Developing Assessments of Students' Science Reading Processes and Explanatory Models to Measure Learning Outcomes from Instruction with Text-Based Investigation Modules

Project READI Technical Report #18

Willard Brown, Cynthia Greenleaf, Stacy Marple,
Frances Reade, Irisa Charney-Sirott, Ursula Sexton,
Gina Hale, Mo-Lin Ko, Katie James,
and Candace Burkett

Citation for this Report: Brown, W., Greenleaf, C., Marple, S., Reade, F., Charney-Sirott, I., Sexton, U., Hale, G., Ko, M., James, K. & Burkett, C. (2016). *Developing assessments of students' science reading processes and explanatory models to measure learning outcomes from instruction with text-based investigation modules.* READI Technical Report #18. Retrieved from: www.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 materials in the READI modules were developed based on enacted instruction and are intended as case examples of the READI approach to deep and meaningful disciplinary literacy and learning.

Introduction

In order to inform the iterative design process of text-based investigation modules, as well as to gauge the promise of this approach to science instruction, the READI science design team developed pre-post assessments related to specific science investigation modules. The assessment model is designed to generate data about the impact of the text-based investigation modules on student literacy and science learning. Therefore the assessments' tasks parallel the science reading-argumentation tasks in the READI MRSA and Water science modules but are reduced in scope and limited to individual work.

The assessments draw on previous work to develop Integrated Learning Assessments that was carried out by Greenleaf and collaborators at the UCLA CRESST Center for a study of the impact of Reading Apprenticeship professional development on science teaching (Silver, et al., 2011). These Integrated Learning Assessments (ILAs) were designed to enable an in- depth investigation of whether and how the integration of literacy instruction with high school biology coursework benefits student learning in both subject areas. The ILA model followed the advice of prominent national commissions that assessment should be based in knowledge of student cognition and developed to directly inform instruction and learning (NRC, 2001) and drew on research showing the power of formative assessment in enhancing student learning (Black & Wiliam, 2009; Stiggins, 2005).

In developing the ILAs for the earlier study, we applied an expanded view of assessment validity, as advanced by the CRESST criteria (Linn, Baker, & Dunbar, 1991; Herman, 1998) and the Standards for Educational and Psychological Measurement (1999). Consistent with Knowing What Students Know (NRC, 2001), the measurement approach started with a conception of student learning and the constructs to be measured– in this case those dealing with biology content and literacy skills, specified them in terms of both expected content and levels of cognitive demand, and created and validated reusable assessment models addressing content and literacy domains simultaneously. The ILA approach combined assessment strategies developed individually by CRESST and the Strategic Literacy Initiative at WestEd (Baker, 1994; Chung, O'Neil, & Herl, 1999; Baker, Aschbacher, Niemi, & Sato, 2005; Schoenbach, Greenleaf & Murphy, 2012).

In particular, the ILA, and its derivatives for the READI Project, drew on the Curriculum Embedded Reading Assessment (CERA) developed by the Strategic Literacy Initiative at WestEd to provide formative tools for assessing student reading processes and comprehension. The CERA was based on models of reading comprehension derived from expert/novice studies and built on the methodology of a think aloud protocol for eliciting the mental processing accompanying complex tasks (Ericsson & Simon, 1980, 1993; Paris & Flukes, 2005; Veenman, Van Hout-Wolters, & Afflerbach, 2006). The CERA assessment was designed to include both measures of students' efficacy within the curriculum in which Reading Apprenticeship was embedded and measures of students' ability to apply productive reading processes to novel texts and situations (Gallagher & Pearson, 1989). Measures of metacognitive awareness and the strategic deployment of intellectual resources are key targets of the intervention and therefore prominent aspects of the CERA (Baker & Brown, 1984; Flavell, 1979). Similarly, engagement – students' ability to focus on meaning, avoid distraction, actively interact with the text, and take an active and vital disposition toward the reading task is central to students' continued progress in reading ability (Wigfield & Guthrie, 2000).

In the CERA, measurements of reading comprehension, metacognitive dispositions and strategic processing, knowledge of discipline-based texts and textual practices, engagement with text and task, and stamina are elicited through a performance task. Students are asked to read a complex, instructional text and Talk to the Text as they do so, leaving their authentic thoughts and comments in the margins. After reading, students are asked to respond to metacognitive prompts about their reading processes and to summarize the text. Rubrics guide teachers in making sense of the annotations and student responses. The ILA extended this approach to multiple sources and a written explanation task (Baker, 1994; Chung, O'Neil, & Herl, 1999; Baker, Aschbacher, Niemi, & Sato, 2005), since science is about the construction of theories that explain how the world operates, explanation and argumentation are at the heart of science learning (Boulter & Gilbert, 1995; Duschl & Osborne, 2002; Erduran, Simon & Osborne, 2004; Pontecorvo, 1987; Schwarz, Neuman, Gil & Ilya, 2003). Within school-based science writing, explanation tasks are a dominant genre of student writing (Martin and Miller, 1988; Prain & Hand, 1996), and also lend themselves easily to on-demand assessment conditions.

When the ILA was used as part of a randomized, controlled trial of Reading Apprenticeship professional development, we found that students who annotated outperformed students who did not on measures of reading comprehension. In addition, we found annotations that were indicative of discipline-specific reading strategies. These included connecting to prior biology knowledge, questioning scientific methods, attending to and evaluating evidence, analyzing graphs and diagrams, and considering science implications beyond the scope of the text. For more detail on the ILA and its use in various studies, see Silver, et al., 2011.

Like the Integrated Learning Assessments, the pre/post assessments we developed for Project READI modules are comprised of interrelated science reading and explanation tasks. These assessments ask students to annotate the multiple texts presented on a phenomenon, as well as carry out culminating tasks. In the case of the Integrated Learning Assessments, these tasks included multiple choice comprehension questions and a culminating essay intended to demonstrate how students have synthesized the information in the texts. The READI science assessments differ in some respects, due to the learning goals of interest. The READI pre-post assessments targeted the Student Learning Goals defined by the project. Thus, students are asked to carry out close reading with annotation of a set of science texts presenting information about a scientific phenomenon, to develop an explanatory model for the science phenomena synthesized from information presented in the text set, and, for the assessment accompanying the high school MRSA investigation, to compose a recommendation for potential courses of action drawing on their own model and grounded in evidence from the text set. Different from our prior Integrated Learning Assessments, because we wanted students to learn to read visual representations, including explanatory models, we asked students to compose an explanatory model rather than an explanatory essay.

READI Assessment Design Criteria

Tasks. Several design criteria influenced the development of the assessment task specification.

- **Close reading:** To assess development of student's close reading process as well as gain insight into modeling and argumentation while reading, the task asked students to make notes or annotate the texts while reading.
- **Synthesizing science information:** The task description explicitly asked students to draw on information from the multiple texts in the text sets.
- **Science in the public interest:** The task introduction advanced a claim that the science phenomenon was demonstrably in the public interest.
- **Multiple science representations:** The task explicitly asked students to use both words and visuals in their constructed response.
- **Science Modeling:** The task asked students to create a model of a science phenomenon.
- **Scaffold for science model:** The task offered scaffolds about the purpose and composition of science models – that they are explanatory and they may include words and visuals.

- **Duration:** The one-hour duration focused the task on synthesizing information into an evidence-based model which is largely an explanation task.

Texts. These design criteria also influenced the selection of texts for the pre and post assessments for the READI Water and MRSA science modules.

- **Close Reading:** We selected texts for 6th grade and 10th grade science students for which students would likely have background knowledge but need to read the texts to respond to the task.
- **Synthesizing science information:** To assess students' ability to synthesis information across texts to construct domain knowledge, we included texts that offered both distinct and corroborating information about the science phenomena. No single text provided sufficient information to respond to the modeling and recommendation task.
- **Science in the public interest:** To parallel the READI science modules we selected topics with demonstrable public interest and included texts that provided evidence that the science phenomena warranted public interest.
- **Multiple science representations:** To assess students' reading of multiple science representations we included at least one visual text in the text set.
- **Science Modeling:** to assess students' modeling while reading we selected phenomena with multiple elements and relationships. Also, we selected texts with enough information about the system to form an evidenced based mental model.
- **Scaffold for science model:** We were concerned that students' unfamiliarity with the concept of a science model would lead to underperformances on the pretest. Therefore, we included at least one text that is an example of a science model in the text set.
- **Duration:** The one-hour duration limited the scope of the text set to no more than four brief texts to facilitate students' completion of the assessment task in a one-hour class period.

Two additional criteria contributed to the selection of texts.

- **Overarching themes:** we selected topics and texts for the pre and post assessments that included overarching themes/big ideas/enduring understandings present in the text-based investigation modules.
- **Authenticity:** To assess students' reading of texts they might encounter, we used unmodified texts, except that we excerpted texts from larger documents.

Analogous topics. To meet these criteria, we looked for analogous topics to link to specific text-based investigation modules to serve as assessment tasks. We developed an assessment on Malaria for the high school MRSA investigation, one on pesticide resistant lice for the middle school MRSA module, and an analogous assessment topic of the Carbon Cycle for the Water investigation module. The resulting assessments are appended (see Appendix A – Assessments).

As in any complex design process with multiple and competing criteria, each pre and post assessment text set and task met our various design criteria incompletely. Limiting the duration of the assessment to a single class period limited the extent of constructed responses we could require. For example, in the READI text-based investigation modules, students both created explanatory models and engaged in argument about the quality of the models. In the pre and post assessments on the other hand, we asked students to construct the science model--the claim about how the phenomena works--but we did not ask them to back up their models with an explicit argument for the model. An exception to this was the pre and post assessment on Malaria, for which high school students were asked to engage in argumentation in a third task – to make a recommendation for reducing deaths from Malaria and Africa and account for why the recommended course of action might work.

Synthesizing science information across multiple texts and scaffolding students' understanding of science modeling also presented challenges to simultaneous satisfaction of the design criteria. Students who had not yet been asked to develop an explanatory model (the pre-investigation assessment) could not be expected to know what a model is or might look like, or even to see how visual modalities could support their representations of how a phenomenon worked. By including a model related to the phenomenon, we offered such scaffolding. However, including a partial model of the system also situated in a single text much of the information required for developing an explanatory model, thereby working against the need to synthesis across texts (see the life cycle model for the Plasmodium bacteria included in the Malaria text set, Appendix A).

Synthesizing science information and science modeling also presented a challenge if we were to stay within a single lesson period for assessment duration. We expect that readers will need to read substantial amounts about a topic to develop a reasonably grounded explanatory model. Synthesizing science information relies on reading multiple, information rich texts. To meet the duration criterion, a practical expediency was to limit our impact on curriculum pacing, and yet offer sufficient information about the science system. Accordingly, we selected concise texts focused on explaining the phenomena. Yet concise texts limit the spectrum of close reading work necessary to distinguish relevant and irrelevant information because most of the information we included in these texts was relevant to the explanation/modeling task.

The 2012 Malaria pre/post assessment begins with a page that presents the annotation and modeling tasks. Students are asked to engage in close reading and annotation of a set of four science texts presenting information about a scientific phenomenon and to develop an explanatory model for the phenomenon synthesized from the information presented in the text set. The assessment prompts offer scaffolds about the purpose and composition of science models – that they are explanatory and they may include words and visuals. The next pages present the four texts, one text per page. Texts 1, 2 and 4 contain verbal text only. Text 3 is a visual model of the life cycle of the bacteria that causes Malaria. After the texts are presented, the model prompt is reiterated and space for students' responses are provided. Based on observations made about the 2012 version, a revised version of the Malaria pre/post assessment was created for 2013 which included more visual text and more information on the ecology of Malaria. Texts 2 and text 3 (the visual model) from the previous (2012) version were combined into a single composed text. An additional visual text (a Malaria map) was appended onto text 4. The tasks and overall structure of the assessment prompts did not change.

Administration of Pre/Post Assessments

During the 2011-2012 and 2012-2013 school years, we administered topic-related assessments directly before and after students engaged with the text-based investigations (Water in middle school, MRSA in middle and high school). The assessments were subsequently de-identified and scanned for analysis. These data sets are described **in the Table 1.**

Table 1: Pre/post assessment data sets

Data Set # and Date	Assessment Administered	% Module completion	Grade	Classes	Number of Samples Collected	Number of Pre/Post Matched Sets
1) Spring 2012	Malaria	100 (20 days)	11th	2 total Physiology	60	26
2) Spring 2012	Malaria	60 (11 days)	9th	3 total	69	33
3) Spring 2013	Malaria (II)	100 (23 days)	9th	1 total Biology	34	16
4) Fall	Carbon Cycle	80	6th	1	40	18

2012		(15 days)				
5) Spring 2013	Carbon Cycle	60 (10 days)	6th	1	46	19
6) Spring 2013	Lice	100 (17 days)	6th	2 total	53	25
7) Spring 2013	Lice	85 (14 days)	7th	1	29	13

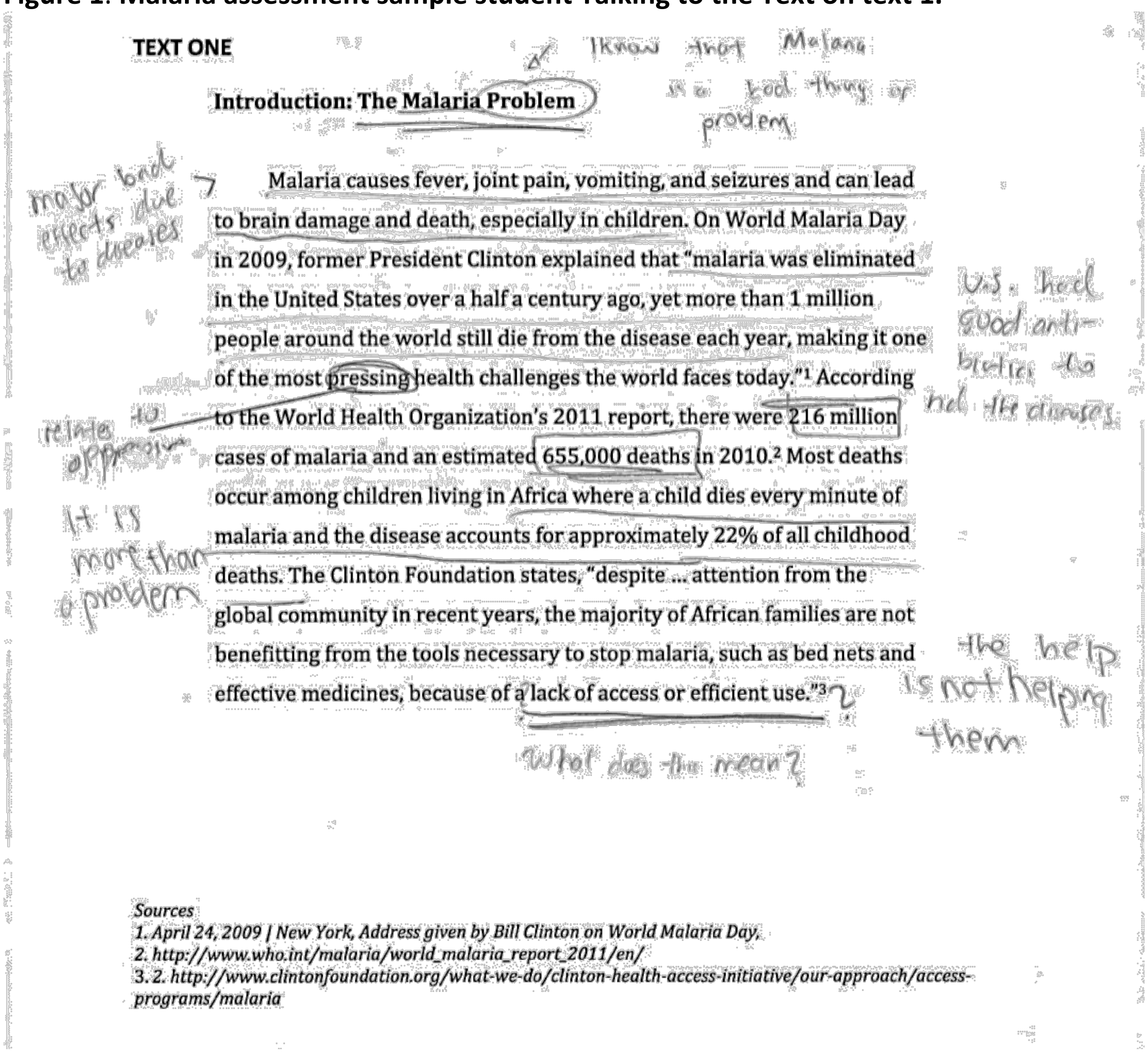
The Malaria assessment was administered pre- and post-implementation of the MRSA module in six classes in the spring of 2012 and the revised version of this assessment was administered in two classes in the spring of 2013. However, the high school MRSA investigation module was implemented to various degrees of completion in different classrooms. One teacher implemented the module for two weeks in four 9th grade Honors Biology classes in 2012. In this time, students completed about one third of the module, experiencing only one of the multi-text modeling and argumentation tasks in the module. Another teacher implemented the entire module in two 11th grade physiology classes in 2012 and in two regular 9th grade biology classes in 2013. The students in these latter classes experienced three multi text modeling and argumentation tasks. The Carbon Cycle assessment for the READI science Water module was administered in one 6th grade science class before and after students worked on the module in 2012-2013. In this case students completed about three quarters of the module. The middle school version of MRSA was implemented in two middle school classrooms during the 2012-2013 school year. In these classes, the pre/post assessment on pesticide resistant lice was administered.

Because the high school Malaria assessment was administered most frequently and across several grade levels and classes, it offered the richest data set from which to develop the model scoring and annotation coding strategies which we report below. Once developed, we applied the scoring strategies to the scoring of the Carbon Cycle and Lice assessments. To complement this work and further support teachers, we adapted these annotation scoring guides for use by teachers as ongoing formative assessment during the teaching of Text Based Investigations. In the remainder of this report, we describe the iterative process of scoring development, for both the models and the student annotations, using the Malaria assessment data from the 2011-2012 and 2012-2013 implementations of the complete MRSA module.

Development of Model Scoring Rubrics and Annotation Codes for Malaria Assessment

Across the pre and post assessments developed by the science design team, students were asked to carry out two interrelated science reading-argumentation tasks, 1) reading and annotating the source texts, and 2) developing a model explaining the phenomena presented in the source texts. See Figure 1 and Figure 2. The READI science design team therefore worked iteratively to develop a two-part approach to analyze and interpret the student work on the assessments, consisting of: 1) codes for analyzing students' close reading annotations, and 2) rubrics for analyzing students' constructed science models. We began by developing a rubric for scoring students' models.

Figure 1. **Malaria assessment sample student Talking to the Text on text 1.**



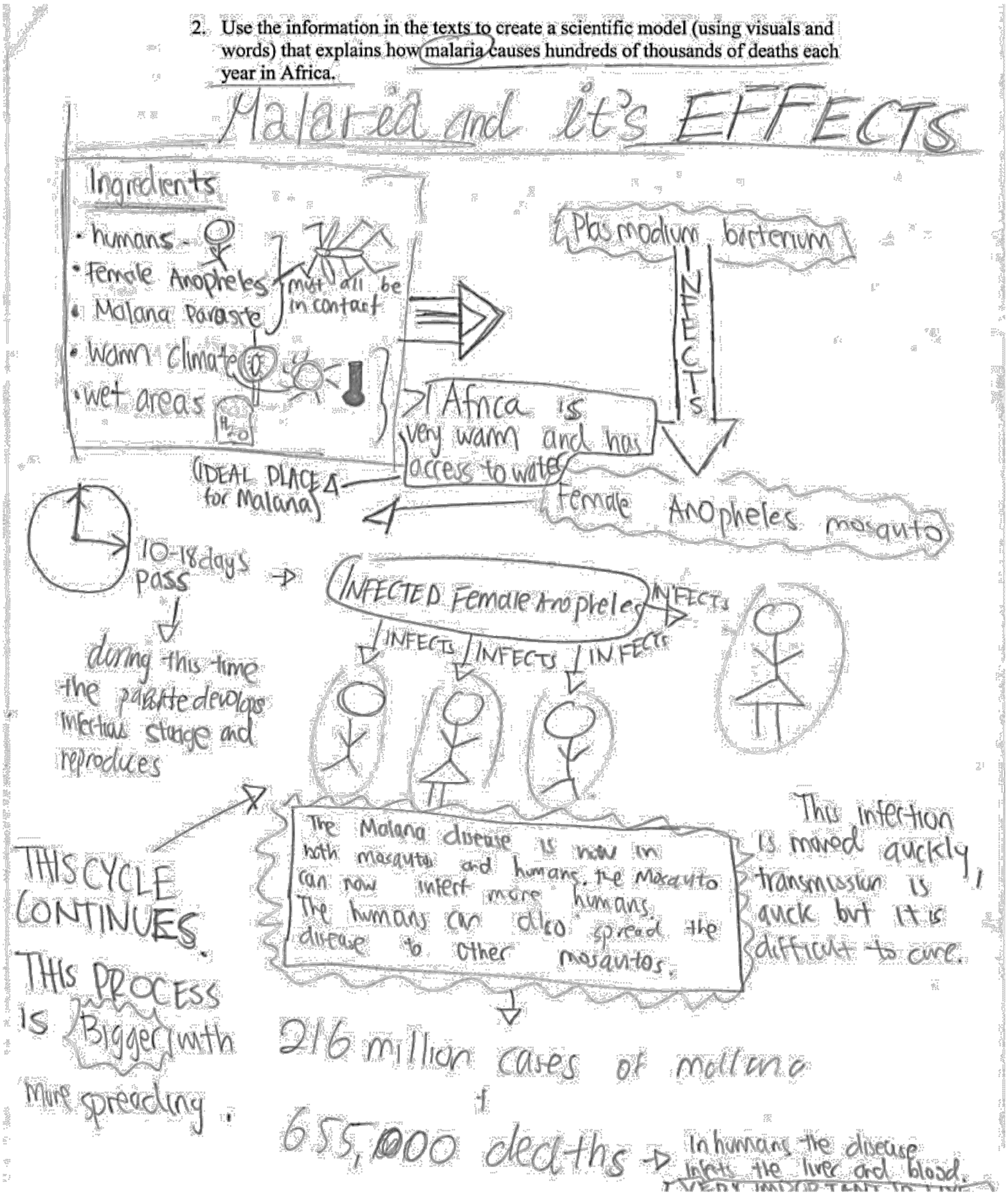
TEXT ONE

Introduction: The Malaria Problem

Malaria causes fever, joint pain, vomiting, and seizures and can lead to brain damage and death, especially in children. On World Malaria Day in 2009, former President Clinton explained that "malaria was eliminated in the United States over a half a century ago, yet more than 1 million people around the world still die from the disease each year, making it one of the most pressing health challenges the world faces today."[1] According to the World Health Organization's 2011 report, there were 216 million cases of malaria and an estimated 655,000 deaths in 2010.[2] Most deaths occur among children living in Africa where a child dies every minute of malaria and the disease accounts for approximately 22% of all childhood deaths. The Clinton Foundation states, "despite ... attention from the global community in recent years, the majority of African families are not benefitting from the tools necessary to stop malaria, such as bed nets and effective medicines, because of a lack of access or efficient use."[3]

Sources
1. April 24, 2009 | New York, Address given by Bill Clinton on World Malaria Day.
2. http://www.who.int/malaria/world_malaria_report_2011/en/
3. 2. http://www.clintonfoundation.org/what-we-do/clinton-health-access-initiative/our-approach/access-programs/malaria

Figure 2: **Malaria assessment sample student model**



Model Rubric Development Process

We approached the design process for the model rubric from a combination of top down and bottom-up perspectives.

Top down:

- **Causal Model:** We developed a causal model for the science system drawing on information from the text set to identify the components of the system and the relationship between the components. We also analyzed the relationship between the causal model and the information texts as to the amount of synthesis and inference responses the task might entail.
- **Expert Responses:** We constructed multiple expert responses to the modeling task to explore the potential range of representations for the models.
- **Targeted Learning Goals:** We drew on the READI core constructs for science (Evidenced aspects of the system, Cohesion and parsimony, and unifying concepts), identify potential dimensions for the rubric and initial indicators within these dimensions.

Bottom Up:

- **Grounding:** We examined multiple student samples to establish that the rubric dimensions are grounded in student practice.
- **Iterative Refinement**: We iteratively analyzed student work and refined the rubric to insure that the rubric language is precise and concise.
- **Scoring Calibration:** Five science module design team members analyzed a randomly selected set of ten students' pre and post assessments (five from each implementation site) to calibrate the scoring of the rubric and establish inter-rater reliability.

Malaria Model Rubric

The first iteration of the Model Rubric is shown in Figure 3. As the science team articulated this model for Malaria, we simultaneously kept in mind the goal of developing a generalizable approach for scoring students' science models of other phenomena. Thus, the dimensions labeled State of human and mosquitoes; Movement of plasmodium; and Aggregate effects might be represented more generically as Elements; Interactions; and Effects.

Figure 3: Malaria Model Rubric

DIMENSION 1: State of human and mosquitoes with respect to plasmodium infection

	1	**2** (Macro)	**3** (Macro)	**4** (Micro)
A. Humans (H)	Inclusion of human in model	Humans can exist in EITHER infected or uninfected	Humans can exist in BOTH uninfected and infected by malaria	Level 3 AND describes 2 or more stages/process of malaria propagation inside the humans
B. Mosquitos (M)	Inclusion of mosquito in model	Mosquito can exist in EITHER infected or uninfected	Mosquito can exist in BOTH uninfected and infected by malaria	Level 3 AND describes 2 or more stages/process of malaria propagation inside mosquitos (*i.e. zygotes, cyst, sporozoites)*
C. Context (C)	Y/N			

DIMENSION 2: Movement of plasmodium

	1	**2**	**3**	**4**
A. Movement of plasmodium through mosquito feeding on human blood	No mention of infection and disease as stages or how it moves from host to host. *NO identification of plasmodium as 3rd species*	Evidence of ONE-WAY transmission (either way through 'bite' or 'spread'), but only ***that*** it happens.	Evidence of BI-DIRECTIONAL transfer of plasmodium *(Ok if detail for one direction is stronger than the other)*	BI-DIRECTIONAL movement and clear description of the parasite moving between the organisms: 1) Movement is bi-directional 2) Explicit mention of what actually moves between (liver, zygote, sporozoite)* *Plasmodium has to be named as separate species*
		The words infection, disease, malaria, etc. can all be used to describe the 3rd species		
B. Continuity & transmission of plasmodium (*directionality)*	No mention of H → M or M → H spread	Either mention of H → M or M → H	Both H → M and M → H	Discusses H → M and M → H as continuous cycle (through verbal or through visual – process repeats, circular)
C. Intervention	Y/N			

DIMENSION 3: Aggregate level effects

	1	2	3
Aggregate level impact of malaria	No mention of aggregate level outcome	Discusses the large-scale impact by going beyond the human ←→ mosquito relationship, but does not clearly or accurately delineate how human ←→ mosquito *leads to* large scale impact	Integrated description of how human ←→ mosquito interactions leads to aggregate level impact on large population. This must include: 1. many mosquitos and individuals have malaria 2. one single mosquito can infect multiple individuals 3. multiple individuals can transfer plasmodium to many mosquitos

Iterative Development of Science Annotation Codes

A key component of the READI approach to evidence-based argument is close attention to text. As described in Project READITechnical Report #17, the Text-Based Investigation modules for science classrooms work to establish and draw on close reading of text, as evidenced through individual and small group annotation of the sources provided to students. Annotations reflect ways in which readers Talk to the Text and can show evidence of the mental processes readers engage in to make meaning of texts, as well as their comments (cognitive and affective) about the text, the content, and/or their own state of knowledge and understanding (Schoenbach, Greenleaf & Murphy, 2012, pp. 108-110). The pre-post assessments were designed in parallel to the modules. Students were asked to Talk to the Text and thereby show their thinking in the margins of the assessment texts. The resulting annotations become a source of information about the ways students are engaged with, and processing, the texts provided. The science design team undertook an iterative process to develop an annotation coding system to analyze impact of the modules on students' processing and engagement while reading.

A long history of process tracing studies have verified the repertoire of strategic processing moves that accompany reading to understand (Ericsson & Simon, 1980, 1993). Methods of process tracing have included think aloud protocols during reading, as well as retrospective think alouds to recover traces of text processing (Afflerbach & Cho, 2009; Afflerbach, 2000; Afflerbach & Johnston, 1984; Paris & Flukes, 2005; Pressley & Afflerbach, 1995; Veenman, Van Hout-Wolters, & Afflerbach, 2006). In addition, text processing studies have utilized verbal protocols as well as eye tracking technologies to associate directional eye gazes and dwelling on parts of texts with hypothesized processing (Braten & Stromso, 2003; Coté, Goldman & Saul, 1998; Goldman & Durán, 1988; LaBerge & Samuels, 1974; McConkie, 1997; van Dijk & Kinstch, 1983; Wolfe and Goldman; 2005). As the introduction to this technical report indicated, we are drawing on this tradition of studies to take as data the comments and marks readers make as they interact with text, when they are asked to do so as part of an assessment task.

It is important to distinguish the Talking to the Text routine promoted in Reading Apprenticeship as metacognitive conversation with text, from the many procedural techniques for “annotation” proliferating in literacy pedagogy. Students are often taught in school to use specific icons or marks to indicate when they have a question or where the vocabulary in a text is unfamiliar or unknown. This sort of proceduralization easily gives way to perfunctory rather than thoughtful completion of school tasks. In contrast, Talking to the Text is an open-ended, unconstrained interaction with the text meant to capture and make visible the thinking processes readers engage alongside (on the margins of) the text as they read. Talking to the Text is therefore idiosyncratic, by virtue of being a record of an individual’s interaction with a text, and includes a combination of markings (arrows, circles, underlining, highlighting) and commenting. In Talking to the Text, students are encouraged to make authentic comments (their real questions, reactions, and conjectures) rather than extracting verbatim quotes from the text itself, as in Cornell Notes. Thinking, rather than indexing, is the focus of this routine.

We recognize the annotations students make when encouraged to Talk to the Text to be an incomplete record of the mental processes readers engage in, just as think aloud protocols are unlikely to capture all of what occurs at the speed of electron transport in the working mind. However, these annotations leave traces or artifacts that provide clues to how students are proceeding through texts. Through careful definition and operationalization of codes, we are able to identify strategic moves such as summarizing, in the comments students make along the margins of texts. Such strategic moves may be associated with different intentional processes. A student may be summarizing what she understands thus far, in order to remember important information in the text. On the other hand, she may be summarizing in order to judge the adequacy of her understanding of the text. We have no way to disambiguate these possible processes from the trace left on the page. While we cannot know the intention behind a reader's margin comments when they Talk to the Text, we can describe these comments and identify types of interactions between reader and text. We can use these comments and marks on the page to look for such important comprehension factors as self-explanation, a focus on constructing understanding and a reader's attention to evidence. With this set of understandings we set out to develop a low-inference coding scheme that would characterize the observable interactions between student readers and the science texts in the assessments.

Methodology

Similar to the way we approached the model rubric development, we approached the design process for the annotation coding system from a combination of top down and bottom-up perspectives. To understand the impact of integrating reading for understanding with scientific practices through Text-Based Investigations, we asked, How are students interacting with text when asked to Talk to the Text? Using annotations as data about student reading and reasoning processes required us to develop a system for coding the raw marks students left on the page. For the purpose of analyzing student samples of Talking to the Text, we identified discrete marks and comments as the unit of analysis for coding. We made decision rules that defined the beginnings and endings of each mark or comment, each of which constituted one annotation. We then coded each discrete annotation. Figure 1 illustrates a sample of Talking to the Text on the Malaria assessment.

To develop the codes for annotations, we were informed by theoretical constructs from the field of literacy, particularly descriptions of comprehension supporting strategic processes (Afflerbach & Cho, 2010). We drew on our previous studies and development of the Integrated Learning Assessments for which we had developed annotation scoring rubrics and processes (Greenleaf, et al., 2011; Silver, et al., 2011) related to biology reading strategies and multiple forms of texts. In addition, we drew on the Reading Apprenticeship framework and its focus on metacognitive conversation (Schoenbach, Greenleaf & Murphy, 2012) and our own reading process analyses as we ourselves carried out the reading and modeling tasks in the assessments. Additionally, we drew on Cromley's work on science diagram comprehension (Cromley, Snyder-Hogan, & Luciw-Dubas, 2010) and the READI Student Learning Goals for science to ensemble a list of potential annotation codes. This gave us an initial set of theoretical concepts and potential categories of codes.

At the same time, we worked from an initial open coding of the raw annotations on the page to begin to develop labels, moving toward a focused coding and category development as we iteratively examined and reexamined the data (Charmaz, 2008a, b; Miles & Huberman, 1994; Miles, Huberman & Saldana, 2013). The coding process was thus theoretically informed, but grounded in the data (Glaser, 2011).

Thus the work of annotation coding proceeded in an iterative fashion, across multiple samples of student assessments and sets of researchers and progressed over two years toward a set of reliable codes that could be used by multiple coders to gauge student learning from the READI science Text Based Investigation modules. In Phase one, the READI science design team worked first with the high school Malaria 2012 and 2013 student assessment samples, developing an initial set of codes. The resulting initial code book was then discussed, debated and honed by a larger, cross sectional group of READI project members. A subgroup then took the resultant refined code book and tested, debated, refined and negotiated what was possible to understand from students' comments and marks. In Phase two, the coding team employed Atlas.ti to complete a multi-tiered consensus coding protocol, using the high school Malaria 2012 and 2013 and to the Middle school Lice 2012-2013 data. We will now turn to describing each phase of this work in detail.

Initial Annotation Code Development Process

Working with the Malaria student assessments collected in 2012, we analyzed several class sets of student annotations using these potential annotation codes, grounded the codes in actual examples, and resolved these codes into five distinct types:

1. Kind of Annotation (Comment or Mark),
2. Association of Annotations (with running text or with visuals/diagrams),
3. Degree of Student Voice,
4. Close Reading Processes, and
5. Scientific Reading Processes.

As we coded and developed these categories, we simultaneously developed functional definitions and examples for the first four types. What resulted was a set of initial descriptive codes for reading processes. Always in view was the utility of the coding system and examples for use by teachers in formative assessment.

In working toward, a functional description and definition of the fifth type, Scientific Reading Processes, we noticed in student annotations many instances of students interacting with the science texts in scientific ways that were not well characterized by the preliminary codes we had developed for this type. We then drew on the model rubric we had developed for analyzing students' pre and post assessments to characterize the kinds of work indicated by the students' annotations. We found clusters of examples of student annotations focused on understanding the elements of the phenomenon and other annotations focused on understanding how the elements of the phenomenon related to one another. In addition, we noticed annotations indicating attention to the implications of the phenomenon, the significance of the phenomenon, and responding to the challenge presented by the phenomenon. Through ongoing analysis of student annotations, we formed definitions and found examples for reading processes related to developing a causal model, which we viewed as indicators of nascent modeling and a particular important form of evidence of the impact of our work with students in reading science for (science) understanding.

Nascent Modeling Science Reading Codes

- Attending to the phenomenon
- Clarifying/Inquiring about the phenomenon
- Attending to the elements of the explanatory model
- Clarifying/Inquiring about the elements of the explanatory model
- Attending to relationships (links) in the explanatory model (relationship between elements and links)
- Clarifying/Inquiring about the relationships (links) in the explanatory model (relationship between elements and links)
- Generating a course of action based on the explanatory model
- Attending to the scope and significance of the system

We defined several terms for common reference by various coders, along with the definitions of specific codes, in a Coding Organizer related to the Malaria assessment. This glossary included the following terms:

- **Phenomenon**: Anything about the science topic (issue of malaria in the world)
- **Explanatory model**: causal account for phenomenon: details *how* malaria is spread and leads to human deaths
- **Elements** (parts of the explanatory model): humans, mosquitoes, environmental context

- **Links** (relationships within explanatory model between elements) - how the plasmodium parasite is transmitted (i.e. when uninfected mosquito biting infected human, infected mosquito biting uninfected human)

Initial Annotation Analysis for 2012 Malaria Sample

We took several passes at coding the complete set of 2012 Malaria pre/post annotations. The first draft of the codebook was completed by three coders in Sept, 2012 and subsequently. Table 2 shows the pre and post mean values for each code for the entire sample, comparing each code mean value for the three classrooms combined.

Table 2. Mean Values for Pre-Post Annotations across 2012 Malaria Assessment Sample

Student Annotation Type	Text	MEAN PRE TEST	MEAN POST TEST
Total Number Marks	1	5.29	6.81
	2	9.43	8.82
	3	2.84	2.38
	4	0.69	1.11
Total Number of Comments	1	1.45	2.79
	2	1.75	2.94
	3	0.70	0.90
	4	0.62	0.72
Total Number Paired Marks and Comments	1	0.86	1.73
	2	1.04	1.88
	3	0.40	0.44
	4	0.40	0.51
Number Marks Associated with Text	1	5.37	6.73
	2	8.71	8.57
	3	2.40	2.17
	4	N/A	N/A

Number Comments Associated with Text	1	1.33	2.63
	2	1.55	2.92
	3	2.86	0.88
	4	N/A	N/A
Number Marks Associated with Visuals/Diagram	1	N/A	N/A
	2	N/A	N/A
	3	N/A	N/A
	4	0.67	1.13
Number Comments Associated with Visuals/Diagram	1	N/A	0.00
	2	N/A	0.00
	3	N/A	0.00
	4	0.62	0.70
Total Number Comments in Students' Voice	1	1.14	2.50
	2	1.41	2.84
	3	0.65	1.08
	4	0.74	0.74
Total Number Comments in Author's Voice	1	0.33	0.06
	2	0.35	0.14
	3	0.09	0.00
	4	0.12	0.00

On average, the students marked Text Two (see Apendix A– Malaria Assessment) the most, both for pre and post tests, with 9.43 and 8.82 average markings respectively. We expected a higher number of annotations for this text, being a longer text than Text One (the next longest text) by four sentences. Next, in decreasing order and consistently across pre and post assessments were Text Two with 5.29 (Pre) and 6.81 (Post) marks; Text Three with 2.84 (Pre) and 2.38 (Post) markings; and, Text Four with 0.69 (pre) and 1.11 (Post) markings. Scoring also indicates that students increased their annotations on Texts One and Four after instruction in the MRSA module, reduced markings on Text Two slightly, and maintained about the same number of marks on Text Three.

In contrast, Comments increased consistently from pre to post assessment annotation for all students and for all texts. This is important since Comments more explicitly reveal student thinking in response to the text than do marks such as underlining or circling, which require much more inference to code. It also suggests that students were increasingly showing their thinking alongside text in the context of the MRSA Module implementation which they had experienced, where Metacognitive Conversation and Talking to the Text was encouraged.

The use of Paired Marks and Comments was the least used annotation strategy by students. Frequently, annotations on margins referred to ideas on the text, but only a portion of the sample used lines or brackets, arrows or a mark that was paired to the text and their comment. Students increased the use of this annotation strategy during post assessment annotation across all four texts. Text One increased from 0.86 to 1.73 average paired marks and comments; Text Two, from 1.04 to 1.88 average paired marks and comments; Text Three had the least change from pre to post with 0.4 average to 0.44 pairings; and Text Four increased from 0.40 to 0.51 average paired marks and comments.

On average, students made more marks than comments. The student annotations included many different types of markings, but the average for each text and across the sample had approximately 17-25% less comments than marks (e.g. Text One had 1.33 average comments as compared to 5.37 average marks). Text Four, the visual representation, was the least annotated text. We predicted that students' unfamiliarity with reading visual texts would result in lower annotations on such texts and believe this accounts for the relative paucity of annotations. Even so, the increase of marks and annotations on this text is evidence of some success in helping students view such science texts as requiring reading and comprehension.

Finally, student comments were coded for the degree of student voice found in the annotation. The data shows that students on average increased by 50% the use of their own voice with a 70% decrease in verbatim (author's voice) comments from pre to post assessment annotations in the expository texts (Texts One, Two and Three). For Text Four, the visual text, the mean number of student comments in students' own voice remained constant (0.74) and the mean number of verbatim comments decreased from 0.12 to 0. We view student voice as a marker of knowledge transformation, as students take the text as written and transform it in the process of making meaning.

The initial pass at analyzing the Malaria assessment through annotation coding thus reinforced the science design team's approach to Text-Based Investigations in science, in which reading and literacy strategies are modeled for making sense of text and transforming text propositions into one's own understandings. These initial findings showed an increase in frequency of student annotations on the texts from pre to post, as well as increased frequency of comments in the student voice pre to post (as opposed to comments displaying verbatim language from the text), suggesting student development through the MRSA module along expected lines: more visible and in depth engagement with text and annotation. The coding of assessments highlighted for us the importance of student voice in science reading and comprehension processes.

Refinement of the Annotation Codes

Refining Close Reading and Scientific Reading Annotation Codes

To refine the annotation coding, we turned to the 2013 sample of Malaria assessments that were collected in AJ's 9th grade biology class. In previewing the 2013 Malaria pre-post Assessments we noted an increased frequency of nascent modeling in the annotations compared to the 2012 Malaria pre-post assessments. We anticipated that the 2013 Malaria pre-post assessments might offer better opportunity for refinement of the annotations codes than the 2012 Malaria pre-post assessments. Therefore, to expedite refinement of the codes, we began analyzing the 2013 pre-post assessments. Three coders independently coded 17 pairs of pre-post assessments. The 2013 pre and post data included 305 comments of which 273 indicated close reading and 270 indicated close scientific reading, providing plentiful opportunity to confirm or refine the annotation codes. The analysis led to discovery of no additional codes nor any substantive changes to the code specifications (rules for operationalizing).

Final Refinements of Annotation Codes and Codebook

Having iteratively developed a set of codes for analyzing students' annotations on science texts, we worked across institutions and subject area teams to test the robustness of the code book, and open it to further refinement. In March 2014, the team working on annotations analysis expanded from the core coders previously involved to include thirteen team members from across Project READI. This group met four times to hone understandings of the goals for the code book and further refine the definitions to increase understanding across a wider project group. We expanded the data set to include the annotations on science EBAIMs pre/post tests developed by the NIU team in collaboration with the READI science design team. This team was then reduced to two researchers from UIC and five researchers from SLI, two of which were core members of the Science team. This smaller group met 14 times between April and August and coded the Science EBAIMS, and then moved back to the Malaria data. The group also worked to consolidate and refine the codebook further and come to agreement on code definitions.

This process both enlarged the circle of coders and through an iterative process tightened the operational definitions of each code in order to support achievement of consensus across coders and reliability across coding groups (Mays & Pope, 1995, Miles & Huberman, 1994). We assigned samples to pairs of coders for consensus coding, conducted reliability checks, reconciled different interpretations of the data, and modified the codebook. In addition, we brought the codes to the California Teacher Inquiry Network members, and began working with them to think about how these codes might be used in the classroom as part of formative assessment. Literature and History teams worked to create codes for historical thinking and literary analysis in place of the science reading codes. (See READI Science Annotation Guide, Appendix B, and described below.)

As we worked across data sets and teams, we moved toward a systematic coding process that could support low-inference descriptions of students' marks and comments on the text, when requested to Talk to the Text. Finally, a group of 5 researchers conducted a final coding of the entire Malaria data set using a similar process as that described above, but with less modification of the code book. In practice, this constituted a form of consensus coding process that resulted in modification of the codebook to reach reliability, a process prescribed in Hruschka, et al, 2004.

The final codebook included 32 primary codes grouped into five types: kind of annotation, association of annotation, degree of student voice, close reading processes, and science reading processes. The final codebook differed from the initial in two fundamental ways. First, the dependencies among the codes were reduced. In the initial codebook the student voice code was a necessary precursor to considering close reading codes. The dependency on student voice was intended to underline the differences between capturing authentic reader/text interactions through Talking to the Text and notetaking procedures more broadly defined. Initially, we had not coded any comments that were verbatim copies of the original text to identify close reading processes. When working with the cross-disciplinary group, we realized that paraphrasing and summarizing as close reading processes do not always evidence student voice, so the dependency of close reading on student voice was removed. See Figure 4. In addition, we further defined requirements for what counted as student voice, and it was operationalized as follows: Mark as student voice if a paraphrase features at least 25% of words changed or there is a substantial change in sentence structure from the original. This decision rule allowed us to demarcate the linguistic transformations on text that indicated students had done something with the text other than simply copy it onto the margins.

Another shift that our cross-site collaboration yielded was in how we accounted for comments connected by a mark (line, arrow, shape etc.) to text or image. Initially, we would did use this text to understand if the comment could be considered close reading, and/or scientific close reading. However, much as the case with paraphrasing and summary, we realized that readers often utilize the verbatim text substantively in Talking to the Text and indicate this with connecting marks. We therefore chose to take into account the connected text when assessing a comment for close reading and scientific close reading. This was operationalized as follows: IF coded with the comment & mark associated code (if the comment was clearly connected to associated text) use the marked text to determine what close reading or scientific close reading code type to apply.

Second, the number of primary codes was reduced from 44 to 32. See Table 3. Six close reading codes were eliminated. Ten science reading codes were collapsed to four. Two science reading codes were eliminated. Two new science reading codes were added. See Table 3. In the close reading codes, there were three types of codes that were winnowed: codes around specific word and term use, codes identifying concepts in the text, and codes around puzzlement. All of these moved from a finer to a coarser grain, which both better reflected the data itself, and also allowed for less ambiguity in the application of the code book. Perhaps the most important, and tricky to resolve, was moving the three inquiry codes (Identifying/ expressing confusions and/or roadblocks, using context clues to build understanding, Inquiry questioning for knowledge building) into the code Inquiry question for knowledge building. Inquiry is central to our teaching and learning model and thus we feel it is critical to catch all types of inquiry and to support teachers to take note of any variety of student inquiry. However, student comments in this data sample did not support the variety of inquiry stances we often hear in the classroom and see in student work. Thus we came to operationalize this broader code as follows: An inquiry question is a question posed that suggests comprehension rather than confusion with the meaning of the text AND asks for information beyond what is given explicitly in the text. The question can be posed in the form of a statement ("I wonder if..."). If it simply asks for the meaning of word, code as identifying unknown vocab.

The large cross site discussion helped the team at SLI to see that some of the original science reading codes were capturing information that was contained in the close reading codes. This observation supported a change to focusing the science reading codes on the science in the text that was attended to in the student comment. Science modeling is the overarching focus of the modules the assessments were designed to compliment. Thus the initial code book broke modeling down to a finer grain than was represented in the data. For these reasons codes such as Attending to phenomena and Attending to elements of the model were collapsed into a single code.

As a whole, these modifications to the code book helped to clarify the work for the coders and to bring the most important concepts that we designed into the MRSA module to the fore. They resulted in a code book that was both educative of important reading processes (and was subsequently taken up and used by teachers and professional development facilitators in such a way) and useful as an assessment methodology. As an analysis tool, many of the codes continued to elude high degrees of inter-rater reliability. This fact drove a complex consensus based coding system to be developed (described below) for the final analytic work.

Figure 4: Initial and Final Code Dependencies

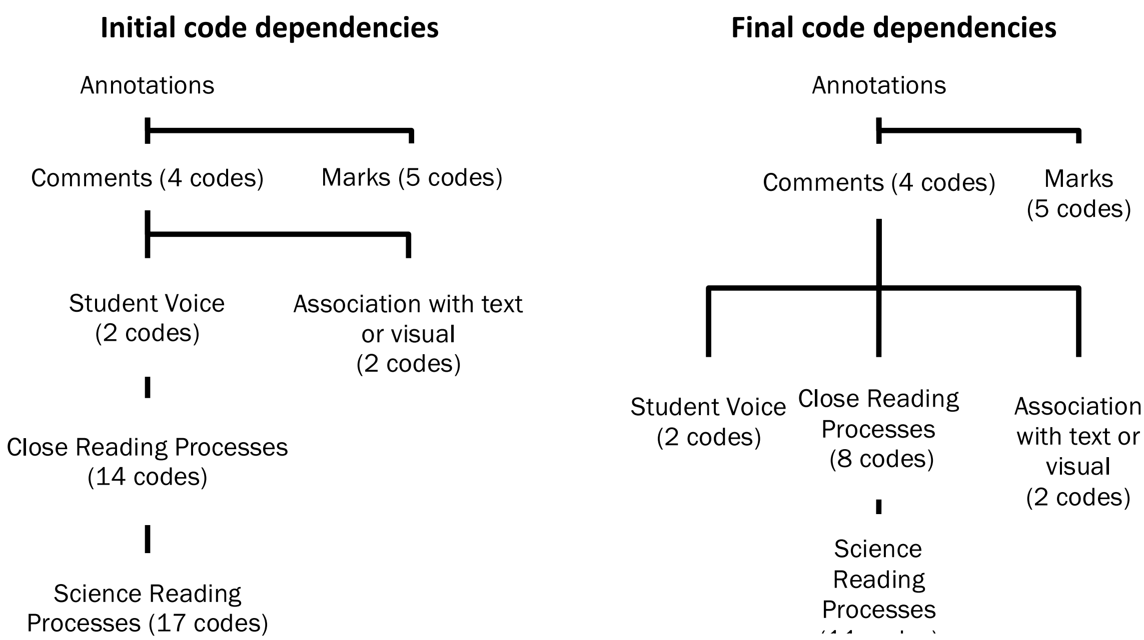


Table 3: **Initial and Final Codes**

Initial codes	Final codes
Type 1: type of annotation	**Type 1: type of annotation**
marks: *underlining*, *highlighting*, *circling/boxing*, *connecting lines / arrows*, *symbols* Comments: *single words*, *Statements*, *Questions*, *paired with marks*	marks: *underlining*, *highlighting*, *circling/boxing*, *connecting lines / arrows*, *symbols* Comments: *single words*, *statements*, *questions*, *paired with marks*
Type 2: association of comments	**Type 2: association of comments**
Running text — *Visual*	*Running text* — *Visual*
Type 3: degree of student voice	**Type 3: degree of student voice**
Students voice — *Verbatim*	*Students voice* — *Verbatim*
Type 4: close reading processes	**Type 4: close reading processes**
Identifying key vocabulary	Identifying unknown words
Identifying unknown vocabulary	
Attempting to define unknown vocabulary	*Defining vocabulary*
Identifying main ideas of the text	
Paraphrasing and summarizing	*Paraphrasing and summarizing*
Identifying familiar concepts	
Making connections with knowledge	*Making connections with knowledge*
Making connections within text	*Making connections within text*
Making connections text to text	*Making connections text to text*
Predicting/inferencing	*Predicting/inferencing*
Identifying/expressing confusions -&/or roadblocks	
Using context clues to build understanding	
Inquiry questioning for knowledge building	*Inquiry question for knowledge building*
Other close reading process	
Type 5: science reading processes	**Type 5: science reading processes**
Identifying conventions of a science text	
Identifying cross-cutting/unifying concepts in science	*Attending to cross-cutting concepts*
	Attending to scale, quantity and proportion
	Attending to stability and change
Attending to science in text	*Attending to science*
Clarifying/inquiring about science in text	
Attending to the phenomena	*Attending to the phenomena/elements*
Clarifying/inquiring about the phenomena	
Attending to the elements of the explanatory model	
Clarifying/inquiring about the elements of the explanatory model	
Attending to relationships in the explanatory model between elements and links	*Attending to interactions*
Clarifying/inquiring about the relationships in the explanatory model between elements and links	
	Generating a model
Evaluating /questioning of science model or explanation	
Attending to arguments	*Attending to arguments*
Clarifying arguments-inquiry-questioning of science arguments	
Evaluating arguments	*Evaluating arguments*
Attending to scope and significance of the system	
Supporting an assertion	*Supporting an assertion*
Generating a course of action	*Generating a course of Action*

Using the Refined Annotation Coding and Model Scoring Processes to Explore Student Learning from Text-Based Investigation

Process

To systematize the approach, we investigated various qualitative data analysis systems. Student annotations, since they often are situated close to one another in visual space, present a substantial challenge to the use of such analysis systems. After much investigation, we selected Atlas.ti qualitative analysis software because we were able to select individual student comments and code them, keeping the frequencies of code types even when student annotations overlapped one another. Using Atlas.ti, we are able to select students' handwritten marks, underlining and comments in the margins of electronic texts, code these traces for the evidence they provided regarding student interaction with the reading tasks and processes, and develop quantitative reports demonstrating changes within and between groups of students over time.

The annotations code book resolved into two distinct categories of codes: Marks (type 1) and Comments (types 3, 4 and 5). Marks are an important trace of student interaction with the text, however they are wholly unambiguous. Thus we had a single coder go through the entire data set to record all the underlining, boxing and circling that was evident. In our reporting we chose to roll these separate codes up and report them together.

Comments proved more challenging to code, particularly when looking at close reading (type 4) and scientific close reading (type 5). The code book works, in a sense, as a pyramid of increasing complexity. Discrepancies between coders in the type 3 codes were generally clerical in nature. It was the complexity of the interactions captured in types 4 and 5 that led us to continue with a consensus coding process for analysis. In particular there were challenges discerning between the three "making connections" codes. We continued to use the three codes, with the understanding that the codes were not meant to describe a student's internal thinking and knowing processes, about which we could make no confident claims. Rather, these codes aimed describe how a student was working with the texts presented. Therefore, the aim was to take a very literal read of the text – and if the comment did not contain explicit information from within the text set, we coded the comment "making connections to outside knowledge." Furthermore, because we have no way to know what knowledge the student brought to the text set from prior experiences, we chose to group these potentially different types of "making connections" into a single "making connections" code. Similarly, predicting/inferencing and summarizing/paraphrasing were often coded differently by the members of the final coding team. To mitigate these challenges and still produce an analysis we felt confident was explanatory of the patterns of student engagement with text, we engaged in the consensus coding process as follows.

To code the comments, we used the consensus coding protocol, as described above: Each person coded an assigned set of samples, then met with a partner or in a trio to reach consensus. We conducted a reliability round, bringing all five coders together for discussion of a sample at the midpoint of coding to ensure that all the individual coders were still applying codes in a normative way, and then continued the consensus coding. The consensus groups were changed every five sets of samples, to keep any pair or trio from drifting from the other coders. Samples were anonimized but each set included both the pre and the post samples for the same five students.

Use of Atlas.ti allowed for a clean export of the final codes into Excel for statistical analysis. In addition, coders were able to make memos on the student work as they went. To facilitate working in the PDFs of student materials, each coder was provided with an analysis subset containing just the student assessment samples they were to code. After coding, the pairs or trio merged their independently coded set of assessments into a single Atlas.ti file. Any discrepancies were discussed and resolved. The Atlas.ti files containing the final coding results for each subset of student assessment samples were then merged into a complete Atlas.ti document containing all the consensus coded assessment samples. Once all the samples were aggregated, the results were exported by student and by code to various Excel files.

The coding team first completed all the coding of the Year 2 and Year 3 Malaria samples. Next we coded the annotations in the Lice data set, the pre/post assessment for the middle-school MRSA unit. The data set consists of 82 samples from 2 Chicago READI teachers' Year 3 classrooms, with 38 matched pre/post pairs. These teachers, RL and KM, enacted the READI MRSA module in their classrooms, for 14 and 17 days respectively, and the Lice assessment was given as a pre/post test (see the Year 4 Annual Report for more on these classrooms). This data set was coded by two researchers who had worked on the Malaria data coding set, using a consensus pair coding protocol, coding individually then meeting to reach consensus.

With the annotations coding complete, we scored the scientific models students had created as part of the assessment. The models from the 2012 classrooms had already been scored; using the rubric we developed that year. Two coders rescored 8 of the 60 Malaria models that had been scored in 2012, reaching 85% fidelity to the previous coding. Given this high level of fidelity, we decided not to rescore the remaining models. We then proceeded to code the 2013 Malaria models and the Lice models. As in the Annotation coding process, the coders scored all the models independently and met to reach consensus on disagreements.

Analytic Sample

To analyze the impact of Text-Based Investigation on student learning, we focused on a single teachers' implementation of the MRSA module in high school science courses over two years. Over these two years, AJ had implemented the entire modules at two different grade levels and had administered the Malaria assessments before and after the module. The final Malaria data set from AJ's implementation of MRSA is comprised of 95 student samples from three classes gathered over these two years. The Year 2 set is from two eleventh grade physiology classes taught by AJ, a participant in the California TIN, in which we conducted 20 days of interventions. This set has 60 total samples, with 26 matched pre/post pairs. The set from 2013 is from a single ninth grade high school biology class also taught by AJ; we conducted 23 days of intervention in this class. This set has 34 samples, with 16 matched pre/post pairs.

Results from Annotation and Model Analysis

To assess the impact of the MRSA Text Based Investigation on student learning, we asked, did student Talking to the Text practices change, and if so, how? We also wanted to understand the impact of the MRSA implementation on students' science modeling, as an indicator of science learning. One measure of change is the classroom average frequency of each code and code type on students' pre and post assessment. This analysis could uncover broad shift in students' Talking to the Text. For example, in the 11^{th} grade and 9^{th} grade data, the overall number of marks and comments increased from pre to post (See **Tables** 4 and 5). Similarly, the number of comments evidencing student voice, close reading processes and science reading processes increased. These results suggest the promise of the MRSA investigation for student learning and the utility of the annotation codes as a measurement tool.

Analysis of the model scores can help us to address the impact of the MRSA implementation on student science modeling. One measure of change is the classroom average of the model summary and individual item scores. The shifts in scores, while uniformly positive, reached rho<0.05 reliability for a subset of the items. This result suggests that the shifts in modeling were small compared to the sensitivity of the measurement tool. This result provides some grounds for theorizing about the observed outcomes. For example, it may be that the MRSA investigation provided these students with more support for close, scientific reading than for modeling. Alternatively, gains in reading engagement may precede gains in modeling from text comprehension as in the MRSA design.

Table 4: Pre/Post MRSA Unit: Changes in Students' Science Reading and Modeling
11th Grade Physiology Students (n = 26, matched pre-post samples)

Annotation Codes	Average Pre	Average Post	Average Change
Marks	10.9	16.9	6.0*
Underlining, boxing, circling	9.3	11.9	2.7
Connecting Marks	1.5	5.4	3.9*
Symbols	0.3	0.4	0.2
Comments	3.7	9.2	5.5**
On verbal text	3.6	8.5	4.8**
On visual text	0.1	0.7	0.6*
In student voice	2.4	8.8	6.4***
Close reading processes	2.5	6.8	4.4**
Identifying unknown words	0.0	0.3	0.3
Defining vocabulary	0.0	0.2	0.2
Summarizing, paraphrasing	1.8	2.4	0.6~
Connections outside texts	0.2	1.2	0.9**
Connections within texts	0.4	1.2	0.8**
Connections across texts	0.3	0.3	0.0
Making predictions / inferences	0.3	1.5	1.1*
Questioning	0.2	2.4	2.2***
Science close reading processes	2.3	6.2	3.9**
Science	0.0	0.1	0.1~
Phenomena & elements	1.8	4.9	3.0*
Element interactions	0.4	1.2	0.8*
Scale, quantity and proportion	0.2	0.5	0.3
Model generation	0.1	0.0	-0.1
Model Scores			
Model summary score (0-16)	5.1	6.9	1.8*
Model elements score (0-6)	2.6	3.3	0.7~
Model interactions score (0-6)	2.2	3.0	0.8~

Table 5: Pre/Post MRSA Unit: Changes in Students' Science Reading and Modeling
Practiced 9th Grade Biology Students (n = 16, matched pre-post samples)

Annotation Codes	Average Pre	Average Post	Average Change
Marks	11.8	24.3	12.5*
Underlining, boxing, circling	8.8	16.5	7.8*
Connecting Marks	2.6	6.9	4.3~
Symbols	0.6	1.3	0.7~
Comments	6.8	11.4	4.6
Comments on verbal text	6.7	10.2	3.5
Comments on visual text	0.1	1.3	1.1
Comments in student voice	5.9	10.7	4.8*
Close reading processes	5.4	9.6	4.3~
Identifying unknown words	0.1	0.2	0.1
Defining vocabulary	0.0	1.0	1.0*
Summarizing, paraphrasing	3.1	4.6	1.4
Connections outside texts	0.5	1.7	1.2
Connections within texts	0.8	1.8	0.9
Connections across texts	0.1	0.5	0.4*
Making predictions / inferences	1.2	2.1	0.9
Questioning	1.1	1.3	0.3
Science close reading processes	5.4	8.8	3.4~
Science	0.1	1.1	0.9~
Phenomena & elements	4.2	6.4	2.3
Element interactions	1.0	1.3	0.3
Scale, quantity and proportion	0.0	0.1	0.1
Model generation	0.0	0.2	0.2
Model Scores			
Model summary score (0-16)	3.8	5.4	1.6
Model elements score (0-6)	1.6	2.6	1.0~
Model interactions score (0-6)	1.6	2.1	0.5

Another application of annotation coding and model scoring is to explore the potential relationship between students' reading processes and quality of models. This question is significant to assess the promise of Text Based investigation for simultaneously supporting reading and modeling in science. An analysis of the correlation of individual students' comments and model scores might provide evidence of a relationship between reading processes and modeling in the pre/post data. To explore this potential relationship, we included pre and post data from both the 9th grade and 11th grade Malaria data sets, 84 samples in all. We conducted this analysis using the frequency of comments, student voice comments, close reading comments and science reading comments. A positive correlation between comments and model summary scores was uncovered. Figure 5 shows the correlation between student voice comments and the model summary scores ($r = 0.36$, $p < 0.000$). As shown in Table 6, the correlation between varied comment frequencies and model summary scores ranged from 0.28 to 0.36, with student voice comments showing the highest degree of correlation.

To further explore the potential relationship between reading processes and modeling, we analyzed the correlation between the change in summary model scores and change in the number of student voice comments, from pre to post. See Figure 6. We included the pre-post change data from both classes, total $n = 42$. The correlation, $r = 0.23$ and $rho = 0.15$, was positive but did not reach 0.05 significance. A larger sample might shed light on the potential correlation between changes in students' reading and modeling. On the whole, these exploratory analyses suggest that there may be promise for the Text Based Investigation approach to simultaneously support student learning in science reading and modeling. The analysis also suggests the utility of the measurement tool for assessing student learning.

Figure 5: Voice Comments and Model Summary Score Relationship: Pre and Post Aggregate Study

For this analysis the data from the pre and post assessments of both classes (11th Grade Physiology Students and 9th Grade Biology Students) is aggregated. ($r = .36$, $N = 84$, $p < .000$)

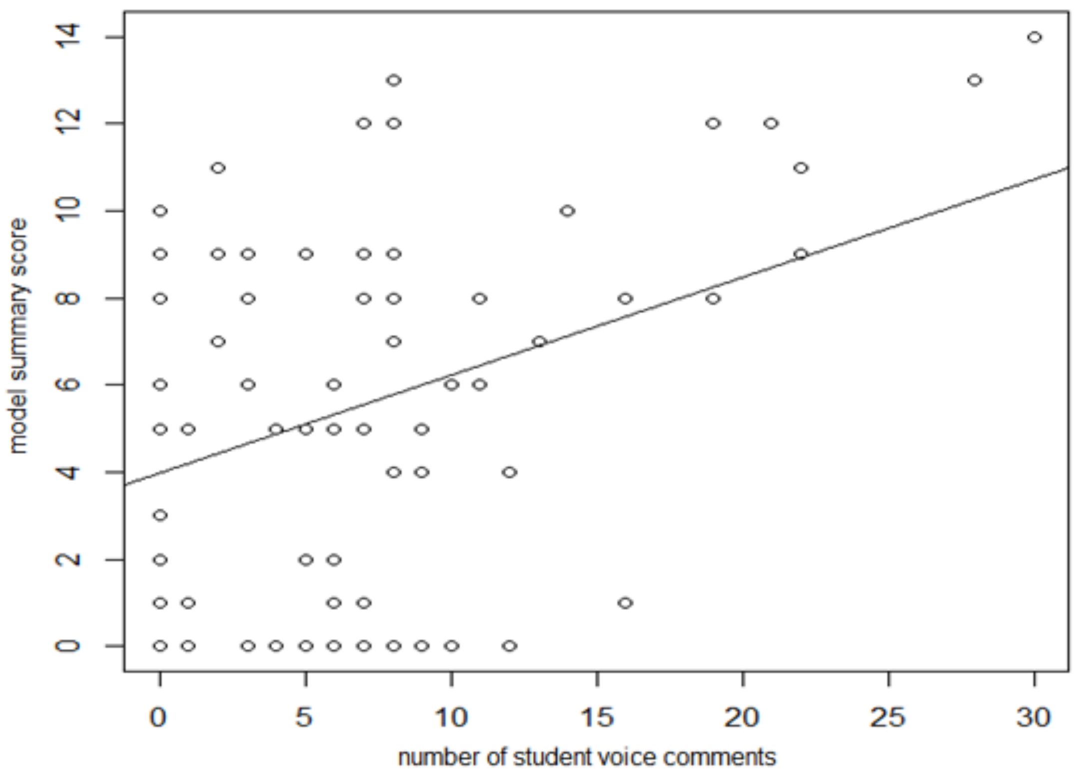
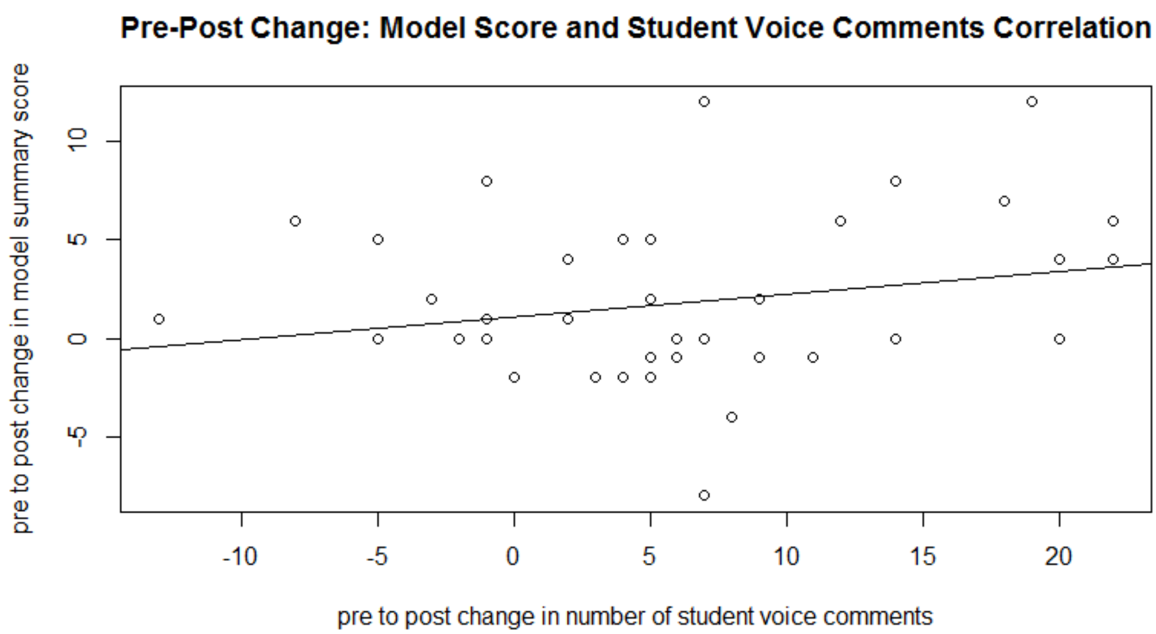


Table 6: Comparison of four comments types and Model Summary Score Relationship

	r	slope
Comments	0.33**	0.20**
Comments in Student Voice	0.36***	0.22***
Close Reading Processes	0.29**	0.21**
Science Close Reading Processes	0.28*	0.23*

Figure 6: Correlation between the Change Model Summary Score and change in number of student voice comments, pre to post.

For this analysis the data from the pre and post assessments of both classes (11th Grade Physiology Students and 9th Grade Biology Students) is aggregated. (r = .23, N = 42, p = .15)



READI Science Annotation Assessment Guide for Teachers

Finally, as part of the Science Assessment development, and as mentioned above, we developed a READI Science Annotation Assessment Guide to provide guidance for using student annotations in ongoing formative assessment of READI Science Learning Goals. In particular, we aimed to support teachers in noticing the kinds of intellectual work indicated by student annotations and to identify student strengths as well as opportunities for growth. (See Guide, Appendix B).

The front-end material in the Guide consists of the explanations of the five types of codes linked to student annotations: kind of annotation, association of annotation, degree of student voice, close reading process and science close reading process. This introduction also offers guidance as to how analysis of student annotation can serve as formative assessment of student's science reading for understanding. The body of the Annotation Assessment Guide consists of elaborations of the five types of codes as well as examples of each kind. The guide is grounded in the iterative analysis of student annotations in pre and post assessments related to the READI Text Based Investigation modules that were implemented in 2011-2013.

To test the utility of the Annotation Guide for teachers, we developed a series of professional learning experiences to support science teachers learning to use the annotation assessment guide. Science teachers in the California Teacher Inquiry Network brought student work from their own instruction to analyze, using the Guide. The professional learning experiences consisted of a series of inquiries, beginning with teachers reading a complex text in their subject area while annotating (Talking to the Text), then working with colleagues and the Guide to analyze their own annotations. We then engaged Teacher Inquiry Network teachers in using the annotation assessment guide to carry out the same set of inquiries to formatively assess their students' science reading with the texts they brought to the professional development session. We asked teachers to provide feedback, including suggestions of types of close reading or scientific close reading that were missing from the Guide, for refinement of the Guide. This process offered the opportunity to obtain teacher feedback and make observations of the Guide in use, to inform refinement of both the professional learning experiences and these material supports for assessment. Of interest, we used the Science Annotation Assessment Guide across subject areas, asking teachers of History/Social Studies and Literature to generate parallel codes to describe historical and literary close reading processes. The assessment activities, used over two sessions of the Inquiry Network, proved highly generative for teachers in all subject areas in coming to deeper understandings of the ways their disciplines shaped the process of reading for understanding.

Conclusions

To support teacher implementation as well as our learning about the promise of the Text Based Investigations approach, the Science Design Team of Project READI developed several pre post assessments and an assessment guide. Further, we iteratively developed ways of systematically scoring and coding the pre post assessments for evidence of student learning. We developed a system for scoring students' models based on our core constructs in science, and highlighting the roles of elements, interactions, and aggregate effects in accounting for (modeling) a science phenomenon. By analyzing pre/post assessment data from a single teacher's implementation of the MRSA module over two years, we found evidence that students were reading differently, and developing higher quality models of phenomena, from pre to post. In particular, the number of marks and comments increased from pre to post, and more importantly, the number of comments evidencing student voice, close reading processes and science reading processes increased. In addition, we found a correlation between students' comments and their model scores. These results suggest the promise of the MRSA investigation for student learning and the utility of the model scoring and annotation coding as an assessment tool.

Moreover, we believe the work carried out to develop a method of systematically coding student annotations on science texts for evidence of student learning constitutes a significant contribution to the field. Based on our explorations of these data as well as our previous studies (Greenleaf, et al., 2011), we believe that annotation mediates close reading. Student annotations are indicators of how students are reading and what they are learning. When elicited through Talking to the Text, students' comments and marks can serve as indicators of student's intellectual engagement with science text. Therefore, student annotations offer evidence for assessing students' progress on READI Science Student Learning Goals, particular Goal #1 (close reading of science texts) and to a lesser degree Goals #2-6. The iterative process of code definition, refinement, and application to varied data sets carried out by the team has led to a robust analytical approach to data of this kind.

The resulting annotation coding system can be used summatively on texts that are part of the pre/post assessments as well as formatively on texts that students are using during instruction. In fact, after sharing the initial code book with the Teacher Inquiry Network, many teachers worked with the codes in their classrooms, and with other teachers in their schools. In light of this, and in collaboration with the science design team teachers in California we, developed the teacher guide described above. The coding guides have also been used by teachers with their students as a rubric for self-assessment, a utility we did not foresee but which presents a space for further exploration.

References Cited

Afflerbach, P. P. (1990). The influence of prior knowledge on expert readers' main idea construction strategies. *Reading Research Quarterly,* 25, 31-46.

Afflerbach, P. (2000). Verbal reports and protocol analysis. In M. Kamil, P. Mosenthal, P. Pear-son, *&* R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 163-179). Mahwah, NJ:Erlbaum.

Afflerbach, P., & Cho, B. Y. (2009). Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading. *Handbook of research on reading comprehension*, 69-90.

Afflerbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behavior,* 16, 307-322.

American educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. American Educational Research Association.

Baker, E. L. (1994). Making performance assessment work: The road ahead. *Educational Leadership*, *51*(6).

Baker, E., Aschbacher, P., Niemi, D., & Sato, E. (2005). *CRESST Performance Assessment Models: Assessing Content Area Explanations.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. *Handbook of reading research*, *1*(353), V394.

Baker, E. L., Freeman, F., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 131-153). Englewood Cliffs, NJ: Prentice Hall.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5-31.

Boulter, C. J., & Gilbert, J. K. (1995). Argument and science education. In P. J. M. Costello & S. Mitchell (Eds.), *Competing and consensual voices: The theory and practice of argumentation*. Clevedon: Multilingual Matters.

Braten, I., & Stromso, H. (2003). A longitudinal think-aloud study of spontaneous strategicprocessing (luring the reading of multiple expository texts. *Reading and Writing: An Inter-disciplinary Journal, 16,* 19.5-218

Charmaz, K. (2008a). Constructionism and the grounded theory method. *Handbook of constructionist research*, 397-412.

Charmaz, K. (2008b). Grounded theory as an emergent method. *Handbook of emergent methods*, 155-170.

Chung, G. K. W. K., Harmon, T. C., & Baker, E. L. (2001). The impact of a simulation-based learning design project on student learning. *IEEE Transactions on Education, 44,* 390398.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*(1), 1-53.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *102*(3), 687.

Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse. *Studies in Science Education*, 38, 39-72.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the use of Toulmin?s Argument Pattern in studying science discourse. *Science Education, 88*(6), pp.915-933.

Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review, 87,* 215-253.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA:MIT Press. (Original work published 1984)

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist,* 34, 906-911.

Gallagher, M. C., & Pearson, P. D. (1989). Discussion, comprehension, and knowledge acquisition in content area classrooms. *Center for the Study of Reading Technical Report; no. 480*.

Glaser, B. G. (2011). *Getting out of the data: Grounded theory conceptualization*. Sociology Press.

Goldman, S. R., & Durán, R. P. (1988). Answering questions from oceanography texts: Learner, task, and text characteristics. *Discourse Processes*, *11*(4), 373-412.

Herman, J. L. (1998, December). The state of performance assessments. *The School Administrator*, pp. 17-18, 20-22.

Herman, J. L., Baker, E. L., & Linn, R. L. (2004). Accountability systems in support of student learning: Moving to the next generation. *CRESSTLINE*, Los Angeles, CA: CRESST.

Herman, J. L., Baker, E. L., & Linn, R. L. (2004, spring). Accountability systems in support of student learning: Moving to the next generation. CRESST Line. Retrieved from: http://www.cse.ucla.edu/ products/newsletters/CLspring2004.pdf

Hruschka, D. J., Schwartz, D., John, D. C. S., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, *16*(3), 307-331.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing inreading. *Cognitive Psychology,* 6,293-323.

Linn, R., Baker, E., & Dunbar, S. (1991) Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

Martin, K., & Miller, E. (1988). Storytelling and science. *Language Arts*, *65*(3), 255-259.

McConkie, G. W. (1997). Eye movement contingent display control: Personal reflections andcomments, *Scientific Studies of Reading, 4,* 303-.3I6.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative data analysis: A methods sourcebook*. SAGE Publications, Incorporated.

National Research Council. 2001. *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment. Pelligrino, J., Chudowsky, N., and Glaser, R., editors. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Paris, S., & Flukes, J. (2005). Assessing children's metacognition about strategic reading. In S.E. Israel, C. C. Block, K. L. Bauserman, & K. Kinnucan-Welsch (Eds.), *Metacognition in literacy learning: Theory, assessment, instruction, and professional development* (pp.121-139). Mahwah, NJ: Erlbaum.

Pontecorvo, C. (1987). Discussing for reasoning: The role of argument in knowledge construction. *Learning and instruction: European research in an international context*, *1*, 239-250.

Prain, V., & Hand, B. (1996). Writing for learning in secondary science: Rethinking practices. *Teaching and Teacher Education*, *12*(6), 609-626.

Schoenbach, R., Greenleaf, C. & Murphy, L. (2012). *Reading for understanding: How Reading Apprenticeship improves disciplinary learning in secondary and college classrooms*. San Francisco, CA: Wiley.

Schwarz, B. B., Neuman, Y., Gil, J., & Ilya, M. (2003). Construction of collective and individual knowledge in argumentative activity. *The Journal of the Learning Sciences*, *12*(2), 219-256.

Silver, D., Hansen, M., Herman, J., Silk, Y., & Greenleaf, C.L. (2011). IES integrated learning assessment final report. (CRESST Report 788). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *The Phi Delta Kappan*, *87*(4), 324-328.

VanDijk, T., & Kintsch, W. (1983). *Strategies for discourse comprehension.* New York: Aca-demic Press.

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and learning*, *1*(1), 3-14.

Wigfield, A., & Guthrie, J. T. (2000). Engagement and motivation in reading. *Handbook of reading research*, *3*, 403-422.

Wolfe, M., & Goldman, S. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction, 23,* 467-502.

date ___________________________

name __________________________

teacher _________________________

Introduction

Malaria is a serious disease in many parts of the world. It has many "causes" linked in a chain of events. Scientists try to prevent the disease by breaking links in the chain.

Task

1. Read the texts on the following pages and make notes in the margins about your reading, thinking and problem solving processes.

2. After you have read the texts, respond to the following, using information from your reading:

   A. Use the information in texts two, three and four to create a model, using visuals and words, that explains how malaria could cause millions of deaths each year in Africa. (You may add to the model in text 4, but yours may also look different).

   B. Based on what you know now, explain what might be done at different points to stop the transmission of malaria and use evidence from your reading to explain why these might work.

SPACE WILL BE PROVIDED AT THE END FOR YOU TO COMPLETE YOUR RESPONSES.

TEXT ONE

Introduction: The Malaria Problem

Malaria causes fever, joint pain, vomiting, seizures and can lead to brain damage and death, especially in children. On World Malaria Day in 2009, former President Clinton explained that "malaria was eliminated in the United States over a half a century ago, yet more than 1 million people around the world still die from the disease each year, making it one of the most pressing health challenges the world faces today."[1] According to the World Health Organization's 2011 report, there were 216 million cases of malaria and an estimated 655,000 deaths in 2010. [2] Most deaths occur among children living in Africa where a child dies every minute of malaria and the disease accounts for approximately 22% of all childhood deaths. The Clinton Foundation states, "despite … attention from the global community in recent years, the majority of African families are not benefitting from the tools necessary to stop malaria, such as bed nets and effective medicines, because of a lack of access or efficient use." [3]

1. April 24, 2009 | New York, Address given by Bill Clinton on ***World Malaria Day***
2. http://www.who.int/malaria/world_malaria_report_2011/en/
3. *2. http://www.clintonfoundation.org/what-we-do/clinton-health-access-initiative/our-approach/access-programs/malaria*

TEXT TWO

How is malaria spread?

Malaria is caused by *Plasmodium* bacterium. These parasites infect successively two different hosts: humans and female *Anopheles* mosquitoes.

The parasites are transmitted to people who are bitten by infected female *Anopheles* mosquitoes. In humans, *Plasmodium* multiplies in the liver and then invades the red blood cells. Successive generations of parasites grow inside the red cells and destroy them, releasing daughter parasites that continue the cycle by invading other red blood cells. These blood-stage parasites, called "gametocytes" (G. gamete + kytos, cell) cause the symptoms of malaria, which begin 6-10 days after infection.

When a female *Anopheles* mosquito bites an infected human, she takes the person's infected blood for a meal. During this meal, if gametocytes are picked up by the female mosquito, they may start another, different cycle of growth in the mosquito's gut. After 10-18 days, the parasites develop the infectious stage, called "sporozoites" (G. *sporos*, seed + *zōon*, animal), which reproduce in the mosquito's salivary glands.

When this infected *Anopheles* mosquito bites another human, the sporozoites are injected into the human's blood along with the mosquito's saliva. Thus the mosquito acts as a vector, transmitting the disease-causing parasite from one human to another.

Adapted from http://www.cdc.gov/malaria/about/biology/index.html and http://www.who.int/mediacentre/factsheets/fs094/en/

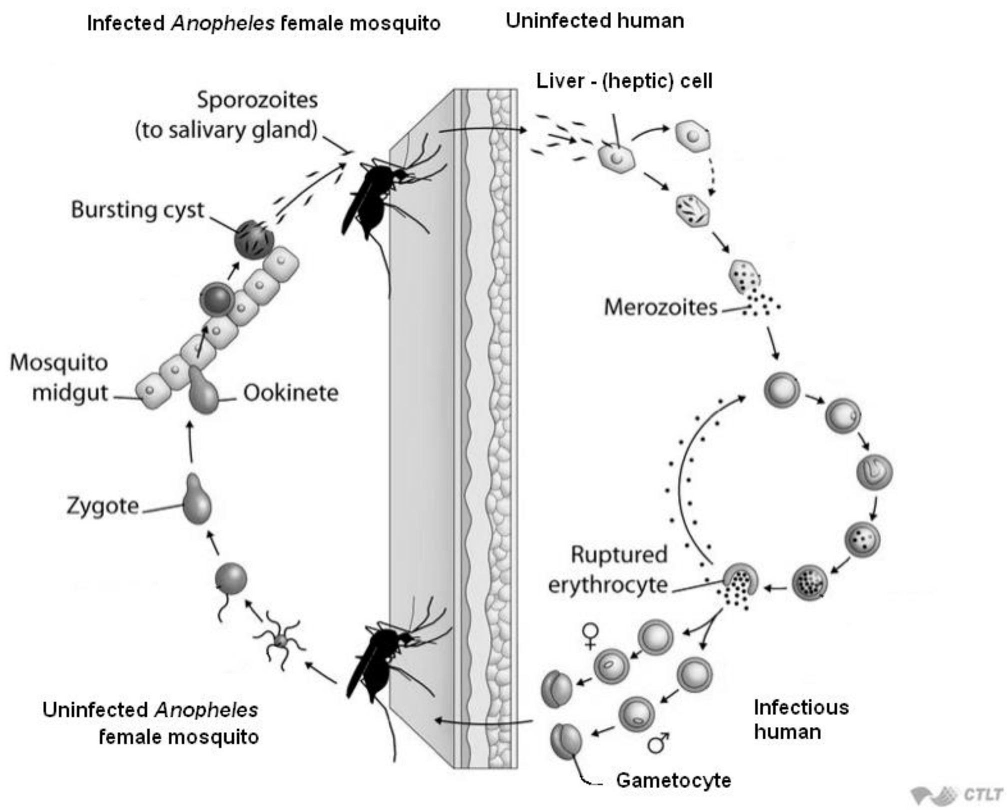
TEXT THREE

Ecology of Malaria

For malaria transmission to occur, three conditions must exist:

- Female *Anopheles* mosquitoes must be present, which are in contact with humans, and in which the parasites can complete half of their life cycle
- Humans must be present, who are in contact with the *Anopheles* mosquitoes, and in whom the malaria parasites can complete half of their life cycle
- Malaria parasites must be present.

TEXT FOUR



Life Cycle of Malaria parasite – Adapted from: http://ocw.jhsph.edu

2A. Use the information in texts two, three and four to create a model, using visuals and words, that explains how malaria could cause millions of deaths each year in Africa. (You may add to the model in text 4, but yours may also look different).

2B. Based on what you know now, explain what might be done at different points to stop the transmission of malaria and use evidence from your reading to explain why these might work.

Date ____________________________

Name ___________________________

Teacher __________________________

Introduction

Science is about understanding changes in the natural world and developing solutions to challenges. In this learning experience you will do both. The topic is about how head lice are becoming pesticide resistant.

Task

1. Read the texts on the following pages. Make notes in the margins about your reading, thinking and problem solving processes.

2. Construct a scientific model that explains how head lice have, over time, become commonly pesticide resistant. Use the information in the texts. Use both visuals and words in the model.

3. Explain why the scientific model you constructed is a good model.

Space is provided after the texts for you to complete your responses.

Text 1

Head Lice: Treating Parasites That Go to Your Head

If simply the thought of head lice makes you feel a little itchy, imagine how those affected by a head lice infestation feel when they learn what has been creeping around them and causing symptoms such as intense itching or irritated scalps. Even less comforting is the fact that lice parasitic infestations are on the rise.

According to the Centers for Disease Control and Prevention (CDC), 6 million to 12 million people a year suffer from head lice infestation, and it is estimated that more than $100 million is spent annually to combat this problem. Head lice tend to affect younger, school-aged children, but teens also can get them; and girls tend to get head lice more than boys due to their longer hairstyles.

Upon close examination, the most common signs of head lice include: an itchy scalp, red bumps, small skin tears and evidence of the egg casings (or nits) attached to the hair shafts, as well as live lice. The nits hatch within seven to 10 days and live about 30 days, during which they reproduce to spread the infestation.

Common, over-the-counter, topical treatments for head lice include chemical pesticides, such as permethrin and synthetic pyrethroids. However, resistance to standard pyrethroid treatments has become widespread and is well documented in the United States, the United Kingdom, Israel and the Czech Republic. In one study, patients using both permethrin and synthetic pyrethroids for 10 minutes and then washing it out (the standard treatment), only killed 5 to 7 percent of the head lice.

To help children avoid head lice, they should not share combs, brushes, hats, barrettes or any other personal care items with anyone else, regardless of whether they have lice or not. Also, it’s important to examine everyone in the household when there is a case of head lice, just to be sure that the bugs have not been transmitted.

Sources: Adapted from: American Academy of Dermatology (Academy) http://www.news-medical.net/news/20100305/Dermatologist-addresses-common-parasitic-infestations-and-lates t-treatment-options-at-Academy-Annual-Meeting.asp

Excerpts from: http://www.news-medical.net/health/What-is-head-lice.aspx.

Text 2

Head Lice Resistance to Pesticides

Pharmacists and doctors have relied on chemical pesticides to kill head lice. Some of the chemical pesticides are over the counter products, and others are prescriptions. Their purpose is to affect the nervous system of the lice, to disrupt their ability to move and eat, or to kill bacteria that lives in their gut, which provides nutrients to them. If the bacteria die, the lice die. However, these may now be unwise choices for treatment, in light of potential lice resistance to these chemicals.

Resistance is the development of mechanisms to survive potentially deadly onslaughts. Many organisms that can cause disease have become resistant to many antibiotics. It should be no surprise that rapidly reproducing insects, such as head lice are developing resistance to the pesticides used to kill them.

Resistance has become a 'growing problem' since the 1970's, as patients resort to using multiple treatments of chemical pesticides, which can also potentially and needlessly expose children to toxic chemicals. By 1999, several of the chemical pesticides were reported as virtually useless in England, while, in the U.S. 81% of patients using pyrethrin against head lice could not get rid of the lice. Also, more than 58% of people in the U.S. who treat against lice without success the first time have treated themselves with higher doses of chemical pesticides and have done it more frequently. Resistance seems to be affected by:

- how large is the spread of infestation,
- the type of chemical pesticides used,
- the variety of mechanisms by which lice resist chemical pesticides, and
- the pattern of use of chemical pesticides in different countries.

Scientists hypothesize that there are various resistance mechanisms that head lice develop, such as changes that take place in the amino acids of cells in the nervous system of the lice, so the chemical pesticides' purpose is no longer effective; or by slowing down the absorption and metabolism of the pesticides into their bodies, allowing lice to live longer and to lay eggs; and by successful mutations in their DNA being passed on to succeeding generations for survival.

Sources:

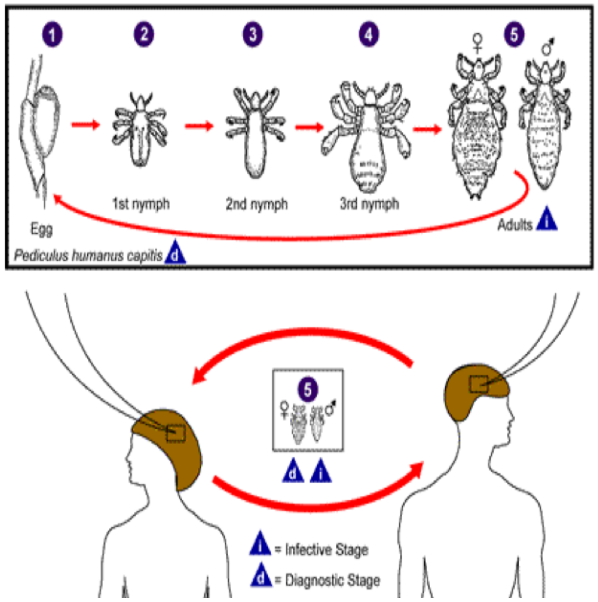
Pray, W., Head Lice: New Approaches May Help Overcome Pediculicide Resistance, *US Pharm*. 2010; 35(3):10-15. Retrieved from:
http://www.uspharmacist.com/content/c/19874/
http://www.ncbi.nlm.nih.gov/pubmed/14651472

Text 3

Head Lice Life Cycle

Pediculus humanus capitis, the head louse, is an insect parasite that lives only on the outside of human hosts, particularly on hair close to the scalp (1 mm).

The adult female lice lay around 7 to 10 eggs a day and attach them to the hair using a glue-like, water-insoluble substance. Most eggs are laid at night and can survive for more than 2 weeks. The common site for these eggs or nits is the back of the head or back of the ears. The heat and the moisture of the human head help to incubate the eggs. Because people have a constant body temperature, female lice reproduce continuously throughout the year.



Each adult louse lives for around 30 days. Within 7 to 10 days the nymph emerges from the eggs and feeds on blood from the scalp. Another 7 to 10 days and three moulting stages makes the nymphs adult lice. New adult females start laying eggs soon after day 10. Consequently, the total life span of a head louse from egg through adult averages about 25 days.

To survive, a newly hatched head louse must have a blood meal within minutes of birth. Each louse takes several meals of blood each day and die if they are removed from the head for more than 2 days.

At any given time a person with an infestation has no more than 10 to 12 live head lice but over a 100 eggs or nits.

Source adapted from: from URL: http://www.news-medical.net/health/What-is-head-lice.aspx

2. Construct a scientific model that explains how head lice have, over time, become commonly pesticide resistant. Use the information in the texts. Use both visuals and words in the model.

3. Explain why the scientific model you constructed is a good model.

Date ___________________________

Name __________________________

Introduction

Scientists have collected data that prove that Earth is warming up. Scientific models and experiments show that increasing amounts of carbon in the atmosphere causes this warming. One of the most important scientific questions today is "How are humans impacting the carbon cycle and what can be done about it?"

In your folder are four texts that will help you understand the carbon cycle and how humans are impacting it.

Your Tasks

1. Read the texts in your folder and write notes in the margins about your reading, thinking and problem solving processes.

2. After you have read the texts, respond to the following:

    Use information from the texts to create a detailed model, using visuals (pictures, diagrams, graphs, etc…) and words that explain how humans impact the temperature of the Earth.

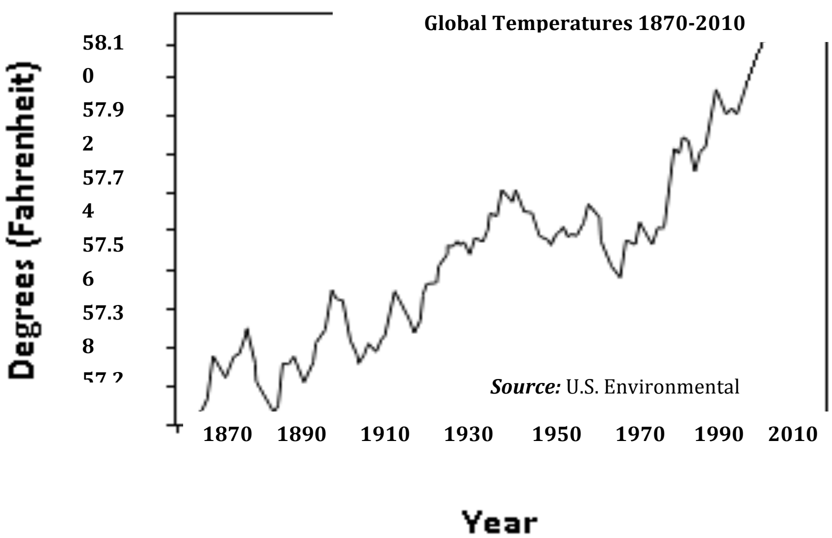
- You can read the texts in any order you wish.
- The information in the texts will help you create your model.
- When you have finished reading and writing notes on the texts take the blank paper that is in your folder and use it to create your model.
- You can look at the texts and your notes when you are creating your model.

Name __________________________

Changes in Global Temperatures

The temperature of the land and the oceans is measured by weather stations all over the Earth. Scientists have access to all of this temperature information. At the end of the year, they take the average of all of these temperatures. This average is called the global temperature of the Earth. Small changes in the average global temperature create big problems for living things. A recent study by the Royal Society of biological sciences found that warmer temperatures are related to higher extinction rates. *

Scientists have recorded temperatures around the globe since 1880. The graph below is a visual model made using this data. It shows the average global temperature of the Earth from 1880 to 2010.



* doi: 10.1098/rspb.2007.1302 *Proc. R. Soc. B 7 January 2008 vol. 275 no. 1630* 47-53

Name ________________________

Carbon Balance

Carbon is the backbone of life on Earth. We are made of carbon, we eat carbon, and our economies, our homes, our transportation all use carbon. On Earth, carbon is stored in the ocean, the atmosphere, in living things, and in the earth as rocks, soil, and fossil fuels. We call places that store carbon "sinks."

Carbon moves between sinks through the carbon cycle. Carbon can be moved out of one sink and into another, but it never gets destroyed or goes away. For example, when people use gasoline in their cars, they are moving carbon from the lithosphere to the atmosphere.

Carbon dioxide is gas made up of one atom of carbon and two atoms of oxygen. Carbon dioxide traps heat in the atmosphere. Without it and other carbon gases, Earth would be a frozen world. When there is more carbon in the atmosphere, the Earth warms.

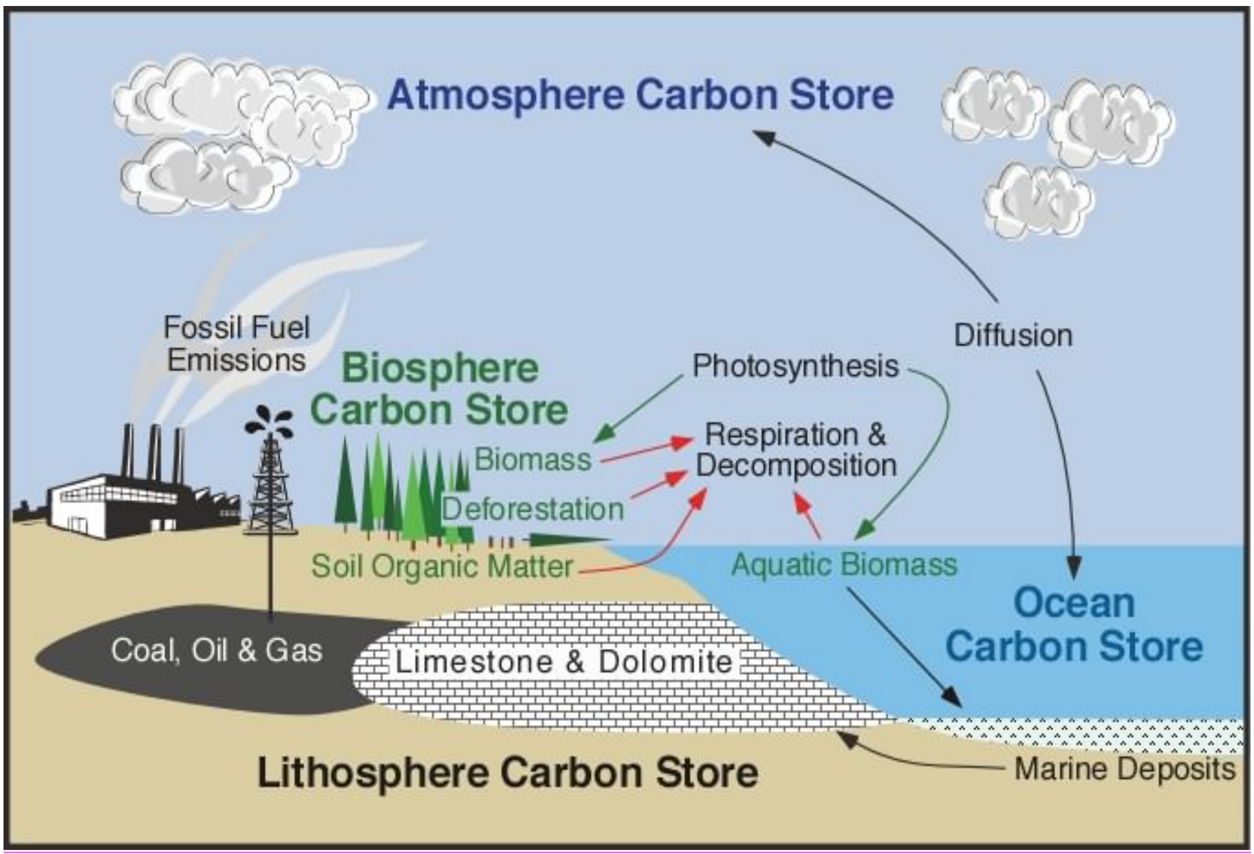
Name ___________________________

Carbon Sinks

The diagram below is a visual model of the major carbon sinks, or "stores".

Carbon is found

(1) in living things in the ***biosphere;***
(2) as the gas in the ***atmosphere;***
(3) in soils in the ***geosphere;***
(4) as fossil fuels and rock in the ***lithosphere***
(5) in the oceans, or ***hydrosphere.***



http://www.physicalgeography.net/fundamentals/9r.html

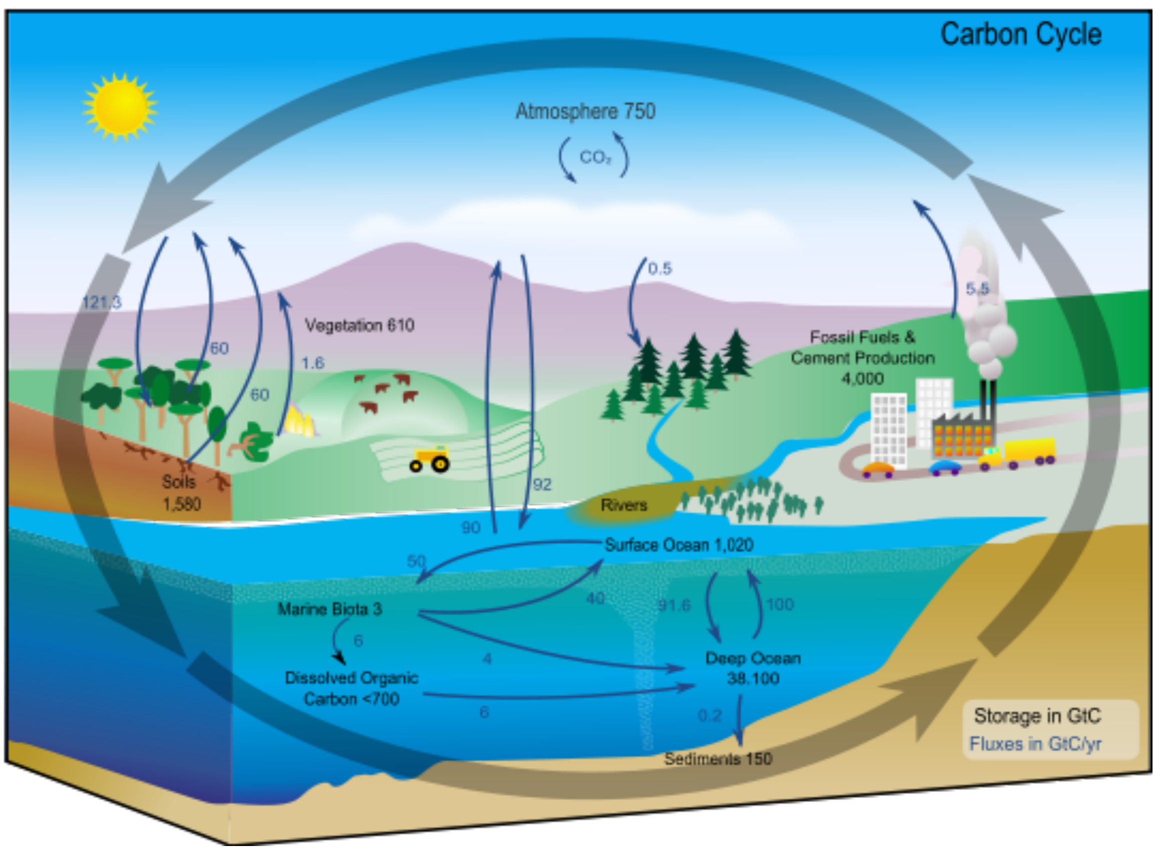
Name ___________________________

The Carbon Cycle

The carbon cycle exchanges carbon among the biosphere, lithosphere, geosphere, hydrosphere, and atmosphere of the Earth. It is one of the most important cycles of the Earth and allows for carbon to be recycled and reused by all living things.

Over billions of years, the carbon cycle seems to maintain a balance between the atmosphere and the other four sinks. This balance has kept Earth's temperature relatively stable and capable of supporting life, unlike any other planet in our solar system. But the ways humans use carbon has moved carbon from a few sinks into the atmosphere. This is changing the balance between sinks and impacting Earth's temperature.

Visual Model of the Carbon Cycle



The black numbers show how much carbon is stored in different sinks, in billions of tons ("GtC" stands for gigatons of carbon).

The grey numbers indicate how much carbon moves between reservoirs each year.
http://en.wikipedia.org/wiki/File:Carbon_cycle-cute_diagram.svg

Appendix B

Annotations Code Book: Developing Assessments of Students' Science Reading Processes and Explanatory Models to Measure Learning Outcomes from Instruction with Text-Based Investigation Modules

Code	Definitions	Guidelines
Type 1: Type of Annotation		
Marks	Use for each type of mark - these are exclusive of one another: underlines, highlights, circling/boxings	*If the underline reaches to the end of one row of text and resumes on the beginning of the next, count it as one mark. *If the line stops for a punctuation mark, count it as one mark. *If the underline stops skips a word and then resumes, count them as distinct underlines. *If a string of words are each underlined whether or not there is a gap in the underlines, the string is counted as one mark.
Connecting mark	Between two or more places in the text; between text and annotation; between marked text and annotation; between annotation and annotation; between any two or more lines of text; between any two parts of text or between other marks or writing	These marks may be long or short. Can include arrows between text, brackets, carots, parentheses
Symbols	Symbols associated/near a text. Can be floating punctuation or other non alpha numeric items	*Code punctuation marks that are not attached to words also other non alphanumeric characters, such as stars, asterisk and other scientific symbols *Do Not Code doodles or pictures unrelated to topic in the margins
Comments (all writing in the margins should get one of the codes below)		
Single Words	May include a single word, a single word and a punctuation, or a short list of single words separated by slashes or commas	an article + word (a lot, the __) and two words connected by a hyphen are considered one word.
Sentences and sentence fragments	Sentences or sentence fragments making an assertion	Score an assertion followed by a … 'right?' … or 'correct?' as a statement. Albeit a tentative statement. (a wondering thought connected to the statement)
Questions	Sentence or sentence fragment posing a question MUST BE co-coded with either Single Word, or Sentence code	May be signaled by interrogatives, verb structure, or question mark.
Combinations of Comments Paired with Marks	Use this code to indicate instances where it is clear that the writing in the margin is associated with marks in the text and therefore should be understood together when deciding how to apply type 4 & 5 codes	Code if: *Comment is directly attached (by a connecting mark to marked text) *Comment has a clear referent to a specific sentence or phrase that is marked in the text. *Comment is directly attached by a connecting mark to unmarked text AND has a clear referent to a specific sentence or phrase. Referents could be a replicated word, summary, or synonym to the proximal text *Use low inference as to how the writing is connected to the marked text

Type 2: Association of Annotations		
Associated with text	Annotations made on, linked to or nearby verbal text	
Associated with visuals/diagrams	Annotations made on, linked to or nearby visuals/diagrams or the verbal elements associated with the visual. Including any on captions, legends, labels etc.	
Type 3: Occurrence of student voice in annotation. CODE FOR ALLL COMMENTS		
Verbatim Excerpts / Author's Voice	Verbatim/near verbatim or excerpts from text. "Near verbatim" includes examples where student writes a selection of text but drops words or phrases, adds or drops articles, engages in single synonym word replacement	Include comments that are verbatim/near verbatim excerpts from the text, or condensed from the text - primarily authors words and structure, some words/phrases from text omitted
Comment foregrounding the student's voice/thinking	A comment expressing a students thoughts about the text, contexts or themselves. Mark student voice if a paraphrase/summary of text features at least 25% of words changed or there is a substantial change in sentence structure from the original	Almost all comments not coded as 'Author's voice' can be coded 'student voice.' Strong indicators are words, ideas, phrases in comments that are not in the text. May include use of punctuation marks (?, !) or other symbols.
Inaccurate representation	Comment expresses an idea not substantiated by the text nor our knowledge of the natural world	MUST co-code with Student voice comment
Level 4: Comments Indicating Close Reading Processes Can be co-coded with either student or author voice IF coded with type 1, comment & mark associated, use associated text to help assess what code type to apply Can be multiply coded		
Identifying unknown vocabulary	Comment expresses puzzlement regarding a word, term or symbol that is clearly unknown to the student	If comment is simply a "?", do not score as identifying unknown vocabulary because it is ambiguous whether student is expressing curiosity or confusion.
Defining vocabulary	Comment asserts/advances a meaning for an word, term or symbol	*Comment identifies a roadblock and expresses a meaning OR comment explicitly and overtly claims a meaning for a word/term, otherwise score as a paraphrase *May show identifying root words, looking ahead for meaning in the text for a definition, etc.
Paraphrasing and summarizing	Annotation is a paraphrase or summary (beyond the word level)	*Comment explicitly/overtly signals that it is a paraphrase/summary *OR (low bar criteria exception) comment IS an accurate paraphrase or summary. * DO NOT CODE if comment is a verbatim recopying of text- this is not considered a paraphrase nor a summary. Use definition of "verbatim" from type 3.

Making connections codes caveat: These codes are not meant to describe a student's internal thinking and knowing processes, rather they describe how a student is working with the texts presented. Therefore the aim is to take a very literal read of the text – and if the comment does not contain information within the text set, use Making connections: information outside of text set. This acknowledges that we have no way to know what knowledge the student brought to the text set from prior experiences		
Making Connections: information outside text set	Comment displays student connecting an idea in the text to idea(s) not present in any of the texts in the set	*Low bar *Can Co-code with any other type 4 code except other making connections codes
Making Connections: within text	Comment displays student making connections between two or more ideas present within the text	*Connection must be between ideas in two different sentences *Connecting lines can qualify as connections between ideas in the text*Can Co-code with any other type 4 code except other making connections codes
Making Connections: text to text	Comment displays student making connections between two or more ideas between two or more texts	*Connection must be between ideas in two different texts in the text set *High bar- needs to be clearly from the other text in the set, OTHERWISE code Making connections: outside text set *Can Co-code with any other type 4 code except other making connections codes
Predicting / Inferencing	Comment makes a prediction / hypothesis / inference / tentative claim	*Can be co-coded with paraphrase/summary- i.e.: if an inference is imbedded inside a summary *Must assert an outcome/connection/position/opinion not explicitly described in the text. *Could also be tentative claim. *Must display students expanding on the ideas text in any form (if comment does not extend ideas in text, code for paraphrase/summary)
Inquiry Questioning for knowledge building	Comment poses an inquiry question with the purpose of knowledge building.	*An inquiry question is a question posed that suggests comprehension rather than confusion with the meaning of the text AND asks for information beyond what is given explicitly in the text. *The question can posed in the form of a statement ("I wonder if...") *If comment expresses confusion or implies a misunderstanding, scores are identifying/expressing confusion. *If it simply asks for the meaning of word, code as identifying unknown vocab.
Metacognitive Comment	Comment shows student awareness of their own thinking process, or makes an evaluative statement regarding their own knowledge	*** use this code broadly to help define. It would be helpful to me to have more clarity on this construct- metacognitive could include both comments about self-monitoring of understanding as well as other things like self-monitoring of emotions, etc.**

Type 5: Indicators of Scientific Reading Processes		
Comments on Science in text	Indicates that student displays a SCIENCE reading stance (in addition to close reading stance) or engaged with science in the text that is NOT part of the explanatory model or the phenomena in focus	*Code if comment indicates student is situating themselves as a reader of a science text* Can co- code with type 5 cross cutting concepts or scientific argument arguments* Use for things like paying attention to text features that are science specific, attention to size and quantity that is not at the level of the system or phenomena, *NEVER co-code to other Nascent Modeling codes
Nascent Modeling		
Comments on the phenomena or specific elements of the explanatory model	Indicates student is showing thinking about the topic of the text set. We are using phenomena as a proxy for topic. Indicates student is showing thinking or engagement with a component or of the explanatory or causal model. Not limited to an the understanding of that component as conscribed by the explanatory or causal model design.	*Code if the comment evidences a science reading stance AND is on the text topic *DO NOT code if comment is on an element of the explanatory or causal model (see Type 5 modeling codes)
Comments on the relationships in the explanatory model (relationship between elements)	Indicates student is showing thinking or engagement with the links or relationships between elements in the explanatory model for the phenomena.	*Code for any comment or combination comment/mark that expresses attention to the ways in which two or more elements of the explanatory or causal model are understood to interact in the causal or explanatory model *DOES NOT need to be a causal relationship *Can Co-code with type 5 cross cutting concepts or scientific arguments *NEVER co code with other nascent modeling codes
Modeling		
Generate a course of action based on the explanatory model	Annotation expresses a course of action about the problem and implicitly/explicitly indicates the basis from the explanatory model.	
Identifying Cross-cutting / unifying concepts in science	Annotation expressly note a connection to a cross cutting/ unifying concept in science.	*If evolution is biological evolution, score as schema development, or as science topic knowledge. *Only 3 of the cross cutting concepts are strongly evoked by module and assessments text sets (#3. Cause and Effect, #4 Scale, proportion and quantity, and #7 Stability and Change) Of these #3 and #4 are addressed by the nascent modeling codes. So on Attending to scale is pulled out as a code of its own.

Attending to Scale Proportion and Quantity	A comment about the scale, quantity or proportion of the system and associated with text evidence about the scale, quantity or proportion.	*Must be on a phenomena or system (not "this paragraph has a lot of words") *Must be about measurable items (e.g., "seriousness" is not measurable) *Comment can be qualitative or quantitative in nature *Text can be qualitative or quantitative, but topic must be measurable
SUPPORTING AN ASSERTION: Annotation offers evidence and/or principled reasoning for own claim)	[Undefined due to lack of sufficient evidence in the data]	This must be double scored for previous argument codes.
Generating a hypothetical/mental Causal Model	[Undefined due to lack of sufficient evidence in the data]	
Stability and change		
Attention to Argument		
Attending to ARGUMENTS: Noting in texts arguments advancing models and explanations (ungrounded in pre/post, no examples yet)	A comment associated with a portion of text that is an aspect of an argument (incl. claim, warrant, backing, etc.).	*If function of the comment is clarifying, score a clarifying arguments. *If function of the comment is evaluative, score as evaluating arguments. *Text associated with comment must be an argument (Claim with kind of support). * If it is a claim without evidence, consider if it is an explanation.
EVALUATING ARGUMENTS Evaluating-questioning of science argument.	Evaluating the model as it is in the text based on own schema (not evaluating the authors argument, if any) Asks questions about the nature of the model/phenomena, its processes and interactions, based on own schema.	*May be related to recommendation/ solution *Annotation reflects a skeptic stance.