

ABSTRACT

Exploratory Development of a Tool to Measure and Analyze Structural Knowledge (ASK) in High School Biology Courses

Research Program Topic: Education Technology “NCER-EdTech”;

Application Goals: Development and Innovation “Development”

Purpose: This three year project will develop and evaluate an online enhanced version of existing software (*ALA-Reader*) that converts students’ high school biology essays into network graphs. These graphs will encourage metacognitive reflection to support improved content knowledge, reading comprehension, and self-regulated learning.

Setting: Research will take place in Years 1 through 3 in a public virtual high school and with high risk students in a traditional high school. Pilot research of promise will be conducted in Year 3 in these schools and in two matched control schools.

Population: Over the 3 years of the project, the students in this study will include a diverse group of from 800 – 1,000 high school biology students from Pennsylvania.

Intervention: The software will run in a web browser from the internet. Students will be assigned a biology content writing task weekly and will use the software to submit their writing assignments which will immediately convert their work into a network graph. Students will then be asked to reflect on this representation and on the content of their writing. Meta-analysis of writing-to-learn has reported an effect size of .52 when feedback is provided, and an effect size of .77 when metacognitive reflection is used.

Development Design and Methods: Working with high school biology teachers, in Year 1 the investigators will develop and pilot writing prompts aligned to course readings and collect student essays to automate the software through an iterative design process. In Year 2, teachers in this school and another school will use the writing prompts embedded in the software. A pilot study in Year 3 will make treatment-to-control comparisons and year-to-year comparisons to determine the evidence of promise.

Key Measures and Outcomes: Key measures of the validity of the software are (1) teachers’ and students’ acceptance level and use of the software, (2) expert content analysis of the students’ essays compared to the group average network graphs to determine how well the network graphs align with the students’ understanding of the content, and (3) students’ success on the state standardized tests for biology and reading comprehension compared with those who did not use the software.

Data Analytic Strategy: In Years 1 through 3, the project will use observation, interview, and survey approaches to determine if the software measures reliable and valid. State biology and reading test data will be collected and analyzed all three years for trend analysis and will serve as covariates in Year 3. The pilot study in Year 3 will use a quasi-experimental design with two treatment and two control schools (with about 140 students per site per year) to determine the effects of the intervention on state biology and reading tests as a measure of promise. Site variability will be controlled to some degree using covariates in the analysis. Mediating and moderating factors will be also considered including ability level, students’ achievement goal orientation, epistemological beliefs, beliefs about writing, and self-regulated learning.

**Exploratory Development of a Tool to Measure and Analyze Structural
Knowledge (ASK) in High School Biology Courses
The Pennsylvania State University
Budget Narrative**

Senior/Key Personnel
Salaries and Wages

Principal Investigator, Roy Clariana – Academic 30% (2.70 months); Summer 33.3% (1.00 months). The PI will oversee all phases of the development project and Year 3 pilot investigation including directing the graduate students' work tasks, the design and development of the online tool, conduct of all project meetings, direction of the usability trials, liaison with schools, and budget oversight. He will be the primary research protecting records and conducting analysis of collected data and will be the primary author for reports and publications.

Co-principal investigator, Adrian Barb – Academic 11% (1 month). The Co-PI will primarily conduct the actual development of the online tool and will maintain the server and databases of essays and scores. He will also participate in school site visits and will author publications.

Graduate Assistants – Grade 12 half time (20 hours per week), fall and spring for Year 1 (2 GAs) and Years 2 & 3 (4 GAs). The Graduate Assistants hired for Years 1 and 2 will have substantial biology content knowledge in order to analyze and recommend curriculum aligned writing prompts, to communicate knowledgably with the biology teachers in schools; Year 1 and some of Year 2 tasks include scoring biology essays weekly. Year 3 Graduate Assistants tasks will primary involve data collection at school sites including interviews and surveys. Their time includes both technical and project management functions such as data entry, and review and synthesis of the literature.

For project time occurring after June 30 of any given year, the salaries have been adjusted at the approved University rate of 3.0% per year each July 1.

Fringe Benefits – Fringe benefits are computed using the negotiated rates of 31.8% applicable to Category I Salaries and 14.1% applicable to Category II Graduate Assistants (GAs) for the current fiscal year July 1, 2011, through June 30, 2012. If this proposal is funded, the rates quoted above shall, at the time of funding, be subject to adjustment for any period subsequent to June 30, 2012, if superseding Government approved rates have been established. The fringe benefit rates are negotiated and approved by the Office of Naval Research, Penn State's cognizant federal agency.

Travel

Travel support is requested for three purposes: (1) team site visits to participating schools, (2) travel to the American Education Research Association (AERA) national conference to present findings and to attend project-related sessions in order to improve the project, and (3) annual meetings in Washington, DC with the project officer.

All travel will be in accordance with University travel regulations and mileage will be charged at the current rate on the date of travel. Travel estimates are based on costs that were incurred on previous projects of a similar nature for federal and state agencies.

School Visits

- In Year 1, 11 visits to the 21CCCS by four investigators including mileage, food, and hotel accommodation where appropriate - \$3,960
- In Year 2, 8 visits by six investigators to 21CCCS and to SCASD including mileage, food, and hotel accommodation where appropriate - \$8,640
- In Year 3, 8 visits by six investigators to 21CCCS, to SCASD, and to 2 control schools including mileage, food, and hotel accommodation where appropriate - \$17,280

Travel to the American Education Research Association (AERA) Annual Meeting

- In Year 1, 4 team members (PI, Co-PI, 2 GAs) to AERA in Atlanta, GA – allowed \$1,500 per person: \$6,000 total
- In Year 2, 10 team members (PI, Co-PI, GAs, teachers) to AERA in Philadelphia, PA – allowed \$500 per person: \$5,000 total
- In Year 3, 4 team members (PI, Co-PI, 2 Gas) to AERA to a location not yet determined – allowed \$1,500 per person: \$6,000 total

Annual Meeting of PI and Co-PI to meet with the Project Officer in Washington,DC

\$1,840 each year for three years, \$5,520 total

Washington, DC 3 days / 2 nights Meals: \$71/day x 3 = \$213; Lodging \$230/day x 2 = \$460; Car Rental \$162 + \$85.20 mileage = \$247 Total = \$920/per person.

Other Direct Costs

Materials and Supplies

IRB Permission Letters

Because the high school students in this development project are under 18, it will be critical to obtain signed consent forms from their parents. It will be necessary to contact some parents directly by phone call, email, and use regular post to mail forms and return signed IRB permission documents. Estimated cost is \$700 per year, \$2,100 total.

Test and Survey Materials

We will need copies of the biology textbooks and curriculum guides and materials for use by the investigators. Also we will need to purchase some tests and surveys to administer to students and teachers (achievement goal orientation, epistemological beliefs, self-regulated learning). Year 1 - \$1,200, Years 2 & 3 - \$900 each, \$3,000 total.

Consultants (Biology Experts)

Two independent biology content experts to examine student writing submission and the resulting network graphs to determine whether the network graphs have validity. We estimate this to require 2 days work each at \$500 per day, total \$2,000.

Other

Internet file server @ \$4,700 and software @ \$800_ Dell PowerEdge R710 – Price \$5,500

Justification: The server for this project must be reliable and scalable to manage the expected load for at least 5 years. This server will be dedicated to this project and will host the website, and a related database. The website will provide access for both students and teachers to project resources while the database will store the information entered by users. The proposed server configuration utilizes two 1TB disk drives configured in a RAID array and will feature two Intel Intel® Xeon E5530 processors and 8MB of memory. The operating system is Windows Server 2008 R2, Standard Edition. Service maintenance, \$150 in Years 2 and 3, total \$5,800.

Teacher Stipends

We will provide the typical local daily rate for a substitute teacher (\$140 per day) for each teacher day spent on the project as an incentive for the extra work the teachers will be doing on the project. Year 1 and Year 2 Teacher stipends are 4 teachers @ \$140 x 13 days = \$7,280 per year; and Year 3 Teacher stipends are 8 teachers @ \$140 x 13 days = \$14,560. Total \$29,120

Tuition

Computed using the approved tuition charges for a one-half (1/2) time graduate assistant of \$7,315 for fall semester 2011, \$7,315 for spring semester 2012, and \$3,658 for summer session 2012. The charges quoted above are increased by five (5.0) percent for any project period occurring after summer session 2012, and each summer session thereafter.

F&A – ON CAMPUS RESEARCH

F&A rates are negotiated and approved by the Office of Naval Research, Penn State's cognizant federal agency. Penn State's current on-campus rates for research are 49% of MTDC from July 1, 2011, to June 30, 2013. New awards and new competitive segments with an effective date of July 1, 2013, or later shall be subject to adjustment when superseding Government approved rates are established. Per OMB Circular A-21, the actual F&A rates used will be fixed at the time of the initial award for the duration of the competitive segment.

Learning and Performance Systems (Education) / The Pennsylvania State University
Exploratory Development of a Tool to Measure and Analyze Structural Knowledge (ASK) in High School Biology Courses
Institute of Education Sciences
Project Dates: 07/01/2012 - 06/30/2015

	Year 1 07/01/2012 - 06/30/2013	Year 2 07/01/2013 - 06/30/2014	Year 3 07/01/2014 - 06/30/2015	Total
Direct Costs				
Salaries (Category I)				
<u>Clariana, Roy Boris (Principal Investigator)</u>				
30% (2.70 mos.) Academic; 33.3% (1.00 mos.) Summer				
<u>Barb, Adrian Sorin (Co-PI)</u>				
11% (1.00 mo.) Academic				
Total Salaries	47,156	48,570	50,028	145,754
Graduate Assistants (Category II)				
<u>2 Graduate Assistants (Year 1)</u>	33,464	0	0	33,464
Grade 12, Half-Time; Fall/Spring				
<u>4 Graduate Assistants (Years 2 & 3)</u>	0	68,938	71,006	139,944
Grade 12, Half-time; Fall/Spring				
Total Graduate Assistants	33,464	68,938	71,006	173,408
Total Salaries and Wages	80,620	117,508	121,034	319,162
Fringe				
<u>Category I @ 31.80%</u>	14,996	15,444	15,910	46,350
<u>Category II @ 14.10%</u>	4,718	9,720	10,012	24,450
Total Fringe	19,714	25,164	25,922	70,800
Total Salaries, Wages and Fringe	100,334	142,672	146,956	389,962
Modified Total Direct Costs				
<u>Travel to Schools (Year 1)</u>	3,960	0	0	3,960
4 Investigators @ \$90/each visit x 11 visits				
<u>Travel to Schools (Years 2)</u>	0	8,640	0	8,640
6 Investigators @ \$90/each visit x 8 visits x 2 sites				
<u>Travel to Schools (Year 3)</u>	0	0	17,280	17,280
6 Investigators @ \$90/each visit x 8 visits x 4 sites				
<u>Travel - AERA Conference (Year 1)</u>	6,000	0	0	6,000
Atlanta, GA @ \$1,500 each x 4 (PI, Co-PI, 2 G.A.'s)				
<u>Travel - AERA Conference (Year 2)</u>	0	5,000	0	5,000
Philadelphia, PA @ \$500 each x 10 (PI, Co-PI, 4 G.A.'s & 4 Part. Teachers)				

Proposal: 3812

Generated by rac175 on: 09/20/2011

Learning and Performance Systems (Education) / The Pennsylvania State University
Exploratory Development of a Tool to Measure and Analyze Structural Knowledge (ASK) in High School Biology Courses
Institute of Education Sciences
Project Dates: 07/01/2012 - 06/30/2015

	Year 1 07/01/2012 - 06/30/2013	Year 2 07/01/2013 - 06/30/2014	Year 3 07/01/2014 - 06/30/2015	Total
<u>Travel - AERA Conference (Year 3)</u>	0	0	6,000	6,000
To Be Determined @ \$1,500 each x 4 (PI, Co-PI, 2 G.A.'s)				
<u>Travel - Annual DC Meeting PI & Co-PI</u>	1,840	1,840	1,840	5,520
Washington, DC 3 days / 2 nights Meals: \$71/day x 3 = \$213; Lodging \$230/day x 2 = \$460; Car Rental \$162 + \$85.20 mileage = \$247 Total = \$920/per person				
<u>Other Direct Costs - Materials & Supplies</u>	700	700	700	2,100
Secure IRB Permission Letters (paper, envelopes, phone, & follow-ups) @ \$700/yr.				
<u>Materials & Supplies</u>	1,200	900	900	3,000
@ \$1,200 Yr. 1, \$900 in Yrs. 2 & 3 for reference materials & survey consumables				
<u>Consultants - Professional Service (Year 2)</u>	0	2,000	0	2,000
2 Biology Content Experts @ \$500/day x 2 days				
<u>Other - Internet Server (Year 1)</u>	5,500	0	0	5,500
Hardware & Software @ \$5,500				
<u>Other - Server Maintenance (Years 2 & 3)</u>	0	150	150	300
@ \$150				
<u>Other - Teacher Stipends (Years 1 & 2)</u>	7,280	7,280	0	14,560
4 teachers @ \$140/ea. x 13 days				
<u>Other - Teacher Stipends (Year 3)</u>	0	0	14,560	14,560
8 teachers @ \$140/ea. x 13 days				
Total Modified Total Direct Costs	126,814	169,182	188,386	484,382
Other Direct Costs				
<u>Tuition Remission</u>	30,724	64,518	67,744	162,986
Total Other Direct Costs	30,724	64,518	67,744	162,986
Total Direct Costs	157,538	233,700	256,130	647,368
F&A Costs				
<u>F&A Rate: 49.00%</u>	62,144	82,901	92,310	237,355
Total Requested From Sponsor	219,682	316,601	348,440	884,723

Proposal: 3812

Generated by rac175 on: 09/20/2011

Project Narrative

Exploratory Development of a Tool to Measure and Analyze Structural Knowledge (ASK) in High School Biology Courses

Significance: Context

Science literacy for all citizens and advanced science knowledge and skills for future scientists and engineers is key for our future competitiveness. The National Assessment of Educational Progress (NAEP) report released in 2011, *The Nation's Report Card: Science 2009*, groups students into Basic, Proficient, and Advanced levels of accomplishment. The Proficient level represents solid academic performance, with the ultimate goal that all students perform at the Proficient level or higher. The Basic level denotes partial mastery of the knowledge and skills fundamental for proficient work. In 2009, only 21% of U.S. twelfth-grade students attained the Proficient level on the NAEP science portion, and only 1% attained the Advanced level. In that report, 60% of students attained the Basic level and 18% were below the Basic level.

Further, the U.S. ranking in science compared to peer nations is troubling. The Program for International Student Assessment (PISA) coordinated by the Organization for Economic Cooperation and Development (OECD) is an international organization that measures the performance of 15 year olds in reading, mathematics, and science every three years. It is encouraging that although the U.S. average score in science on the PISA in 2009 has improved compared to 2006, in the 2009 PISA, U.S. students only ranked in the middle of the sample of 33 OECD countries, with 12 nations scoring significantly better than the United States including in order, Finland, Japan, Korea, New Zealand, Canada, Estonia, Australia, The Netherlands, Germany, Switzerland, The United Kingdom, and Slovenia; and 9 scoring significantly lower than the United States (PISA, 2009, p.24). The highest scoring country in 2009 was China (Shanghai), a non-OECD country included in this testing (Fleischman, Hopstock, Pelczar, & Shelley, 2010).

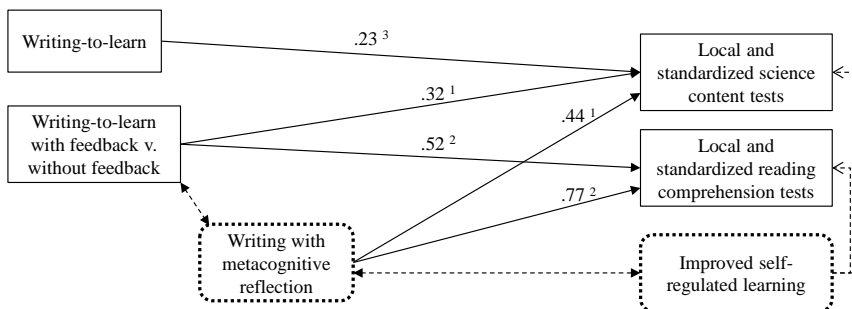
Literacy skills are essential for learning science, but 75% of 12th grade students scored at the basic level or below on writing on the 2007 NAEP. Two recent Carnegie reports, *Writing Next* (Graham & Perin, 2007) and *Writing to Read* (Graham & Hebert, 2010), describe an alarming picture of our students' reading comprehension and writing skills. These reports note that only a quarter of 12th grade students are proficient writers and so are not ready for the workplace. Thirty percent of high school graduates entering college are not ready for English composition courses and so are not ready for college. There will be long-term costs to individuals with low literacy, especially if they drop out of high school. Estimates of the yearly cost of the current literacy deficit to the national economy range from \$3 billion to \$16 billion.

In this proposed software development project, writing, reading comprehension, and learning are closely related. Writing closely resembles thinking and learning and is a demonstration of cognition (Emig, 1977). This project will be conducted with biology teachers in a traditional high school and also in an online high school. This project will integrate two existing standalone software tools (*ALA-Reader* and *Pathfinder KNOT*) that together capture and then visually represent students' content knowledge structure derived from writing assignments. We provide evidence below that this approach is a way to

capture and analyze structural knowledge. The software running in an internet browser will generate real time and at-a-glance graphic representations of students' writing that will be used by them for follow-on metacognitive reflection and by their teachers for visual progress monitoring. If shown to be valid, this software could have wide ranging application across many content areas and delivery approaches.

Significance: Intervention and Theory of Change

The theory of change is based on the premise that writing-to-learn can have a generative effect on learning content knowledge (effect size = .23). Writing with feedback is even better (effect size = .32). Even better than that, reflective writing can scaffold metacognition and self-regulation that positively influences learning (effect size = .44, see Figure 1). Self-regulated learning depends on the learner making correct assessments of what they know and don't know, but novices are not very good at making these estimates, so feedback is critical for learning. The evidence suggests that this intervention will also improve reading comprehension (effect sizes = .52 and .77), so both biology content knowledge and reading comprehension will be measured during the project. Also, students' self-regulated learning will be measured.



1 – effect sizes from Bangert-Drowns, Hurley, and Wilkinson, 2004 (meta-analysis)

2 – effect sizes from Graham and Hebert, 2010 (*Writing to Read* meta-analysis)

3 – effect sizes from Graham and Perin, 2007 (*Writing Next* meta-analysis)

Figure 1. The intervention and theory of change for this project.

This project will ask students to write weekly (i.e., summary writing of the lesson readings or some other writing tasks). Feedback will be immediately provided online by the software and will consist of a highlighted network diagram of the students' writing summaries and the instructor's network graph for that prompt (see Figure 2). The students' metacognitive reflection will consist of a 2-paragraph written response to the network diagram. For example, based on that student's network graph shown in Figure 2, she could reflect that a substantial number of her ideas agree with the instructor and she may list these. She may note that she did not use the term *self sense* and where it may fit and that she associated *neurosignature* with *neuron* rather than with *phantom pain* and may comment on that difference.

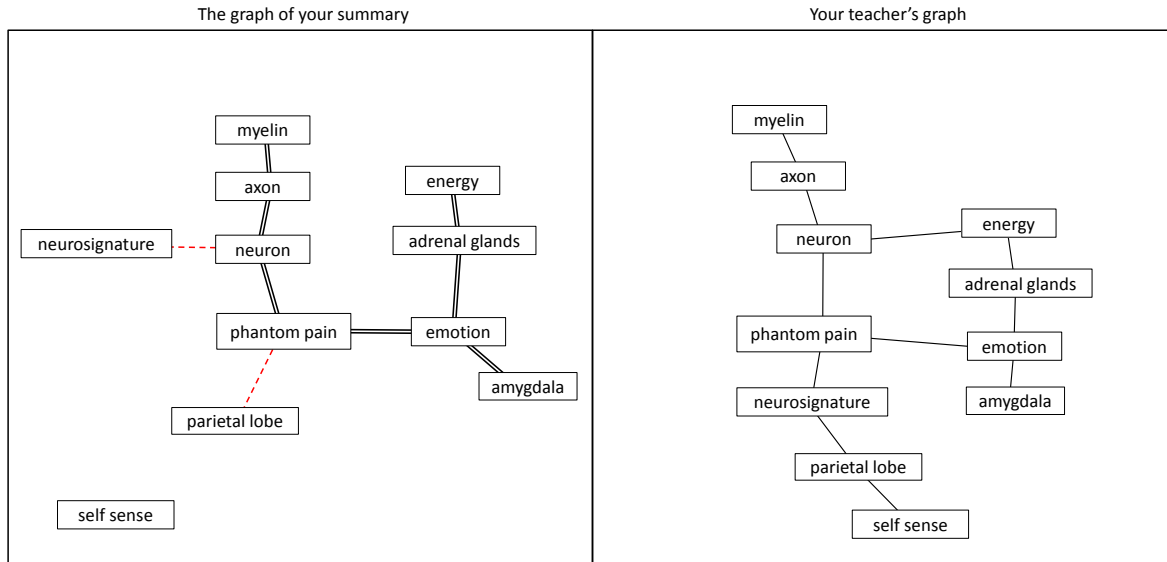


Figure 2. Example of a network graph derived from a student's writing summary and the instructor's model. The students will be asked to compare their network to the instructor's and then reflect on their writing based on this comparison (figures adapted from Ozgungor & Gregory, 2004).

In addition, the software will provide teachers with easy-to-use data for making instructional decisions (e.g., moving on to the next topic or re-teaching a topic). A recent IES Practice Guide, NCEE 2009-4067 (Hamilton, Halverson, Jackson, Mandinach, Supovitz, & Wayman, 2009), examined the evidence for using student achievement data to support instructional decision making. This proposed project will address three of the five goals stated in the Guide (p.8), Goal 1. *Make data part of an ongoing cycle of instructional improvement*, Goal 2. *Teach students to examine their own data and set learning goals*, and Goal 4. *Provide supports that foster a data-driven culture within the school*. Although data-based decision making is an extensive and growing practice (e.g., Quarterly benchmark tests, and so on), the level of evidence to support this practice is classified by IES as *low*, thus showing the need to conduct research in this area.

This three year project will develop the online software based on our existing standalone software and approach and will determine whether this software and approach positively influence students' learning and affects teachers' practice.

Significance: Theoretical and Empirical Basis

In this project, writing, reading comprehension, and learning are seen as intimately related. Following Emig's (1977) view, writing closely resembles thinking and learning and is thus a concrete artifact or manifestation of cognition. In a meta-analysis of the effects of writing-to-learn with or without feedback, Bangert-Drowns et al. (2004) say, *... learning entails active, personal, and self-regulated construction of organized conceptual associations, refined by feedback processes. The same features characterize writing. Writing requires the active organization of personal understandings. The externalization of those understandings in*

symbolic form makes them available for feedback in self-reflection and revision, in review of a record of the evolution of ideas and understanding, and in documentation for public discourse. (p.29)

Thus writing-to-learn with feedback and follow up metacognitive reflection is a reasonable strategy.

There is evidence that writing-to-learn even without feedback or reflection is instructionally effective; a meta-analysis by Graham and Perin (2007, *Carnegie Report: Writing Next*) reported an effect size of .23 for writing-to-learn compared to no prescribed writing. A meta-analysis by Bangert-Drowns et al. (2004) that considered the effects of feedback when writing-to-learn reported that for 47 studies that compared the effects of school-based writing-to-learn with and without feedback, feedback was more effective for academic achievement with an effect size of .32; and for 46 studies that used writing-to-learn with metacognitive reflection or not, metacognitive reflection was more effective for academic achievement with an effect size of .44 (refer back to Figure 1).

A meta-analysis by Graham and Hebert (2010, *Carnegie Report: Writing to Read*) complements the findings of Bangert-Drowns et al. (2004) and extends those findings beyond content knowledge to consider the effects of writing-to-learn on general reading comprehension. Among other findings, Graham and Hebert reported an effect size of .52 on experimenter-designed reading comprehension tests in 19 studies for those students who wrote text summaries compared to other treatment interventions that included reading only, re-reading the passage, reading and studying, and receiving reading instruction. They reported an effect size of .77 on experimenter-designed reading comprehension tests across 9 studies in favor of metacognitive reflection (i.e., categorized as personalization, analysis, or interpretation) compared to reading the text, reading and rereading it, reading and studying it, reading and discussing it, and receiving reading instruction.

These meta-analyses show that the effects of metacognitive reflection accrued over time, often taking months. This suggests that applying a reflection strategy positively alters students' self-regulated learning, a finding that was observed by McCrindle and Christensen (1995). If students become better learners, that is a good thing.

In this project, the novelty of the immediate feedback provided online as a network graph may motivate students to write, at least initially, and provides an artifact for further reflection that in this case is especially related to the structure of their knowledge. This reflection should have a generative learning effect (Grabowski, 2004; Lee, Lim, & Grabowski, 2009) since it requires cognitive activity simultaneously with multiple concepts. Even more important, this feedback should support self-regulated learning by providing a visual knowledge artifact for reflection (Bangert-Drowns et al., 2004; Butler & Winne, 1995; McCrindle & Christensen, 1995).

Writing is common in high school science for both instruction and assessment. Kihara, Graham, and Hawken (2009) report that about half of the high school science teachers surveyed used summary writing or responding to writing assignments at least weekly (see Table 1), and this is far more prevalent for online high school science courses. Thus asking teachers and students to complete written assignments will not be an unusual request.

Table 1. Frequency of writing in high school science classes (Kiuahara et al., 2009).

	Less than once weekly	At least once weekly	Once/week	Several times /week	Daily
worksheets	16%	84%	35%	40%	9%
short answer	19%	81%	38%	36%	7%
respond to a reading	46%	54%	33%	15%	6%
summarize a reading	58%	42%	22%	16%	4%
lists	63%	37%	21%	13%	3%
step-by-step	65%	35%	20%	9%	6%
journal entry	75%	25%	9%	6%	10%
5-paragraph essay	92%	8%	6%	2%	0%

(about 350 high school science teacher responses)

This project will ask students each week to summarize or respond to assigned readings (of biology content) because these types of writing samples are able to be analyzed by the existing *ALA-Reader* approach. Writing text summaries is one relatively well established “writing for learning” approach (Bartlett, 1932; Hidi & Anderson, 1986; Maclellan, 1997; Yu, 2009). Summary writing is effective for improving reading comprehension (Bensoussan & Kreindler, 1990), for science content acquisition (Friend, 2002; Radmacher & Latosi-Sawin, 1995; Selinger, 1995), and is even related to general academic success (Maclellan, 1997). A comprehensive literature review completed in Year 1 of this proposed project plus feedback from teachers and students may lead us to use other kinds of writing that are more effective, more acceptable to the participants, and that are also able to be analyzed by the software.

The Foundations of the Online Software

What is the nature of knowledge and how can it be measured? Besides familiarity and tradition, investigators’ and practitioners’ beliefs about the nature of knowledge determine the set of tools that they are willing to use to measure it, and using a tool will eventually alter understanding and practice (Sternberg & Preiss, 2005). Our connectionist bias (Rumlehart, Smolensky, McClelland, & Hinton, 1986) follows Anderson’s (1984) view that structure is the essence of knowledge (p. 5). Knowledge structure may be a facet of declarative knowledge (Mitchell & Chi, 1984), but Jonassen, Beissner, and Yacci (1993) go further to suggest that structural knowledge is a distinct type of knowledge, and that apposite structural knowledge is a critical go-between for declarative and procedural knowledge (p. 4). Our focus in this project is on domain knowledge structure (e.g., biology content) but not on team or process aspects of mental models.

We propose that knowledge structure is worth measuring. Measures of content knowledge structure have been empirically and theoretically related to memory, classroom learning, insight, category judgment, rhyme, novice-to-expert transition (Nash, Bravaco, & Simonson, 2006), reading comprehension (Britton & Gulgoz, 1991; Guthrie, Wigfield,

Barbosa, Perencevich, Taboada, Davis, Scafiddi, & Tonks, 2004; Ozgungor & Guthrie, 2004), and to the recent approaches and findings for combining individual knowledge structures to form group mental models (Cureeu, P.L., Schalk, R., & Schrujjer, S., 2010; DeChurch & Mesmer-Magnus, 2010; Johnson & O'Connor, 2008; Mohammed, Ferzandi, & Hamilton, 2010; Pirnay-Dummer, Ifenthaler, & Spector, 2010). For example, meta-analytic regression of 231 correlations collected from 65 independent studies of team performance show that shared team cognition, typically as shared content knowledge structure, accounts for significant incremental variance in performance (DeChurch & Mesmer-Magnus, 2010). There is evidence for a relationship between knowledge structure and various kinds of important and perhaps unusual learning outcomes.

Knowledge structure can be intentionally elicited from an individual in various ways but it can also be derived from existing artifacts such as essays and concept maps (Clariana, 2010). For example, compare-contrast type essay questions have been used to assess relational understanding that is part of knowledge structure (Gonzalvo, Canas, & Bajo, 1994). Jonassen et al. (1993) described 22 methods for eliciting knowledge structure including interviews, concept mapping, card sorting tasks, pair-wise concept associations, free associations, and others. Thus knowledge structure exists in and propagates through such artifacts. The nature, context, and purpose for creating the artifact deserve close attention because of their effect on the quality of the artifact for deriving the knowledge structure residue contained in the artifact (Clariana, 2010). The investigators in this project will pay close attention to the development of the writing tasks and prompts.

The online software developed in this project will be patterned after an existing software and approach called analysis of lexical aggregates (*ALA-Reader*, Clariana & Koul, 2004). The principles that led to the design of *ALA-Reader* were derived from diverse sources including in order of importance: Walter Kintsch's (1994) work on propositional analysis, John Deese's (1965) work on linguistic associations using word association tasks (including its current implementation as Princeton's *WordNet* lexical database), the *Pathfinder Network* literature (Schvaneveldt, 1990), the connectionist literature (i.e., Clariana, Wagner, & Murphy, 2000), and to a lesser degree from the automatic text-summarization literature (e.g., see the collected papers in Mani & Maybury, 1999).

The existing *ALA-Reader* software is not web-based and was designed for researchers, not for students' use. We have conducted several investigations using *ALA-Reader* software to analyze biology content and also management content. There are several investigations in progress that are not described below, but here are the studies that have been published so far.

Clariana and Koul (2004) developed and used the *ALA-Reader* approach to analyze students' essays on the structure and function of the heart and circulatory system relative to an expert's essay. Working individually, participants researched the topic online and created concept maps as notes using *Inspiration* software. Then using their concept map, they wrote a short essay text summary of their concept map (about 300 words). At that time the *ALA-Reader* software used a less effective sentence aggregate approach, not the more recently developed linear aggregate approach. For benchmark comparison, the essays were also scored by 11 pairs of raters and these 11 scores were averaged together into a composite essay score. Then the 11 individual rater scores and the *ALA-Reader* score were

correlated with that composite score. Compared to the composite score, the *ALA-Reader* scores were 5th out of 12, with an $r = .69$. Note that the *ALA-Reader* scores were not used to create the composite score and so this is a conservative comparison that strongly favors the raters' scores over the *ALA-Reader* score.

In a follow up investigation, Koul, Clariana, and Salehi (2005) used the *ALA-Reader* sentence aggregate approach again to score students' essays on the structure and function of the heart and circulatory system. This time working in *pairs* in class, participants researched the topic online and created concept maps as notes using *Inspiration* software. Then using their concept map, participants individually wrote a short essay text summary (about 300 words). The essays were scored by *ALA-Reader* relative to an expert's essay, by premier commercial essay scoring software called *Latent Semantic Analysis (LSA)*, Landauer, Foltz, & Laham, 1998), and by 11 pairs of raters using two different rubrics. As above, the 11 rater scores were averaged together into one composite essay score (a conservative value that favors the raters). Compared to the composite essay score, the *ALA-Reader* essay scores were 5th out of 13 ($r = .71$) and *LSA* scores were 9th out of 13 ($r = .62$). As before, relative to the rater composite score, *ALA-Reader* performed better than seven of the raters and also was better than *LSA*.

This finding for *LSA* essay raw score correlation to the human rater composite score ($r = .62$) is consistent with previous research that reported *LSA* raw score correlations between $r = .55$ to $.65$ (see Attali, 2007 and also Attali & Burstein, 2006). Authors of *LSA* investigation adjust these moderate raw score correlations to a 'true-score' correlation of about $r = .95$. True-score correlations are derived by dividing the raw correlations by the square root of the product of the measure reliabilities; thus a raw score correlation of $r = .65$ is reported as a true score correlation of $r = .90$.

The Koul et al. (2005) results above show that *ALA-Reader* out-performed *LSA* ($r = .71$ compared to $r = .62$, these are raw scores not true scores), this is notable since the investigation was designed using the *LSA* writing prompt that was available on the web as a free demo, and *LSA* is probably the best commercial essay scoring system of the several that are available on the market.

Clariana and Wallace (2007) used *ALA-Reader* to score essays on management theories relative to an expert referent and also to establish and compare group average knowledge representations derived from those essays. As part of their final course examination, undergraduate business majors were asked to write a 300-word compare-and-contrast essay on four management theories from the course. The essays were scored by *ALA-Reader* using both a sentence and a linear aggregate approach. To serve as benchmarks, the essays were also separately scored by two raters who obtained a Spearman rho inter-rater reliability of $\rho = .71$ (this rater-to-rater correlation is not high, but is consistent with most inter-rater reliability values). The linear aggregate approach obtained larger correlations with the two raters (ρ rater 1 = $.60$ and ρ rater 2 = $.45$) than did the sentence aggregate approach (ρ rater 1 = $.47$ and ρ rater 2 = $.29$).

In addition, the group average network representations of low and high performing students were reasonable and straightforward to interpret. The high group was more like the expert, and the low and high groups were more similar to each other than to the expert (see Figure 3).

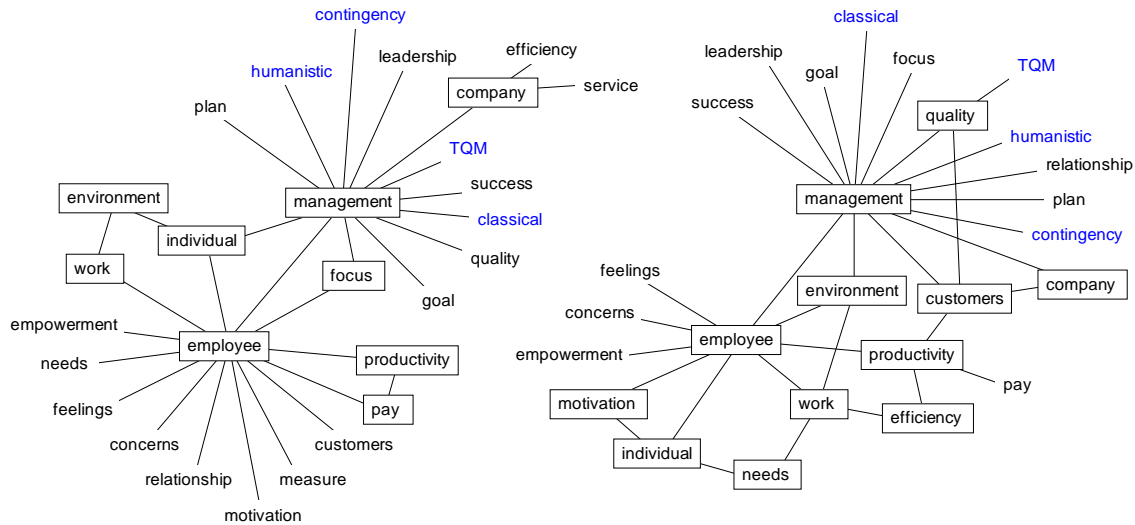


Figure 3. The average group knowledge network representations derived from essays of the low performing students (left panel) and of the high performing students (right panel) from Clariana and Wallace (2007).

Here is an example of a 1-paragraph reflection on these two network graph representations: *“Both the low and high groups’ network structure showed employee and management as the central and most connected concepts. The terms company, environment, individual, productivity, and work were also important concepts for both low and high performing groups. The four super-ordinate management theory categories from the essay prompt (classical, humanistic, contingency, and TQM) were all associated with management. Action terms such as leadership and success were also associated with management, while emotion-related terms such as feelings, concerns, and motivation were associated with employee. Both groups associated productivity to pay, although the high group had a richer representation of productivity that included customers and efficiency. The main differences between the network graphs displayed in Figure 3 are that the high-performers had relatively more terms that were central (e.g., with more than 1 link) and the high group’s association structure was more extensive and complex. This finding is consistent with theory and research on expertise.”*

In a follow-up investigation two years later, Clariana, Wallace, and Godshalk (2009) considered the effects of controlling anaphoric referents (in this case, pronouns) on *ALA-Reader* text processing. Participants in an undergraduate business course again completed an essay as part of the course final examination but this time, participants were given a list of 28 key terms to include in their essays. During analysis, the investigators manually edited these essays to replace the most common pronouns *their*, *it*, and *they* with the appropriate noun referent. The original unedited and the edited essays were processed with *ALA-Reader* using both the sentence and linear aggregate approaches. These data were then analyzed as before using *Pathfinder* analysis. The network average group representation similarity values comparing the original to the edited essays were large (i.e., about 90% overlap) suggesting that pronouns do not profoundly confound the analysis, showing that *ALA-Reader* is fairly robust in dealing with anaphoric referents. In other words, it can handle pronouns. As learned previously, the sentence aggregate approach, the

linear approach provided a better measure of individual essay scores, with a Pearson correlation $r = .74$ with the composite score.

These four *ALA-Reader* investigations described above show moderate correlations between rater essay scores and *ALA-Reader* essay scores. Note that the essays in the first two studies used mostly technical biology vocabulary while essays in the second two used fairly general vocabulary that included multiple synonyms for key terms, such as manager, supervisor, and boss for the key term *management*. In these investigations, the more technical or specific the vocabulary in the essays, the better *ALA-Reader* performed. In addition the linear approach was better than the sentence approach, and this may relate to both the nature of knowledge structure and the linearity of reading and writing expository text.

An interesting aspect of these investigations and a feature of this approach that will be used in this proposed project is the capability to generate both individual and average group network graphs that represent content knowledge structure in a useful and informative way. Network graphs of content knowledge structure in paper-based form have been used as feedback for students (see the dissertation by Taricani, 2002) and for instructors (see the Carleton College Science Education Resource Center best practice guide, 2010). A meta-analysis by Nesbit and Adescope (2006) of the effects of network graphs to support learning (i.e., as concept and knowledge maps) reports that this approach is motivational and instructionally effective across a wide range of interventions, and is especially effective for learners with low ability or low domain knowledge.

Ozgungor and Gregory (2004) used Pathfinder network (PFNET) representation to consider the effects of elaborative questions when undergraduates read a 1,500 word expository science text passage on phantom pain, compared to a reading-only control. The average group knowledge structure of participants in the questions treatment group was more like the average expert representation than was that of the reading-only control (see Figure 4). Based on inspection of these three network graphs, Ozgungor and Gregory reflect that:

The mental representation of the individuals who were in the treatment group displays a relatively well-structured organization of the passage. As in the experts' representation, phantom pain forms the center of the network, with links to other major concepts in the passage. Again, semantically related concepts are linked together in a way that produces a structure very similar to the experts' network. It is organized around at least two important concepts introduced in the passage: emotion and neuron. The terms on the right side of the network are related to emotion. The terms on the left side pertain to neuron. ... Although the treatment group did not link two important, semantically related concepts to each other (parietal lobe and neurosignature), they were able to perceive the relationships among main concepts the relationships of concepts such as phantom pain to neurosignature have been correctly identified by the treatment group but not by the control group. On the other hand, in the concept map representing the knowledge structure of the control group, it is more difficult to detect organizational principles. Emotion, which had a relatively minor value in understanding the content, is located almost as centrally as phantom pain. The hierarchy among the concepts of

neurosignature, self sense, and parietal lobe, is not evident in the control group.... Likewise, the control group's organization of knowledge surrounding the concept of emotion is not as similar to the experts as that of the treatment group. Finally, the control group, on average, has linked entirely unrelated concepts to each other, as in neuron to neurosignature and self sense to neuron. (p.441)

We think that the information captured and displayed in the group average PFNETs provides useful information for further reflection and for instructional decision making.

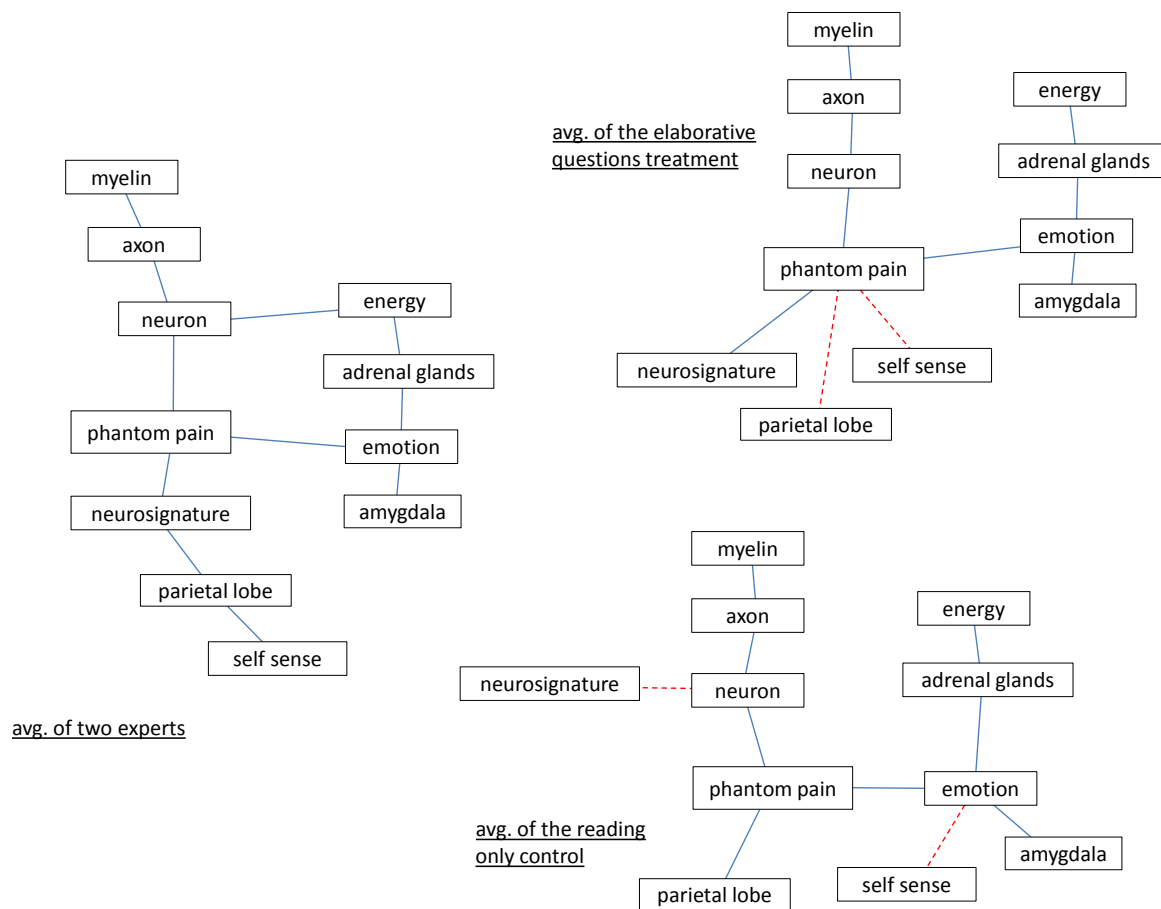


Figure 4. The average network knowledge representation PFNETs for the experts, for the elaborative questions treatment, and for the reading-only control (from Ozgungor & Gregory, 2004). Dashed lines indicate links not present in the experts' PFNET.

There is data from Nash, Bravaco, and Simonson (2006) to suggest how network graphs plus a summary table could be used as an object for reflection and for data-based decision making. In their investigation, 22 high school teachers completed a workshop on Java and object-oriented programming. Individual knowledge structure of this content was measured before and after the workshop using the Pathfinder pairwise approach, and these were averaged to form pre-workshop and post-workshop average group Pathfinder network

graphs (PFNets). Inspection of the pre-workshop network graph (see the left panel of Figure 5) reveals that *object* is the central and most important term for these teachers before the workshop with six links (the number of links to a concept is referred to as the node degree, an important measure of graph structure such as in social network analysis). These six links are shown in the table shown in the right panel of Figure 5. But after the workshop, the importance of the term *object* has decreased (four links) and is more like the expert (with three links).

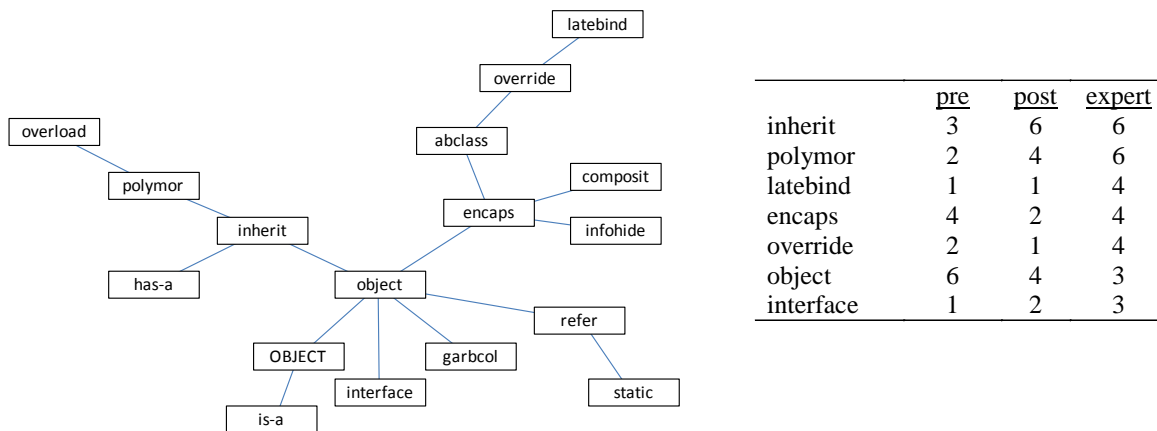


Figure 5. *Pathfinder* network graph representation of the average of 22 teachers collected pre-workshop and a table of node degree (from Nash et al., 2006, p.44).

Further examination of the node degree table shows that the terms *inherit* and *polymer* both have six links in the expert network graph and so are central concepts; the post-workshop network node degrees for the teachers for these two terms are more like the expert than were their pre-workshop node degrees. In other words, the learners' Java and object-oriented programming content knowledge structure reflected in these networks became more like the expert over the course of the workshop. We are able to see this immediately even though we are not familiar with this topic. Thus we believe that if the instructors who know this content are given a representation like that shown in Figure 5 and a node degree table, they could determine quickly which topics to focus on (i.e., in this example, *inherit* and *polymer*) and also possible misconceptions they should address (i.e., *object*).

In this grant project, knowledge structure is elicited from individuals through writing – for example a text summary of a textbook chapter as an in-class assignment or a content-specific essay question on an end-of-unit test. The software will provide the writing prompt and a textbox; students will compose their response on a word processor and then copy-and-paste it into the online software's submit text box. The software will generate and display a Pathfinder network graph of that text summary and of the instructor's network graph. The students' network graphs will include highlights showing the similarity and difference compared to the instructor's network graph. The students can print these off and then compose a response to submit to the instructor (refer back to Figure 2).

The software will archive the participants' writing submissions for further analysis, and instructors can use the software to automatically analyze the archived data to generate visual knowledge structure network graphs and other progress monitoring reports of groups and of individuals, as required to inform class decision making. The fully developed software would allow the instructor to 'see' the group representation within moments of the students' submissions, which will provide useful progress-monitoring feedback for the instructor.

The software can provide individual scores for each student relative to the expert (for example, a student could be told that their essay scored 6 out of 21 possible points). The scores generated by the software will be included in the project analysis and reports and will be compared to scores generated by human raters as a measure of content validity. However, providing "scores" immediately to students could distract from the primary intent and could inhibit rather than promote reflection. Therefore software scores will not be given to students since the emphasis of this project is on the instant visual feedback as network graphs for metacognitive reflection by the students and for instructors to assess learning progress. These scores will be provided to the instructors who may decide how to use the scores.

Significance: Practical Importance

The software developed during this project should encourage and support both writing-for-learning as well as metacognitive reflection. The evidence presented above supports the hypothesis that this would be instructionally effective, with increases in biology content knowledge on teacher-made and state standardized biology tests, on standardized reading comprehension tests, and in improved self-regulated learning. Practically speaking, students will know more biology, will be better readers and writers, and will be better learners – skills needed by 21st century citizens.

This project will develop online software with built-in writing prompts for high school biology. Teachers anywhere can use these immediately with their students with very little training.

After using the software in this way over time, we expect that instructors will modify their instruction. These changes will be described and documented qualitatively and quantitatively. More uses for the software will evolve and be piloted over the course of the project (for example, collaborative writing or applying the software to the text generated in online discussion boards). In such cases, our programmer will create a utility application (app) to move the discussion text into the software making it possible to analyze the discussions. Analyzing online discussions could become another important use of the software.

Significance: Rationale Justifying the Importance of the Software

Knowledge structure is theoretically and practically related to higher-order learning outcomes and performances. It is a different kind of knowledge that is worth further investigation.

In addition, this online software, like its predecessor *ALA-Reader*, is flexible in terms of the content it can handle. The software can generate a network representation of any written text or writing sample given only a list of important terms and at least one

expert writing sample for benchmark comparison. This is substantially different than commercial essay scoring software (e.g., *Intelligent Essay Assessor* from the Educational Testing Service and *IntelliMetric* from Vantage Learning LLC) that cannot generate a network graph but only a ‘score’ and requires very costly setup involving ‘training’ the software with thousands of pages of prompt-specific text and with hundreds of prompt-specific scored writing samples.

But the online software developed in this project will cost almost nothing to set up for a new writing prompt, and there is evidence that it is as effective as commercial software for scoring essays. In other words, a college chemistry teacher, a high school social studies teacher, and an elementary school teacher can all use it. They would only need to create the writing prompt, provide a list of key terms, and provide an expert benchmark essay. When their students submit their writing online, a highlighted network graph would be displayed immediately along with the instructor’s network graph. Students may use these for metacognitive reflection as proposed here, or just as content feedback for planning revisions.

Long term, this project will confirm and extend the research that we have been conducting for several years, allowing us to quickly pilot new text capture, analyses, and comparison approaches. For example, we have only used and described simple force-directed graphs (e.g., in Figures 2 through 5), but Dr. Barb has also proposed adding visual features such as color, directed links, and smaller and larger text sizes to indicate the frequency of words. For example, a *Wordle* (<http://www.wordle.net/create>) visual representation of this proposal is shown in Figure 6. However the *Wordle* representations only show occurrence and frequency of terms in the text, but locations of the words in a *Wordle* visual are random. Our software could show occurrence and frequency, and relationship would be shown by nearness of terms and by linking lines. Structural knowledge is about the relationship among knowledge elements.

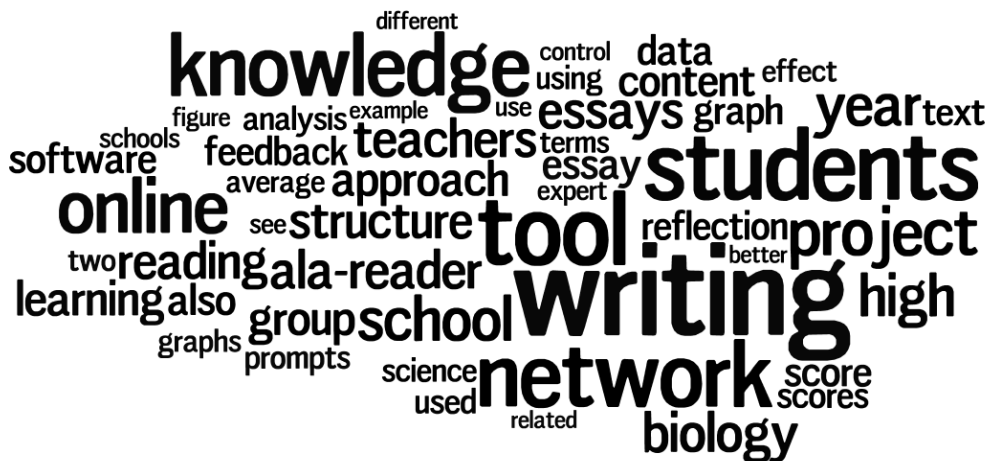


Figure 6. A *Wordle* text representation of this Project Narrative.

The existing *ALA-Reader* software uses a symmetric Boolean approach that disregards the directionality of the text proposition structures. Dr. Barb has added five more text capture approaches to the working online prototype including asymmetric Boolean that captures concept directionality (i.e., dog bites man versus man bites dog). This

directionality information may be important data, so we will compare the results of symmetric versus asymmetric text capture to see which obtains better content validity. We may also integrate a semantic database into the software so it can automatically handle synonyms, metonyms, and roots of words. These and other approaches will be considered as the software is developed and used by the project students and teachers. This three-years-long project will provide an opportunity to further this research and to improve the software even beyond what is proposed in this project.

Methodological Requirements: Sample

The online version of the software will be developed in Years 1 and 2 working with a traditional high school in State College, PA (SCASD) and with the 21st Century Cyber Charter School (21CCCS) in Downingtown, PA. The SCASD high school biology classes are inclusion classes taught over two semesters to about 140 at-risk students (about 40% with special education IEPs). The 21CCCS high school biology course is taught each semester to about 75 online students (150 students per year). This will allow us in Year 1 to iterate the software features frequently as needed and to go through two development cycles of the writing prompts. Due to the differences in biology content and student populations, the writing prompts developed in Year 1 in these two schools may not completely overlap.

In Year 2, we will use the Year 1 foundation to modify existing prompts and to develop more writing prompts as needed. This approach allows us to compare and design for both traditional and online implementations of the software, and gives us interactions with different kinds of students using different biology textbooks and curricular approaches so that the software is applicable to a wider audience.

To examine feasibility in Year 3, we will add two ‘control’ schools – another traditional high school (to be recruited) and another online cyber high school (nearby cyber high schools include Agora Cyber Charter School in Devon, PA and Achievement House Cyber CS in Exton, PA. Other possible participating cyber schools include the Pennsylvania Cyber CS in Midland, PA, the Susq-Cyber CS in Milton, PA, and the Pennsylvania Learners Online Regional Cyber CS in Homestead, PA). Since teachers and administrators change, these schools will be recruited in Year 2 for participation in Year 3. For their participation, these control schools will be offered training on how to use the software after Year 3. After Year 3 we would offer the software to any school for a small license fee (i.e., \$10 per school per year), this revenue would be used to maintain (or upgrade if demand is high) the server, software, and internet access, and to improve the software.

This quasi-experimental design in Year 3 will include four schools, two intervention and two control schools. Because of the lack of random assignment to treatment, covariates will be included in the analysis to control for school (or class/course) differences. The covariates will be at the course or school level from the previous year, and will include Pennsylvania System of School Assessment (PSSA) scores in reading and biology. These state-standardized tests have good reliability and are given to all schools, so are available and are appropriate for these four schools. If the state changes the mandated standardized test during the project, the project will use the new test in consultation with the schools involved. Though this is a high school biology intervention, the evidence from

the three meta-analyses suggests that PSSA reading scores (effect size = .77) will improve more than biology scores (effect size = .42). This quasi-experimental approach should provide any evidence for going to the next stage with a true experimental investigation. If warranted, a proposal for a cluster randomized trial (CRT) of the intervention will be submitted. With that possibility in mind, we would collect data (for example, effect sizes for the different learning outcomes) in Year 3 to support a CRT proposal.

Methodological Requirements: Iterative development

In summer of Year 1 before classes start, investigators will meet with teachers at the two school locations for a three-day workshop to review the project and curriculum and to set writing tasks and prompts for the first few months. The teachers will be primarily responsible for creating the writing prompts appropriate for their students and their content.

In the fall of Year 1, teachers will begin to use these writing prompts in their classes and to collect the text summaries generated. The software will not yet be available to them, but may be phased in during Year 1 as it is developed. The investigators will examine the data regularly and will log their reactions to the text summaries, reflections, and network graphs that are generated. The investigators will meet each month with the 21CCCS teachers to review the students' writing samples that are generated and to revise the past prompts as required, and the teachers will continue to develop biology content-related writing prompts for use the next month. In the regular monthly meeting at the school sites, we will discuss issues such as whether students correctly interpreted the writing prompts thus producing comparable text summaries. If a prompt is ambiguous, we will correct it.

Separately the investigators will analyze the text summaries obtained for each writing prompt to identify key terms and alternate place holders (i.e., synonyms and metonyms) and will establish an expert referent for each prompt. These key terms and expert referent benchmarks are necessary for the full implementation of the tool in Years 2 and 3. The software will be authored and developed and these writing prompts, key terms, and benchmarks will be entered. Development will include the user-interface design and the background text analysis and representation approaches.

The steps in the approach are: elicit (text summaries, other prompts) → capture (number or terms, symmetric, Boolean, other) → analyze (array structure, Pathfinder, cluster-analysis, multi-dimensional scaling, other) → compare (structure of the network graphs, measures of similarity, other). Improving upon the existing design used in *ALA-Reader* must address all of these steps. Some design questions based on survey and interview data include how many terms should be used during analysis and representation and what are the preferred and best writing prompts and metacognitive reflection prompts? As a basis for these decisions, we will synthesize the existing literature and try the most promising strategies. Some design questions based on student data include – what is the best text capture approach, what is the best data reduction approaches beside the Pathfinder network approach already used in *ALA-Reader*, and what is the best way to derive the expert referent? In this way, several best approaches will be identified early in Year 1 based on feedback from students and teachers and the actual student writing samples. For example, three different text analysis approaches with three different data analysis

approaches would produce nine potentially different network graphs of the same set of text summaries. We will ask which is preferred, which is best for metacognitive reflection, and which is best for content revision? The teachers and students will be involved in answering these questions.

We will continue to modify the user interface as well based on student and instructor feedback so that by the end of the spring semester, we have confidence in the interface, in the biology writing prompts and metacognitive reflection prompts, in the prompt specific terms list, in the expert benchmark, and in the analysis approach. In summer, we will fully embed these prompts, terms, and benchmarks so that the software is fully automated by Year 2.

We currently have a partially working mockup of the online software running on a University web server. It has various levels of login permissions (admin, teacher, student, and researcher) with different screen interfaces for each category of user. This mockup allows us to enter a writing sample and a list of key words with their synonyms and metonyms. When you click *Preview*, the software generates and displays the [n x n] array file and a forced-directed network graph of the raw array data. The *Pathfinder* data reduction algorithm is not yet implemented in the mockup. Besides the symmetric Boolean sequence approach used in *ALA-Reader*, the mockup includes six approaches that are available and functioning including the symmetric and asymmetric forms of Boolean, of count sequence, and of maximum inverse word distance. The software runs in an internet browser on Windows and Mac computers, and even on the iPad; most of the processing is done on the server side.

In Year 2, we will begin to use the online software with the 21CCCS and with the SCASD biology teachers. Before school starts in the fall of Year 2, we will implement a 3-day workshop for participating teachers at each school site to introduce the project progress and the plan. The teachers will have hands-on training until they feel comfortable. We will also review the existing prompts including the curriculum alignment of the prompts and will add new prompts. Teachers at both schools will begin to use the software in the fall semester and we will meet at least monthly with teachers face-to-face at the school sites and online via *Skype* video conference.

In Year 3, we will investigate the effects of the fully developed software in the two intervention schools using the software (SASD and 21CCCS) compared to two matched control schools not using the software. In all three years, we will collect PSSA data in Reading Comprehension and Biology to make year-to-year comparisons in case the software development process affects PSSA outcomes. We will also collect survey data from teachers and students including achievement goal orientation, epistemological beliefs and beliefs about writing, and measures of self-regulated learning. Since there may be an expert reversal effect, with low-ability students benefitting most from the intervention (Nesbit & Adescope, 2006), analysis will include this possibility. We will also survey and selectively interview students and teachers regarding their perceptions of the intervention and its effectiveness.

Methodological Requirements: Feasibility of Implementation

When the software is fully developed in Years 2 and 3, the high school students will complete teacher-assigned writing prompts regularly from a list of prompts and will

compose and submit their text summaries and reflections. The software will immediately display their individual network graph. Then they will compose and submit their reflection. It is fairly common for students to submit written work in science courses (Kiuahara et al., 2009) and so the actual activity is not a barrier. This software is based on the existing *ALA-Reader* approach that has some published evidence of validity, thus the software design and development is not a barrier since there is a working model to use as a basis for the software. The 21CCCS has agreed to participate. We have met with the high school biology teachers at SCASD who teach the 10th grade at-risk course sections; they support the grant and are eager to participate, although we do not yet have school board approval. We expect that we can recruit two matched control schools in Year 3. The use of writing-to-learn with feedback and metacognitive reflection is supported by substantial published research. Thus this intervention is feasible.

Methodological Requirements: Pilot Study

The theory of change involves writing-to-learn with feedback and metacognitive reflection. The research reported in the three meta-analyses above suggest that this intervention will positively affect students' learning as measured by the state standardized Biology test and Reading Comprehension test (the Pennsylvania System of School Assessment, PSSA).

To determine the promise of the project, the pilot study conducted in Year 3 will include two intervention schools using the software compared to two control schools not using the software. Year-to-year state-standardized test (PSSA) scores will be collected and analyzed as additional evidence to provide an estimate of the promise of the intervention. Possible change in teacher practice data will be collected through interviews with the teachers throughout the 3 year project period to measure, for example, whether teachers assign more or different writing tasks and if any non-writing activities are discontinued because of more writing. Surveys will be taken of students' and teachers' achievement goal orientation, epistemological beliefs, and beliefs about writing. Self-regulated learning will also be considered. This provides a rich data set to understand the relationship between writing with feedback, metacognitive reflection, and the software and its perceived effect on learning.

Methodological Requirements: Measures

PSSA data for students taking the high school biology course at the participating schools will be collected from the state website and analyzed for each year of the study.

Teacher data will be collected during regular site visits and by interviews and surveys with questions such as: do you think that writing is an effective learning strategy in your online course? Do you think that the metacognitive reflection helped your students learn biology? About how many writing assignments did you assign last year in total for your entire course? How many did you assign this year? Do you feel that having the software score the text summaries for you would influence you to assign more writing tasks? Do you feel that the network graph is a true measure of students' biology content knowledge? What do you think of the network graph? What do you like and dislike about the network graph? We would carefully develop these questions before using them to

ensure unbiased responses. Students will complete surveys and random groups of students will be interviewed regarding the interface and software.

In order to measure the quality of the group average network graphs, besides the teachers' and students' perceptions, we will have external biology content experts read the students' text summaries and examine the group average network graphs to make an expert determination as to whether the graph captured the information in the text summaries. We will also compare group average network graphs for the same writing prompt for low and high subgroups from different schools and different semesters to see if these group graphs are related, similar to the comparison shown in Figure 3 (Clariana & Wallace, 2007).

Personnel

The PI, Roy Clariana, is a professor in Instructional Systems in the College of Education at Penn State University, State College, PA. He has a doctoral degree in Curriculum and Instruction (instructional design and technology emphasis) and has published over 50 experimental investigations, including numerous investigations in K-12 schools. He has a master's degree in biology and taught high school biology for 5 years. He was the primary author of the middle school environmental education curriculum, *TVA: A world of resources*, a curriculum sourcebook that was awarded 1st place in the "Take Pride in America", Federal/Education Division in 1988. He led the project to assess the fundamental knowledge related to resumption of nuclear operations at the Rocky Flats Nuclear Weapons Facility in Golden, Colorado and was the Director of International Operations (UK) for Jostens Learning Corporation collaborating with the British National Council for Education Technology (NCET) in the two year assessment of educational technology in schools in the UK. From 2007 until 2011, he was the Academic Division Head of Education at the Penn State Great Valley graduate campus outside Philadelphia. He conceived, designed, developed, and pilot-tested the *ALA-Reader* approach that is the basis of the software tool to be taken online in this project. The PI is seeking a course release in fall and in spring plus 1-month compensation in summer to implement the proposed project. He will manage budgets and reports, have the lead role working in the schools (logistics, teacher training, planning, interviews, surveys, data collection), and will oversee the graduate assistants (GAs) who will support these tasks. The PI will also conduct the data analysis and will be the main author of the reports. He is 41% of effort and has no other funded grant or grant submissions at present.

The Co-Pi, Adrian Barb, is an assistant professor of Engineering and has a PhD in Computer Science. He has conducted research in knowledge representation, applying data mining techniques like those envisioned in this project, and he has direct expertise in designing online software like the one in this project. He will be the software database programmer but will also participate closely in the teacher interviews, especially those that are related to making iterations of the software. He is 11% of effort and has no other funded grant or grant submissions at present.

The Graduate Assistants (GAs) recruited for the positions will need sufficient biology content knowledge to analyze the high school curriculum, to interact with teachers and students about the content, and to score high school biology essays. This requires a minimum bachelor's degree in biology, but an advanced degree in biology would be preferred. Preference will be given to candidates who have related teaching experience.

Two as yet unidentified biology content experts (\$1,000 each) will be sub-contracted to compare the network graph output and the related students' writing samples used to create the graphs for several of the writing prompts in order to make an external expert determination of the quality of the network graphs and to make recommendations to improve the software output.

Resources

This project requires office space, a meeting room, personal computers and some specific software for the investigators (Pathfinder *KNOT* and some server software), and an internet file server. The University will provide office and work spaces and provides personal computers with all the research-related software we need (i.e., SPSS, data mining software, etc.). The GAs will have shared office space and computers. The project runs on an internet server. Having our own file server provides us with more freedom and control to load and modify the software. The University will provide the internet bandwidth for the project file server. The 21 Century Cyber Charter School (21CCCS) is eager to work with us and the CEO (Principal) has provided a letter of agreement that is attached to this proposal. The biology teachers from the State College Area High School (SCASD) are eager to participate, but we do not have school board approval yet, and so we cannot provide a letter of support from SCASD at this time, although we believe that the district will allow us to conduct this project in their district because of its potential especially for at-risk students.

Appendix A

Letters of Support:

Adrian Barb – Penn State Great Valley Campus

Jon Marsh - 21st Century Cyber School



To whom it may concern

From: Dr. Adrian S. Barb (adrian@psu.edu), Assistant Professor of Information Science

RE: Roy Clariana Institute for Education Sciences Application

Date: September 17, 2011

I am writing to provide enthusiastic support of Dr. Roy Clariana's proposal entitled "Exploratory Development of a Tool to Measure and Analyze Structural Knowledge (ASK) in High School Biology Courses" which is being submitted to the Institute for Education Sciences. Dr. Clariana and I have a common area of interest in domain knowledge representation of novice learners and of experts, and although we approach it from different research bases, our knowledge, skills, and experience in this area complement each other. I am committed to this project as a collaborator, with expertise in knowledge discovery, computer models, and data mining databases, computer algorithms, human computer interaction, information science and technology. You have my support in terms of materials and intellectual resources to make this work a success.

Sincerely,

Adrian S. Barb, PhD.

Assistant Professor of Information Science
School of Graduate Professional Studies
Penn State University
30 E Swedesford Road
Malvern, PA 19355
Phone: (610) 725-5349
Email: adrian@psu.edu



September 12, 2011

Dr. Roy Clariana
Division Head & Professor
Penn State University
Office phone: 610-648-3253
Email: RClariana@psu.edu

Dear Dr. Clariana:

As the CEO of the 21st Century Cyber Charter School, I am happy to provide access to the 21CCCS Biology Instructors for the grant period, with a special emphasis on the first year. We have had numerous working arrangements with Penn State in the past with good results and we look forward to working with you again.

Sincerely,

A handwritten signature in black ink, appearing to read "Jon D. Marsh", written over a circular stamp or mark.

Jon D. Marsh
Chief Executive Officer
484-875-5454 (direct)
jmarsh@21cccs.org