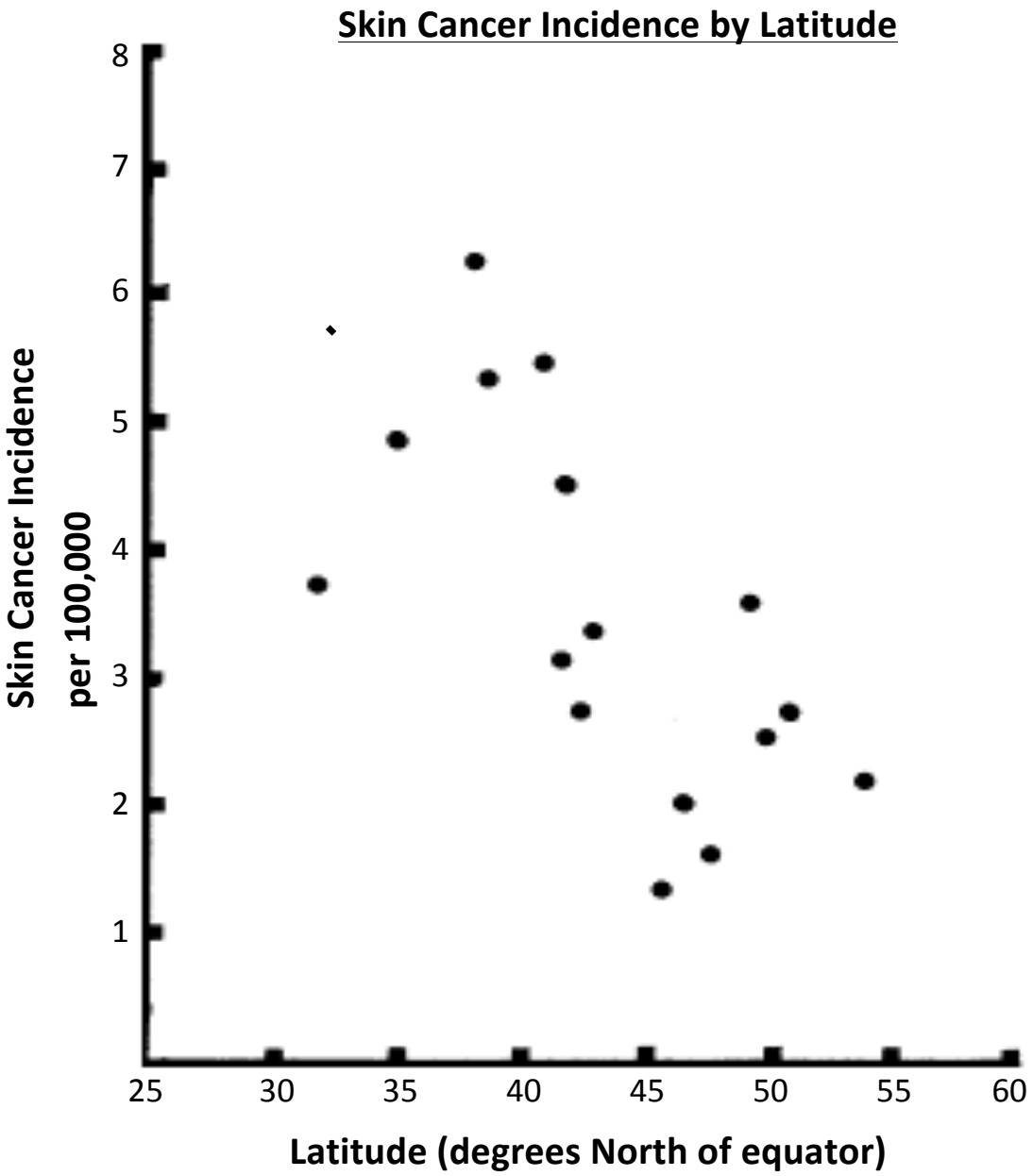


Fig. 1. –The relationship between skin cancer incidence per 100,00 population (age-standardized) and latitude among 16 North American male populations.

Source: Adapted from
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2010099/pdf/brjcancer00145-0099.pdf
Retrieved June, 2013

Your Skin Protects You

If you took off your skin and laid it flat, it would cover an area of approximately 21 square feet. The skin is by far the body's largest organ. Skin forms the barrier between what's inside us and what's outside. Surprisingly, in some places, this barrier is less than a millimeter thick. Skin protects us from many external forces. It acts as a waterproof, insulating shield. It also protects the body against extreme temperatures, damaging solar radiation, and chemicals. Skin releases antibacterial substances that assist in the prevention of infection. Additionally, it manufactures vitamin D for converting calcium into healthy bones.

Skin is composed of three layers. The outermost skin layer is the epidermis. The middle skin layer is the dermis. This layer includes collagen, elastin (this makes skin stretchy), and nerve endings. The innermost layer is subcutaneous fat. This layer contains tissue that acts as an energy source, cushion, and insulator for the body.

While skin is tough, it can't completely protect you from everything. For example, many experimental studies of plants and animals and clinical studies of humans have demonstrated harmful effects of exposure to UVb radiation. UVb radiation is a type of solar radiation. Using both human skin cells and a mouse model, researchers have found that when the skin is hit with normal amounts of UVb radiation, the rays cause small amounts of damage to genetic molecules in the skin cells. Luckily, our bodies are good at repairing this typical amount of damage; however, problems arise when there is excessive exposure to UVb radiation.

Your skin does have some defenses against solar radiation though. Melanin is a special pigment produced in the epidermis. Melanin helps to protect us from solar radiation. The advantage of having less melanin is that it is easier for the body to make vitamin D. The downside is that it decreases our chances of protection from UVb radiation.

Source: Adapted from http://science.nationalgeographic.com/science/health-and-human-body/human-body/unmasking-skin/ Retrieved June, 2013

Sunburn

Sunburn is caused by damage to genetic molecules in skin cells . Sunburn happens when the body directs blood to skin to try to repair or remove damaged cells. The damaged cells are replaced with healthy ones. This additional blood flow is the reason skin becomes red.

Sunburns frequently include painful burning sensations (see Figure 1). The severity of sensations depends on the severity of the burn. Severe sunburn is called sun poisoning. Sun poisoning can lead to infection and shock. In extreme cases, it can even cause death.

Fig.1 This is an image of a woman with a severely sunburned back

Sunburned skin cells that aren't removed can result in skin cancer. Normally, abnormal or damaged cells are cleared away and replaced with new cells, but this doesn't always happen. If a sunburn is severe enough, it becomes less likely that all damaged skin cells will be removed. Suffering just one serious sunburn in childhood or adolescence doubles the chances of developing skin cancer later in life. These chances also double if a person has five or more sunburns at any age.

Research investigating damaged skin cells is critical. More damage to genetic molecules leads to more severe sunburn. The amount of damaged genetic molecules can be used as an indicator of sunburn severity.

Source: *Adapted from* http://www.utsandiego.com/news/2012/Jul/08/cause-of-sunburns-painful-inflammation-discovered/ Retrieved June, 2013.

Name: _____

Date: _____

School: _____

Period: _____

Writing task

Using this set of documents, write an essay **explaining what leads to differences in the risk of developing skin cancer**. Make sure to connect the ideas within your explanation to the differences in the risk of developing skin cancer. Be sure to use specific information from the documents to support your ideas and conclusions.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Name: _____

Date: _____

School: _____

Period: _____

Multiple Choice Items

Based on the documents you read, please select the option that best fills in the blanks to answer the question: "**explain what leads to differences in the risk of developing skin cancer**."

1. People living near 0 degrees latitude experience _____, which leads to more UVb radiation.
   A. decreases in melanin
   B. increases in skin cancer
   C. failure to remove damaged cells
   D. more direct sunlight

2. A more intense sunburn will most likely lead to _____, which causes skin cancer.
   A. decreased UVb radiation protection from the sun
   B. decreased proportion of damaged cells being removed
   C. an increase in melanin production in the skin
   D. increases in direct sunlight exposure

3. Having skin cells that produce very little melanin results in _____, which results in increases in skin cell damage.
   A. increases in UVb radiation exposure
   B. changes in levels of direct sunlight
   C. a decrease in removal of damaged cells
   D. less protection from the sun

4. An increase in UVb radiation exposure causes _____, which leads to a more severe sunburn.
   A. decreased protection from the sun
   B. changes in removal of damaged skin cells
   C. increased skin cell damage
   D. changes in melanin production

5. Using the document titled "Skin Cancer and Latitude," it can be concluded that
   A. more skin cancer occurs the further north you are.
   B. less skin cancer occurs the farther north you are.
   C. rates of skin cancer are the same regardless of distance north.
   D. there is no relationship between skin cancer and latitude.

6. Avoiding UVb radiation
   A. can reduce the risk of sunburn.
   B. can reduce the amount of melanin in the skin.
   C. may decrease the amount of dead cells that are removed.
   D. is impossible near the equator.

7. Increased amounts of direct sunlight lead to
   A. less distance to the equator.
   B. decreases in the amount of melanin produced.
   C. decreases in the likelihood that all damaged cells will be removed.
   D. decreases in frequency and severity of sunburns.

8. If a person went his or her whole life never going between the Tropic of Cancer and the Tropic of Capricorn, skin cancer
   A. will never develop for the person.
   B. could still occur for the person.
   C. will be very likely for the person.
   D. could never be deadly for the person.

9. Higher levels of protection from UVb radiation result in
   A. decreases in sunburn severity.
   B. decreases in damaged cell removal.
   C. increases in nearness to the equator.
   D. increases in direct sunlight.

Explanation evaluation task

Below are explanations written by students like you who are explaining **what leads to differences in the risk of developing skin cancer**. Read the explanations and answer the questions that follow.

Explanation 1.

To understand what leads to differences in the risk of developing skin cancer, you have to understand that skin is our largest organ. The skin has many layers to it, so it's not surprising that there are a lot of things that can go wrong with it. Skin cancer is the most feared skin disease. There are several types of skin cancer, and there are different causes. According to the graph, we should avoid being near the equator where there is a lot of direct sunlight. This is because the sun can cause damage, and sometimes the damage is so bad that it increases the risk of skin cancer. Skin cancer rates can also be higher because you have less melanin in your skin. Fortunately, we have all kinds of ways to protect our skin, such as hats, sunscreen, and sunglasses.

Considering the essay question, what did the student do well in the explanation?

_____

_____

_____

Considering the essay question, what advice would you give the student for improving this explanation?

_____

_____

_____

Explanation evaluation task

Another student explanation for **what leads to differences in the risk of developing skin cancer**. Read the explanation and answer the questions that follow.

Explanation 2.

Skin cancer risk can differ depending on how close you are to the equator. The closer you are to the equator, the more direct sunlight exposure you will get. Direct sunlight has more UVb radiation, and this is harmful. If you live further from the equator, you'll experience less direct sunlight, which means less harmful UVb radiation. A lot of UVb radiation can increase the amount of damage to genetic molecules in skin cells causing more severe sunburns. When a sunburn is really bad, it makes it less likely that damaged skin cells will be removed, and this increases the risk of skin cancer. The fewer bad sunburns you have, the less likely it is that you'll have trouble removing damaged skin cells, and this is good.

Considering the essay question, what did the student do well in the explanation?

_____

_____

_____

Considering the essay question, what advice would you give the student for improving this explanation?

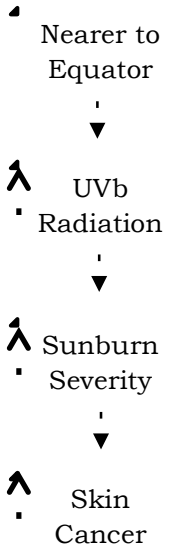
_____

_____

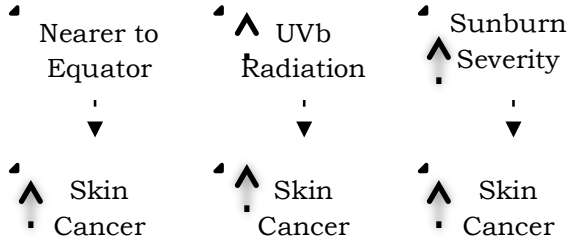
_____

Graphical model comparison

Model 1



Model 2



Above you can see two graphic models. The arrows between the boxes indicate connections between steps in the process. The arrows within the boxes indicate increases and decreases in components of the process. Which graphic model above provides the best explanation for **what leads to differences in the risk of developing skin cancer**?

Circle your answer: Model 1 or Model 2

Why do you think the model you selected is better? _____

_____

_____

Literature Task

Directions for students
Story 1

Today, you will be reading two stories. Tomorrow you will write an essay comparing and contrasting how some of the symbols in the stories help us understand the stories' characters and their worlds.

Please read each story carefully. While reading, it is important to show your thinking by taking notes in the margins or other places on the text. You can read the stories as many times as you like. When you are done reading each story, please respond to the questions.

"Eleven"

by Sandra Cisneros

What they don't understand about birthdays and what they never tell you is that when you're eleven, you're also ten, and nine, and eight, and seven, and six, and five, and four, and three, and two, and one. And when you wake up on your eleventh birthday you expect to feel eleven, but you don't. You open your eyes and everything's just like yesterday, only it's today. And you don't feel eleven at all. You feel like you're still ten. And you are -underneath the year that makes you eleven.

Like some days you might say something stupid, and that's the part of you that's still ten. Or maybe some days you might need to sit on your mama's lap because you're scared, and that's the part of you that's five. And maybe one day when you're all grown up maybe you will need to cry like if you're three, and that's okay. That's what I tell Mama when she's sad and needs to cry. Maybe she's feeling three.

Because the way you grow old is kind of like an onion or like the rings inside a tree trunk or like my little wooden dolls that fit one inside the other, each year inside the next one. That's how being eleven years old is.

You don't feel eleven. Not right away. It takes a few days, weeks even, sometimes even months before you say Eleven when they ask you. And you don't feel smart eleven, not until you're almost twelve. That's the way it is.

Only today I wish I didn't have only eleven years rattling inside me like pennies in a tin Band-Aid box. Today I wish I was one hundred and two instead of eleven because if I was one hundred and two I'd have known what to say when Mrs. Price put the red sweater on my desk. I would've known how to tell her it wasn't mine instead of just sitting there with that look on my face and nothing coming out of my mouth.

"Whose is this?" Mrs. Price says, and she holds the red sweater up in the air for all the class to see. "Whose? It's been sitting in the coatroom for a month."

"Not mine," says everybody. "Not me."

"It has to belong to somebody," Mrs. Price keeps saying, but nobody can remember. It's an ugly sweater with red plastic buttons and a collar and sleeves all stretched out like you could use it for a jump rope. It's maybe a thousand years old and even if it belonged to me I wouldn't say so.

Maybe because I'm skinny, maybe because she doesn't' like me, that stupid Sylvia Saldivar says, "I think it belongs to Rachel." An ugly sweater like that, all raggedy and old, but Mrs. Price believes her. Mrs. Price takes the sweater and puts it right on my desk, but when I open my mouth nothing comes out.

“That’s not, I don't, you’re not...Not mine," I finally say in a little voice that was maybe me when I was four.
"Of course it's yours," Mrs. Price says. "I remember you wearing it once." Because she's older and the teacher, she's right and I'm not.

Not mine, not mine, not mine, but Mrs. Price is already turning to page thirty-two, and math problem number four. I don't know why but all of a sudden I'm feeling sick inside, like the part of me that's three wants to come out of my eyes, only I squeeze them shut tight and bite down on my teeth real hard and try to remember today I am eleven, eleven. Mama is making a cake for me tonight, and when Papa comes home everybody will sing Happy birthday, happy birthday to you.

But when the sick feeling goes away and I open my eyes, the red sweater's still sitting there like a big red mountain. I move the red sweater to the corner of my desk with my ruler. I move my pencil and books and eraser as far from it as possible. I even move my chair a little to the right. Not mine, not mine, not mine.

In my head I'm thinking how long till lunchtime, how long till I can take the red sweater and throw it over the school yard fence, or even leave it hanging on a parking meter, or bunch it up into a little ball and toss it in the alley. Except when math period ends Mrs. Price says loud and in front of everybody, "Now Rachel, that's enough," because she sees I've shoved the red sweater to the tippy-tip corner of my desk and it's hanging all over the edge like a waterfall, but I don't' care.

“Rachel,” Mrs. Price says. She says it like she's getting mad. "You put that sweater on right now and no more nonsense."

"But it's not--"

"Now!" Mrs. Price says.

This is when I wish I wasn't eleven, because all the years inside of me--ten, nine, eight, seven, six, five, four, three, two and one-- are pushing at the back of my eyes when I put one arm through one sleeve of the sweater that smells like cottage cheese, and then the other arm through the other and stand there with my arms apart like if the sweater hurts me and it does, all itchy and full of germs that aren't even mine.

That's when everything I've been holding in since this morning, since when Mrs. Price put the sweater on my desk, finally lets go, and all of a sudden I'm crying in front of everybody. I wish I was invisible but I'm not. I’m eleven and it's my birthday today and I'm crying like I'm three in front of everybody. I put my head down on the desk and bury my face in my stupid clown-sweater arms. My face all hot and spit coming out of my mouth because I can't stop the little animal noises from coming out of me, until there aren't any more tears left in my eyes, and it's just my body shaking like when you have the hiccups, and my whole head hurts like when you drink milk too fast.

But the worst part is right before the bell rings for lunch. That stupid Phyllis Lopez, who is even dumber than Sylvia Saldivar, says she remembers the red sweater is hers! I take it off right away and give it to her, only Mrs. Price pretends like everything's okay.

Today I'm eleven. There's cake Mama's making for tonight, and when Papa comes home from work we'll eat it. There'll be candles and presents and everybody will sing Happy birthday, happy birthday to you, Rachel, only it's too late.

I'm eleven today. I'm eleven, ten, nine, eight, seven, six, five, four, three, two, and one, but I wish I was one hundred and two. I wish I was anything but eleven, because I want today to be far away already, far away like a runaway balloon, like a tiny o in the sky, so tiny-tiny you have to close your eyes to see it.

Name: _____
Teacher/Period: _____
Date: _____

Questions about "Eleven"

In the chart below you will see a list of 10 events. Some of the events happened in the story "Eleven." Some did *not* happen in the story. Please do the following:

1. Read all of the events.
2. Cross out all of the events that you believe **DID NOT** happen in the story. You can reread the story as much as you want when you are doing these tasks.
3. Once you have crossed out all the events that you think **DID NOT** happen, look at the remaining events.
4. Now, put those events in the order that they happened in the story. To do this, put a "1" next to the event that happened first in the story, a "2" next to the event that happened second in the story, and so on until you have numbered all the events that DID happen in the story.

Below, cross out the events you think did **NOT** happen in the story.	Below, number the remaining events in the order that they happened in the story.
Phyllis pointed out the sweater in the coatroom.	
Sylvia did not say whose sweater she thought it was.	
Rachel gave the sweater to Phyllis.	
Phyllis refused to admit it was her sweater.	
Mrs. Price tells Rachel to put the sweater on.	
Mrs. Price is doing a math lesson.	
Rachel feels sick and tries not to cry.	
Rachel put the sweater under her desk.	
Mrs. Price says it is Rachel's sweater.	
Rachel told her parents about the sweater.	

The Butterfly

James Hanley

Brother Timothy's* robes made a strange noise as he strode up and down the passage. The priest's face was red, his mouth twitched, and his fingers pulled nervously at the buttons upon his robe. One could see at a glance that he was angry. He muttered to himself, staring at a strong wooden door before him.

He simply could not understand the boy. Every time the name Cassidy came into his mind, the blood mounted to his forehead. It was the boy's silence that was the enraging thing, his silence! And what was even worse was how happy the boy seemed. Curse him for his silence and happiness. The boy must have no conscience at all.

Brother Timothy stopped and stared at the wooden door. He listened. Not a sound. The boy might be standing behind it now, maybe thinking that he would be let out. Brother Timothy laughed then. That boy would not be let out until he explained himself, until he broke his silence. How to break it down then? Yes, one must try to think.

The priest drew a key from his pocket, opened the door and went inside. The boy was sitting on the bed. He looked up at the Brother, but something in the other's glance made him hurriedly drop his eyes again.

"Well, Cassidy," said Brother Timothy, "Have you come to your senses yet?" The veins in his neck stood out. The silence galled him. "Answer me, I tell you!" he shouted. But Cassidy did not speak.

"Look at me! Yesterday you missed mass, you and this other child Byrne. Did you ask permission to miss mass? Why were you truant? Why are you so unlike the others? And this silence! I will not stand it. You have the devil in you; it's he who has trapped your tongue. But I'll break you. Do you hear me? I ask you for the last time, why did you skip mass?"

Cassidy, a boy of twelve, looked up at the Brother. His lips moved, but he made no sound.

In fury, Brother Timothy struck him across the face.

Then Cassidy said slowly, "Brother Timothy, I told you yesterday."

"You are determined then. Fine. You will remain here, and out of this room you will not go until you open that mouth of yours. And when you have explained to me you will go to confession, and you will confess." The priest strode out of the room.

---

* *Male teachers in some religious schools are called "Brother." Brother Timothy is a teacher.*

The door banged, the key turned.

In the morning when the boy woke the sun was streaming through the window. It filled him with an intense longing for the open air, to be free of this room, free of the sound of those well-known footsteps, from the sight of that face, which mirrored rage and defeat.

Cassidy reflected that he meant no harm. He had simply gone into the forest, had become absorbed in the strange life that abounded in the hedges and ditches, and had not heard the bell. And here he was stuck in this musty room for two days because he would not explain. “But I have explained,” he kept saying to himself. “I have explained.”

From his pocket he took a cardboard box pinned with air-holes. He removed the lid. Slowly a green caterpillar crept out and along his finger. Cassidy watched its slow graceful movements down his hand. He lowered his head and stared hard at it. What a lovely green it was. And one day it would turn into a beautiful butterfly. How marvelous. He stroked it gently with his finger. The sun came out, it poured through the window, filled the room, and the green caterpillar was bathed in the light.

"I think I'll call you Xavier,” he said to the insect, and smiled. For two whole days he had had it in the cardboard box. It made him happy knowing it was there, in the room with him. It made him forget Brother Timothy, forget many things. He knew he would be happy while he had the green caterpillar. If it could speak, he would explain to it why he was kept in the stuffy room by Brother Timothy. Perhaps this green caterpillar did know, perhaps it looked at him.

Cassidy heard footsteps in the corridor. A moment later the door opened and Brother Timothy stepped into the room.

"Well, Cassidy," he said. "Have you come to your senses?" But the boy appeared not to hear him. He was standing with his back to the Brother, the sunlight on his face, and he was gently laying the caterpillar on a bit of moss in the cardboard box.

"Cassidy!" roared the man behind him, and the boy turned round. “What have you there?”

"Nothing-I mean-Brother Timothy, it's a-"

"What! And this is how you spend your time. Aren't you sorry for your sin?"

"Sin? Brother Timothy, I-I mean, it's only a little caterpillar."

If silence had been poisonous, this was worse. “Is this how you think upon your conscience? Is this how you think out your explanation? Outrageous, boy! Give me that at once."

"But it's only a caterpillar, Brother Timothy, a little green one. Soon it'll be a butterfly. It's so green and soft, and it crawls up my finger just like it knew me. Please, Brother -I-

while I was sitting here all by myself it made me happy, I liked having it, I -"

"How dare you!" Brother Timothy grabbed the box and turned out the caterpillar. It fell to the floor and slowly began to crawl.

"You have no right to skip mass and you have no right to be happy or anything else. Do you hear me?" and with a quick movement of his broad foot Brother Timothy trod on the insect and crushed out its life. Cassidy looked up at the Brother. Then he burst into tears.

Name: _____
Teacher/Period: _____
Date: _____

Questions about "The Butterfly"

In the chart below you will see 10 events related to the story "The Butterfly" by James Hanley. Some of the events happened in the story. Some did not.

1. Read all of the events.
2. Cross out all of the events that you believe **DID NOT** happen in the story. You can reread the story as much as you want when you are doing these tasks.
3. Once you have crossed out all the events that you think **DID NOT** happen, look at the remaining events.
4. Now, put those events in the order that they happened in the story. To do this, put a "1" next to the event that happened first in the story, a "2" next to the event that happened second in the story, and so on until you have numbered all the events that DID happen in the story.

Below, cross out the events you think did **NOT** happen in the story.	Below, number the remaining events in the order that they happened in the story.
Cassidy said that the caterpillar made him feel a sense of responsibility.	
Brother Timothy left, locking Cassidy in the room.	
Brother Timothy discovered Cassidy's caterpillar.	
Brother Timothy asked Cassidy why he skipped mass the previous day.	
Cassidy named the caterpillar Francis.	
Brother Timothy made Cassidy put the caterpillar in the box.	
Brother Timothy found that the room with the wooden door was open.	
Brother Timothy paced up and down in the corridor.	
Brother Timothy hit Cassidy in the face.	
Cassidy went to confession.	

Name: _____
Teacher/Period: _____
Date: _____

When writing stories, authors often create symbols that help reveal ideas about the way people are and the way the world is. Symbols are words or phrases in stories that stand for more than what they seem to be. Symbols can be images, actions, objects, or characters (what they do, how they think, how they look, their names).

Today, you will be writing an essay comparing and contrasting how you think symbols in both "Eleven" and "The Butterfly" help reveal ideas about the way people are and the way the world is.

To help you write your essay, please complete this chart. It will help organize your thinking so you can write your essay. Use the stories as you fill in the chart. After 15 minutes, you will be asked to start working on the essay.

Eleven	**Butterfly**	**Compare and contrast**
What do you think are one or two of the symbols in the story "Eleven"?	What do you think are one or two of the symbols in the story "The Butterfly"?	Compare and contrast one symbol from each story.
Explain why you think they are symbols.	Explain why you think they are symbols.	Compare and contrast the reasons you think they are symbols.
In what ways does each symbol show us something important about the main character and their world? Be sure to explain why.	In what ways does each symbol show us something important about the main character and their world? Be sure to explain why.	Compare and contrast how the symbols might show something important about the main characters and their worlds.

Name: _____
Date: _____
Teacher/Period: _____

Now it is time to write your essay.

When writing stories, authors often create symbols that help reveal ideas about the way people are and the way the world is. Symbols are words or phrases in stories that stand for more than what they seem to be. Symbols can be images, actions, objects, or characters (what they do, how they think, how they look, their names).

In your essay, compare and contrast how you think symbols in "Eleven" and "The Butterfly" help reveal ideas about the way people are and the way the world is, or help you understand the characters and their worlds. Use the information you wrote in the chart and the stories to help you write your essay.

Project READi – Literature Group
Unit: Symbolism and Coming of Age
Rubric for Intervention Assessments and Cognitive Studies
Revised Feb. 20, 2012

Apply each rubric category to each story except for the Synthesis category. That category deals with cross-story connections.

Claims
0 – description or summary, restates the question

0.5 – claim/s are provided, but are not accurate in terms of being supportable by the stories

1 – at least one accurate claim is provided

2 – more than one accurate claim provided

3- more than one accurate claim provided; claims support or are connected to one another.

4. more than one accurate claim provided; claims support or are connected to one another and includes nested claims or counterclaims.

Functions of Claims
0 – summarizes the story or stories; or re-states the question without further description of character, world, or effects of language

1 – describes the internal state of a character or characters; or describes the social world of the text or texts

2 – examines how the language of the text functions to explain a character's internal state (i.e. how the language helps the reader to understand the character or characters) and/or social world of the text

NOTE: Claims in the "2" category must address how the language of the text or choices of the text function to create effects. Claims that simply repeat the language used to describe a character's internal state or the social world of the text do not count as claims about functions of language. For example, if a student wrote, "The phrase 'I wanted to be 102' shows how sad Rachel was," the student would score a 1 because they are not focusing specifically on language choices. If the student wrote, "The author uses the age of 102 to show how sad Rachel is, because 102 is so old and seems almost impossible to reach," the student would

score a 2, because they are talking about the functions of the language. Claims about symbols and their effects will most likely be scored as “2.”

Evidence
0 - no evidence provided (either because student did not support claims with evidence or because student had no claims)

0.5 – only inaccurate evidence provided (inaccurate = evidence contradicts the events of texts or reveals a misreading of the text)

1 - one piece of accurate evidence provided in attempt to support at least one claim (that is, evidence does not contradict events of text, is not a misreading of literal text). May include personal or real world knowledge, but is also text-based

2 - more than one piece of accurate evidence provided in attempt to support claim/s. May include personal or real world knowledge, but is also text-based

3 – more than one piece of accurate evidence provided in attempt to support claims. Includes pieces of evidence that belong to same pattern (e.g. “We see Rachel refer to her age three times: when she says she feels three, when she thinks about being eleven, and when she wishes she was one hundred and two”)

4 – comprehensive evidence that identifies and weighs competing textual evidence (i.e., counterlclaims) (e.g. “Although Rachel is portrayed as young, as when she cries after being mistreated, she is also portrayed as older and wiser, as when she recognizes how empty she feels, like ‘pennies in a band-aid box’”)

Reasoning .
0 – no discussion of why evidence supports the claims

1.1 – claims and evidence are logically connected; it is clear why evidence supports claims, but reader must make inference. Student does not make connection explicit

1.2 - claims and evidence are logically connected; it is clear why evidence supports claims, and student has made connection EXPLICIT (e.g. through warrant or backing).

2.1– claims and evidence are supported by some appeal to generalized or cultural norms (e.g. “pink is often associated with femininity”) implicit but reader must make inference. Student does not make connection explicit

2.2 - claims and evidence are supported by some appeal to cultural norms (e.g. “pink is often associated with femininity”), and student has made connection EXPLICIT (e.g. through warrant or backing).

2.3 – claims and evidence are supported by some appeal to literary norms (e.g. “repetition often signals the importance of an image”), but reader must make inference. Student does not make connection explicit

2.4 - – claims and evidence are supported by some appeal to literary norms (e.g. “repetition often signals the importance of an image”), and student has made connection EXPLICIT (e.g. through warrant or backing).

Organization of Ideas
0 - the response has no clear organization;

1 - the response has some organizational framework (e.g., response first talks about text A and then text B, or moves chronologically through text/s).

2 - well organized; connections among either paragraphs or sections of paragraphs are made explicit (e.g. with phrases like “also,” “but,” “in contrast,” etc.).

Language and Syntax
0 - the language and syntax are unclear and difficult to follow throughout most of essay

1 – language and syntax are mostly clear with some lapses

2 - language and syntax are very clear

Symbolism
0 - student makes no attempt to attribute symbolic significance; student does not identify a target in the text that is symbolic

1 - identifies a symbol (e.g. the sweater in “Eleven”) but does not discuss its effects

1.1 identifies symbol and discusses effects not supported by the text (e.g. “the sweater in ‘Eleven’ shows how happy the character is”)

2 - student identifies at least one target in the text that is symbolic; student constructs an abstract proposition in interpreting the target (i.e. image, event, object, state of mind, action, etc.) in the text or texts. The interpretation remains local to the text (e.g. “the butterfly represents the freedom that the boy wants”)

2.1 - student identifies a target in the text that is symbolic; student constructs an abstract proposition in interpreting the target that includes but also goes beyond the immediate world of the text (e.g. “the butterfly represents innocence and freedom of expression that can be crushed by oppressive institutions”)

Coming of Age

0 – student does not discuss the development or evolution of the character or characters in terms of some criteria for coming of age

1 - student does discuss the development or evolution of the character or characters in terms of some criteria of coming of age

2 - student does discuss the development or evolution of the character or characters in terms of some criteria of coming of age AND connects what the character has learned or how the character has changed to the symbolism in the text

NOTE: Discussions about coming of age must go beyond simply identifying a trait of a character (e.g. “Myop seems innocent”) and must discuss some kind of development or realization related to coming of age (e.g. “Myop seems innocent at first, but then loses that innocence when…”).

Synthesis

0 - no connection between stories; stories are analyzed or discussed separately

1 - surface connection (e.g., Main character is a girl in both)

2.1 - explicitly compares or contrasts claims about characters or social worlds across stories (e.g. “Both characters are oppressed”)

2.2 - explicitly compares or contrasts claims about symbols across stories (e.g. “Both stories use symbols of colors to express different things). Supporting evidence may not be in place.

3.1 - explicitly compares or contrasts claims (or claims plus evidence) along with evidence or reasoning about the comparison/contrast (e.g. “Both characters are oppressed. Rachel feels oppressed by her teacher. For example, her teacher forces her to wear a sweater that isn’t hers. Cassidy is oppressed by the priest. For example, the priest hits him when Cassidy won’t explain himself”)

3.2 -explicitly compares or contrasts symbols with evidence or reasoning about the comparison/contrast e.g. “Both stories use symbols of colors to express different things. ‘Eleven’ uses the color red to show sadness. For example, everytime the author shows the sweater, she mentions that it is red, and then

shows Rachel to be sad. ‘The Butterfly’ uses the color green to show life. The caterpillar is said to be ‘green and soft.’ Green usually represents life”)

References

Blaum, D., Britt, M.A., Griffin, T., & Wiley, J. (2014, August). *Thinking about Solutions: The Effect of Policy Questions and Answers on Learning About Causes From Science Texts*. Poster presented at the 24th Annual Meeting of the Society for Text and Discourse, Chicago, IL.

Britt, M.A., Rouet, J.-F., Georgi, M.C., & Perfetti, C.A. (1994). Learning from history texts: From causal analysis to argument analysis. In G. Leinhart, I. Beck, & K. Stainton (Eds.) *Teaching and learning in history*, (pp. 47-84). Hillsdale, NJ: Lawrence Erlbaum Associates.

Britt, M.A., Wiemer-Hasting, P., Larson, A., & Perfetti, C.A. (2004). Automated feedback on source citation in essay writing. *International Journal of Artificial Intelligence in Education*, 14, 359–374.

Goldman, S.R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C. D., Shanahan, C., & Project READI (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework of core processes and constructs. *Educational Psychologist, 51*, 219-246.

Goldman, S.R., McCarthy, K., & Burkett, C. (2015). Interpretive inferences in literature. In E. O'Brien, A. Cook, & R. Lorch (Eds.), *Inferences during reading,* (pp.386-415). Boston, MA: Cambridge University Press.

Graves, B., & Frederiksen, C. H. (1996). A cognitive study of literary expertise. In R. J. Kruez & M. S. MacNealy (Eds.), *Empirical approaches to literature and aesthetics* (pp. 397–418). Norwood, NJ: Ablex.

Griffin, T.D., Wiley, J., Britt, M.A. & Salas, C. (2012). The role of CLEAR thinking in learning science from multiple-document inquiry tasks. *International Electronic Journal of Elementary Education*, 5, 63-78.

Hillocks, G., & Ludlow, L. (1984). A taxonomy of skills in reading and interpreting fiction. *American Educational Research Journal*, 21, 7–24.

Hughes, S., Hastings, P., Britt, M. A., Wallace, P., & Blaum, D. (2015). Machine learning for holistic evaluation of scientific essays. *Proceedings of Artificial Intelligence in Education,* (pp.165-175).

Jaeger, A., Griffin, T.D., Britt, M.A., & Wiley, J. (2015, July). *Making Connections: Improving Student Learning about Climate Change*. Paper presented at the 25th Annual Meeting of the Society for Text and Discourse, Minneapolis, MN.

Langer, J. A. (2011). *Envisioning knowledge: Building literacy in the academic disciplines*. New York, NY: Teachers College Press.

Lee, C. D. (2007). *Culture, literacy and learning: Taking bloom in the midst of the whirlwind*. New York, NY: Teachers College Press.

Lee, C. D. (2011). Education and the study of literature. *Scientific Study of Literature*, 1, 49–58.

Lee, C. D., & Goldman, S. R. (2015). Assessing literary reasoning: Text and task complexities. *Theory into Practice*, 54, 213–227.

Lee, C.D., Goldman, S.R., Levine, S., Magliano, J. (2016). Epistemic cognition in literary reasoning. In Greene, J.A., Sandoval, W.A., & Bråten, I. (Eds.), *Handbook of Epistemic Cognition*, (pp. 165-183). New York: Routledge.

Lemke, J. L. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J.R. Martin & R. Veel (Eds.), *Reading science* (pp. 87-113). London: Routledge.

Peskin, J. (1998). Constructing meaning when reading poetry: An expert novice study. *Cognition and Instruction*, 16, 235–263.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34(2), 344-355.

Shanahan, C., Britt, A., Cribb, G., Goldmna, S., Hale, G., Heppeler, J., Hoard, J., Lawless, K., Manderino, M., Popp, J. S., Puklin, D., Radinsky, J., & Sosa, T. (2013, April). *Enriching history instruction for urban youth through evidence---based argument instruction modules and collaborative design research*. In A framework for conceptualizing reading for understanding: Evidence--based argumentation in history, science, and literature. Symposium conducted at the American Educational Research Association Annual Meeting. San Francisco, CA.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory; Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50.

Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavioral Research Methods, Instruments, & Computers, 31*, 137-149.

Steffens, B., Britt, M.A., & Manderino, M.L. (2014, August). *Know Before You Judge: Tasks That Promote Learning About History.* Poster presented at the 24th Annual Meeting of the Society for Text and Discourse, Chicago, IL.

Wallace, P., Blaum, D., Rupp, K., & Britt, M.A. (2014, August). *Task Model Enhancement as a Means for Improving Representations of Scientific Explanations*. Poster presented at the 24th Annual Meeting of the Society for Text and Discourse, Chicago, IL.

Wiley, J., Britt, M.A., Griffin, T.D., Steffens, B., & Project READi (2012, April). *Approaching Reading for Understanding from Multiple Sources in History and Science: Initial Studies.* Symposium paper presented at the 2012 AERA Annual Meeting, Vancouver, BC.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, 46, 1060–1106.

Wiley, J., Steffens, B., Britt, M.A. & Griffin, T. D., (2014). Writing to learn from multiple-source inquiry activities in history. In G. Rijlaarsdam (Series Ed.) and P. Klein, P. Boscolo, C. Gelati, & L. Kilpatrick (Volume Eds.), *Studies in writing, Writing as a learning activity* (pp. 120-148). Emerald Group Publishing Limited.

Wiley, J., & Voss, J.F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91,* 301-311.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73.