Adv Physiol Educ 35: 59–67, 2011;
doi:10.1152/advan.00016.2010.

Model-based reasoning: using visual tools to reveal student learning

Douglas Luckie,[1] Scott H. Harrison,[2] and Diane Ebert-May[3]

[1]*Department of Physiology and Lyman Briggs College,* [2]*Department of Microbiology and Molecular Genetics, and* [3]*Department of Plant Biology, Michigan State University, East Lansing, Michigan*

Submitted 29 January 2010; accepted in final form 10 November 2010

Luckie D, Harrison SH, Ebert-May D. Model-based reasoning: using visual tools to reveal student learning. *Adv Physiol Educ* 35: 59–67, 2011; doi:10.1152/advan.00016.2010.—Using visual models is common in science and should become more common in classrooms. Our research group has developed and completed studies on the use of a visual modeling tool, the Concept Connector. This modeling tool consists of an online concept mapping Java applet that has automatic scoring functions we refer to as Robograder. The Concept Connector enables students in large introductory science courses to visualize their thinking through online model building. The Concept Connector's flexible scoring system, based on tested grading schemes as well as instructor input, has enabled >1,000 physiology students to build maps of their ideas about plant and animal physiology with the guidance of automatic and immediate online scoring of homework. Criterion concept maps developed by instructors in this project contain numerous expert-generated or "correct" propositions connecting two concept words together with a linking phrase. In this study, holistic algorithms were used to test automated methods of scoring concept maps that might work as well as a human grader.

visual models; concept map; automated grading; C-TOOLS; Robograder

CONCEPTS WITHIN SCIENTIFIC DISCIPLINES are complex abstractions that experts use to analyze and interpret interconnected qualities of the natural world (4, 29, 30, 46). Yet, college instruction often primarily involves passively transmitting large amounts of simple factual information and then testing student recall (2, 13, 34, 42, 51). In response, students do not learn deeply but rather use memorization to succeed in their courses (1, 2, 13, 16, 35, 37, 49). Beyond providing initial guidance on scientific terminology and basic relationships, college-level instruction may greatly benefit students if it helps show them how to build their knowledge of complex systems.

As a main vehicle in scientific thinking, models are focused depictions of systems that help to explain current understanding and specify hypotheses (8). Visual models can be a way to introduce students to expert thinking and help instructors to discover overall student understanding (12, 29, 30, 33). Visual models are illustrations that attempt to simplify and represent a cycle, mechanism, idea, or system. These can include flow charts, diagrams, or sketches that connect images and words with arrows and phrases. The term "visual" is used here to separate illustrative models from mathematical ones. The value of illustrations such as concept maps is that they can challenge each student to grapple with their understanding about the relationships between important ideas in science (33).

It is useful to bring common practices used in science like visual modeling into the classroom (20, 22, 26, 27). Since students often confront both new vocabulary and ideas, concept maps are an excellent tool to address these needs (25, 33, 36). Our colleagues often agree that it would be desirable to use concept maps in their teaching, yet a significant challenge is grading large numbers of them in introductory courses (39, 40). Scoring a single concept map takes considerably more time than computer scoring of multiple-choice exams. To address these challenges, we developed a new online drawing tool, the Concept Connector, to allow students to easily create concept maps. It was designed to provide instant feedback to students who do not immediately see the proper relationships among concepts. Automated feedback features that stimulate more reflective map building by students and allow rudimentary scoring may enable more instructors to use concept mapping.

In this report, we present a case study that tested the use of our visual modeling software in an introductory biology course. Student data were used to evaluate new scoring algorithms that might be useful for automated grading. In our findings from this study, several topological measures showed potential and may be able to help software strategies approach the same scoring accuracy achieved by an instructor. This could enable a scale up of the use of online visual modeling to aid student learning in physiology and elsewhere.

METHODS

Building the drawing tool software, the Concept Connector. The Concept Connector was designed to present classroom problem sets with a concept map drawing area for science students. A concept map contains concepts and linking words or phrases. When two concept words are connected by a phrase, the unit is called a proposition (Fig. 1). The Concept Connector software allowed students to move any preseeded concept words around, add additional concepts, organize hierarchy, and add linking words and lines. The Concept Connector software is a Java applet that is small in size and browser compatible on every current desktop operating system (e.g., Linux, Mac, and Windows). There is a server layer that handles data transmissions from student-operated Java applets to facilitate archival, submission, and automated grading of concept maps. The server layer also controls selective, instructor-specified delivery of different menu options for concept map modification. The overall outcome is a simple and efficient set of user actions provided as menu controls on the Java applet boundary. We used design methodologies (14, 43) to refine the software interface and performance as well as test the Concept Connector with undergraduate science majors in biology, geology, physics, and chemistry courses. This is a report on one case of a class of biology students.

Building the scoring tool software, Robograder. The drawing software's automatic grading feature is called Robograder. The Robograder scoring software feature is, at its most basic level, a script that contains all expert-provided correct and incorrect propositions correlated with defined +2, +1, −1, and −2 values. WordNet is an online thesaurus that Robograder can access to amplify the grading matrix created in a spreadsheet by the instructor or expert (Fig. 1*B*) (23). The

Address for reprint requests and other correspondence: D. Luckie, Dept. of Physiology, Michigan State Univ., 2100 Biomedical and Physical Sciences Bldg., East Lansing, MI 48824-3320 (e-mail: luckie@msu.edu).

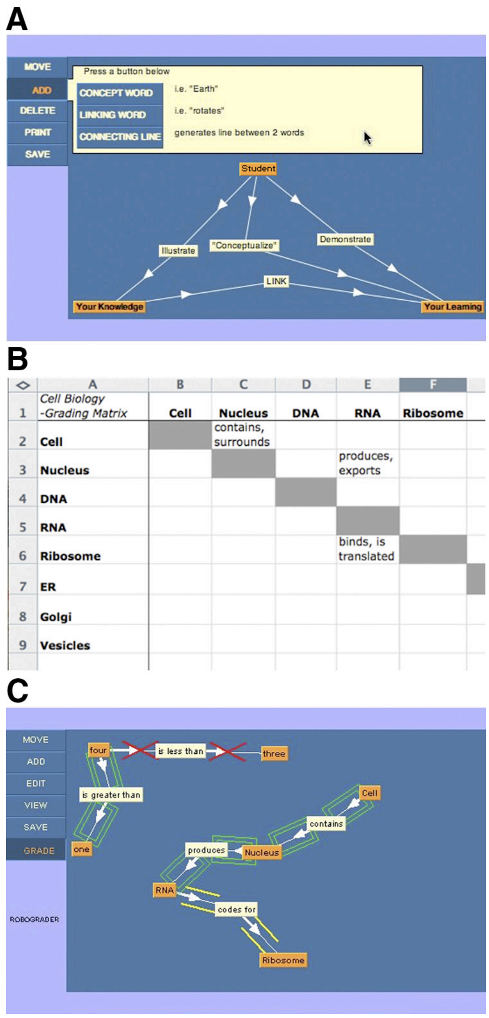

Fig. 1. The Concept Connector software features and tools for instructors. *A*: the Java applet graphic user interface (GUI). These screenshots show the Java applet's GUI (blue areas), how the software draws a concept map, and how the menus appear while in use (http://ctools.msu.edu/). *B*: the instructor creates a grading matrix (with correct and incorrect answers) using an Excel spreadsheet so that Robograder can give students automated feedback when they build their maps online. *C*: Robograder can use both instructor-provided answers as well as a math or other libraries to evaluate the propositions. It can give differing levels of positive (green halos) or negative (red "x"s) feedback as either single or double symbols and can give neutral feedback (yellow lines).

software can do more than just give feedback of correct or incorrect responses but actually can indicate two levels of positive or negative feedback based on whatever the instructor has defined. For example, they can define "correct" linking words as being either superior (+2) or acceptable (+1). Similarly, "incorrect" linking words can also be defined by the instructor as poor (−1) or very poor (−2). When a student building their model requests Robograder to grade their current draft, it highlights propositions known to be correct with double or single green halos to represent +2 and +1 values. Red "x"s with either one stroke or two appear on propositions known to be incorrect (Fig. 1*C*). Whenever Robograder is "not sure," i.e., has no information, it generates yellow rectangular halos on either side of those linking phrases. In automated grading, our ultimate goal is to approach the hierarchical scoring system developed by the Novak and collegues (35–37), yet Robograder, when used in this study, only gave visual feedback concerning the validity of the semantic relationship between linked words in a proposition. As a result, in this mode, Robograder can be used to evaluate nonhierarchical concept maps similar to those developed by Buzan and Buzan (5) and Fisher (19). An extension of the Concept Connector software, a new beta-version of an application called GUIDE, is now under development. It will assist students as they build maps that resemble biogeochemical/nutrient cycles similar to "box diagrams" used by geologists.

Using concept mapping in large introductory science, technology, engineering, and mathematics courses. While the data presented in this case study are focused on a cohort of 76 undergraduate students enrolled in introductory biology, for our research program, over the last 10 yr, we have recruited a cohort of over 1,000 physiology majors enrolled in introductory science major courses in biology, chemistry, physics, and geology. In all courses, before being assigned online homework, students learned to build concept maps by hand in groups with index cards (Fig. 2). Students were taught how to build hierarchical concept maps as developed by Novak and Gowin (37). They were provided with 10 concepts and a few blank cards to add their own. They were challenged to generate a hierarchy, connections, and linking words. During these introductory training experiences, the instructor moved from group to group, giving guidance and challenging weak or unclear elements of the groups' visual models.

In each course, the initial phase of software instruction was an in-class tutorial, often led by both the instructor and a student from the class, on using the Concept Connector web application. There is an easel mode where the concept map is developed, and a gallery mode where students manage an archive of their work to receive automatic scoring feedback and carry out the final submission. Concept map homework assignments in physics included "Where does the heat go in an oven?" and those in geology included "Trace the path of the water cycle." These assignments and further instructions were provided to students directly from the Concept Connector's course assignment menu. The software allowed students to revise their maps after receiving computer feedback. These courses piloted approaches for having students later also work with a partner to complete a final collaborative version of the map. After submitting their first online assignment, students worked in class with their instructor to evaluate several sample online maps to identify strengths and areas to improve in the visual model.

Teaching with concept mapping in the "case study" introductory biology course. In the introductory biology course case study presented in this report, during a 15-wk semester, the concept map assignments were given at the end of each 3-wk-long unit on a topic and served as an organizer/review of all the connections among the material that had been discussed. Students were provided with a fixed number of concept words and a blank concept map drawing area to work in. The four concept map assignments discussed in this report were on the following topics: *map 1*, the carbon cycle (8 words); *map 2*, Mendelian genetics (11 words); *map 3*, natural selection (10 words); and *map 4*, ecosystems (13 words). The change of student performance on *map 1* versus *map 4* is one focus of comparison in our analysis. Student maps were turned in as hard copies and graded by hand. All maps were hand graded by the instructor, and feedback was given to students when each assignment was returned. The instructor would follow up when returning each graded assignment in lecture by placing several example maps (names removed) on a document camera projected on the screen to discuss concept map attributes that were effective and those needing improvement. When used in this report, the term "holistic" refers to a grading method used by instructors that is not purely algorithmic and includes evaluation of the map as a whole, as a creative work, and using some intuition to judge it. In grading maps, the instructor in the biology course followed this general strategy but sampled the work and applied a holistic approach to *1*) look at hierarchy and add 1 point if it seemed reasonable (no dramatic errors), *2*) review a sampling of links made between concepts for validity and award 1–3 points, and *3*) evaluate the student's work based on the map as a whole and award 0–1 point. Expert-generated maps were evaluated by two other biologists in this study. The instructor's scoring of each student map was carefully reviewed

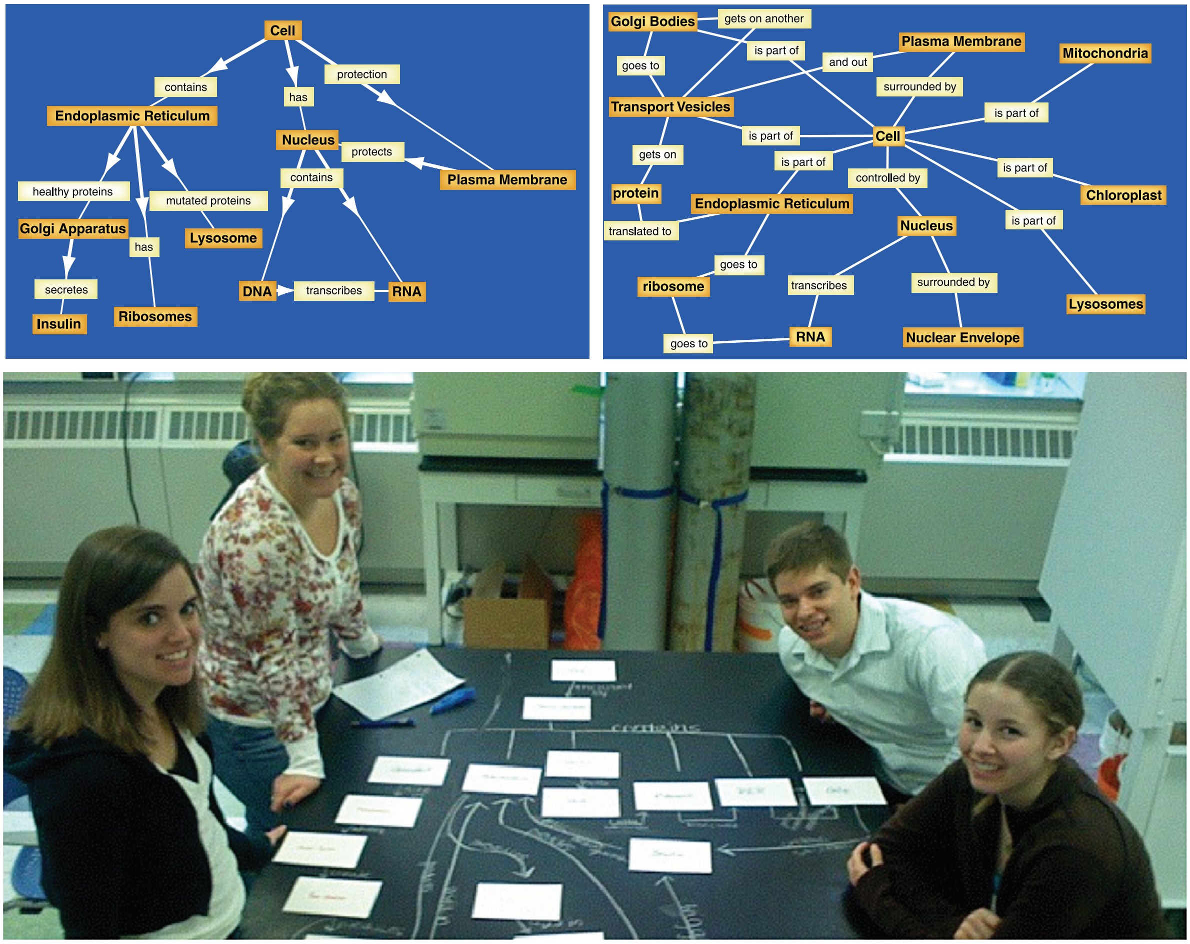

Fig. 2. Sample student work in introductory biology. *Top*: two sample maps are presented from an assignment where students were asked to explain how organelles work in a cell. *Bottom*: students worked in groups to learn to build concept maps with index cards before using online software. Students practiced generating hierarchy, connections, and linking words. Outside of class, students completed online concept map assignments as an integral part of the course.

after the course and was completed during our examination of all the data.

There was no formal pre- or posttesting used in the biology course or in each unit for student learning. Student work and a classroom instructor's grading of concept maps was used to search for trends in student maps as well as to test whether some form of computer grading strategy based on topology would correlate with the results of human grading.

Studying network topology approaches. After the class had ended, the research team tested four topology approaches that scored each student's concept map based solely on its networked structure. These topology approaches are not currently used by Robograder. Methodologies using map network patterns related to hierarchy, "leaves" and "degrees," and cross-linking, "cycles" and "RMS," were evaluated. Concept maps with a greater number of lateral connections would score higher on cycles (internal cross-links) or RMS (extended linkings), whereas maps with many stratified levels of hierarchy would score higher on degrees or leaves. Cycles is literally a count of the number of loops involving three or more concept words on a concept map, whereas degrees is a count of the number of propositions connecting to a given concept word. Leaves is defined as the number of terminal ends in the concept map network. RMS is defined as the root of the mean sum of squared distances between all concept word pairs within a concept map. Data were analyzed using R, a software environment for statistical computing and graphics (version 2.6.2).

RESULTS

The concept map drawing and scoring software at work. The concept map drawing tool allowed students to use menus to build a concept map with words, phrases, and lines with arrowheads. The graphic user interface presented input areas for text and some symbols, a menu system on the left side, and a blue drawing area, and major concept words were differentiated from linking words or phrases by font and color (Fig. 1*A*). Instructors provided the automated scoring software, Robograder, with correct and incorrect linking words or phrase "answers" by building a simple spreadsheet and submitting it (Fig. 1*B*). Once a student completed the first draft of their concept map, they received feedback from Robograder via visual cues of green halos, red "x"s, and yellow lines to help them identify regions of their concept map that might need revision (Fig. 1*C*). Submission to the instructor occurred by pressing a button and sending a digital image via e-mail or to a course management system, but often instructors also required a hard copy be generated and turned in during class for hand grading.

A sample of work by two students that was representative of that seen after students worked individually creating complete concept maps with online guidance provided by Robograder is

Table 1. *Summary statistics for maximum scoring concept maps*

	Percentage of Students With the Maximum Score	Averages for Maximum Scoring Concept Maps			
		Degrees	Leaves	Cycles	RMS
Map 1	25	4.6	1.7	2.7	2.0
Map 2	45	3.9	3.9	1.2	3.1
Map 3	19	4.9	4.1	2.4	2.9
Map 4	59	4.2	4.3	3.1	4.3
Correlation		−0.82	0.47	0.049	0.83

The maximum score was 5 of 5 points assigned by the human grader. Correlation is the Pearson product-moment correlation between the series of average topological measures with the series of percentages of students with the maximum score for each map exercise. RMS, root mean square.

shown (Fig. 2). Note that one student's map does not include a hierarchy, a common result on early drafts (Fig. 2, *top right*). Before using the software online, students initially learned how to create concept maps by working together with index cards and chalk drawing on laboratory benches during a class meeting (Fig. 2, *bottom*).

Comparing human scoring with network topology scoring of maps. Instructor-provided answers allowed the automated scoring of concept maps but only successfully graded 26% of the student-made propositions existing on our server. To increase this percentage, our focused objective in this study was to test whether a computer grading strategy using topology measurements could approach the results of human grading. We tested whether a range of algorithmic software approaches could mimic the human grader's results in terms of measures of degrees, leaves, cycles, and RMS.

Table 1 shows an analysis of concept map scoring data from each of the four map exercises in the biology course. Aligned are the percentages of students given the maximum score on each assignment by the expert faculty member with network topology scoring strategies where software evaluated the same student maps with four approaches. The RMS score had the highest positive Pearson product-moment correlation ($r = 0.83$) to the human grader's scoring and degrees had the most negative correlation ($r = -0.82$). Unpaired *t*-tests between maximum scoring concept maps ($n = 91$) versus other concept maps ($n = 175$) showed significant differences only for measures of RMS and cycles ($P = 0.04$ and $P = 0.006$). The instructor also perceived that the reduced student performance on the third assignment on the topic of natural selection was based on difficulty with managing complex interdependencies. Intriguingly, the decrease in student scores for *map 3* was best signified by RMS counts.

The instructor predicted that those students understanding the central importance of certain critical concept words, such as "photosynthesis" in *map 1*, would likely score the highest on their concept maps. Our analysis of the data tested this hypothesis. Table 2 shows data in the same format as Table 1 of just the subset of student maps that contained the most common "hub" concept words used. The term "hub" is defined here as the concept word with the most links in a particular map. The data appear to support the instructor's hypothesis. The percentage of students achieving maximum scores increased in Table 2 for all four map exercises, and was most significant for *map 1*. Overall, a Fisher's exact test between the consensus and alternative major hub word sets of concept maps for the maximum score, *5* (58 and 33, respectively), versus other scores, *1–4* (82 and 93, respectively, including 3 maps that had been assigned a score of 0 due to an inability to interpret any meaning by the human grader) was significant ($P = 0.01$), and there was an odds ratio of 2:1. A Mann-Whitney *U*-test, however, between all five possible scores of concept maps with the consensus major hub word ($n = 140$) and scores of maps having an alternative hub word ($n = 123$) was insignificant ($P = 0.126$). This discrepancy in significance may be due to the contrasting criteria for assigning a nonmaximal score on a simple ordinal scale. For Table 2, the trend was again found where RMS most positively correlated ($r = 0.68$) with the human grader scores, whereas degrees was most negatively correlated ($r = -0.97$).

Significant linear models and robust correlations were not generally found for the four different map exercises. The only constant aspect of topology measure identified for our exploratory data analysis across *maps 1–4* was a weak but relatively consistent series of correlation *r* values (0.21, 0.22, 0.17, and 0.23, respectively) between the score (*1–5*) and cycles measure (Table 1).

The change of student performance on *map 1* versus *map 4* on all student maps is another focus of comparison in this analysis. The instructor scores indicated that students did much better on the final map than on the first. As shown in Fig. 3, the four automated scoring strategies were compared with these two sets of student maps. The topology count of degrees for each concept map did not vary much from an average score of *4* between the first and last assignments as well as within an assignment when compared among maps scored *1–5* by the instructor. When comparing the first and last assignment, leaves and RMS topology scores more closely mimicked those of the human grader. Interestingly, whereas the average score of cycles did not shift greatly from the first to last assignment, this topology approach did differentiate more than others between maps that scored a *1* versus a *5* within a particular assignment.

A final comparison was made between student performance for the first and last assignment to a combination of both topology measures for hierarchy as well as both measures of cross-linking. As mentioned above, instructor scoring indicated the student performance on the final map, *map 4*, was much better than the first and showed a shift toward higher scores (Fig. 4*A*). Underneath the student performance distributions in Fig. 4 are the respective topology comparisons. In measurements related to hierarchy, the count of number of leaves once again increased more than degrees when *map 1* was compared with *map 4* (Fig. 4*B*), and a similar correlation was clearly found for the increase in RMS score with regard to cross-linking (Fig. 4*C*). Yet, as suggested by this replotting of network topology measures and the associated hot spots, it is not necessary to limit use to only a single measurement for hierarchy or cross-linking. A combined strategy may lead to a more robust approach.

DISCUSSION

This National Science Foundation (NSF)-funded project developed a new assessment tool, the Concept Connector, which consists of a web-based, concept mapping Java applet

Table 2. *Summary statistics for maximum scoring concept maps with consensus Hubs*

	Percentage of Students With the Maximum Score	Averages for Maximum Scoring Concept Maps			
		Degrees	Leaves	Cycles	RMS
Map 1	34	4.5	1.6	2.6	2.0
Map 2	61	3.8	4.1	0.9	3.1
Map 3	24	5.1	4.3	2.6	2.9
Map 4	63	3.9	4.8	2.7	4.2
Correlation		−0.97	0.44	−0.48	0.68

Concept maps with consensus hubs are those maps that share in common the most interconnected consensus major hub word for the map exercise (n = 140). The most consensus hub words for the map exercises were as follows: *1*, photosynthesis; *2*, genes; *3*, natural selection; and *4*, natural selection or ecosystems. The maximum score was 5 of 5 points assigned by the human grader. For those concept maps with consensus hubs, correlation is the Pearson product-moment correlation between the series of average topological measures with the series of percentages of students with the maximum score for each map exercise.

with automatic scoring and feedback functionality. The Concept Connector tool was designed to enable students in large introductory science classes to visualize their thinking online and receive immediate formative feedback. Further details concerning the goals and methodology of this project have been previously published (31, 32).

Development of online tools: software at work. Our research group created a new online tool to help students build visual models. As mentioned above, the Concept Connector drawing software is a Java applet available for download from the ctools server (http://ctools.msu.edu/). We have released two versions of the Concept Connector software: *1*) a tiny concept map drawing applet, which is easy to set up with minimal computer knowledge and can be served from personal Mac, Linux, or PC computers; and *2*) the full suite of Concept Connector software applications with Robograder, which may require the help of a system administrator to deploy on your own Linux PC server. In addition, with some seed money from the NSF, we also created a spin off of the Concept Connector software called GUIDE, which allows students to build a cycle form of concept mapping used by geologists called "box diagrams." Since box diagrams are very restricted in the format and number of correct answers possible, Robograder can give students feedback for 100% of their propositions. There's a call for the use of cyclic concept maps in education literature, so this new software may be a very helpful tool (41).

While the availability of online concept mapping software is quite limited and software that attempts to automatically score a student's concept map is even more rare, several other groups have attempted similar work. The "Reasonable Falliable Analyzer" software developed by Dr. Tom Conlon at the University of Edinburgh (15) gives numeric scores and hints to the student and enables the student to appeal their score. This software is not online and is limited to older Macintosh operating systems. Another is "Betty's Brain," developed at Vanderbilt University (3), where students must build concept maps to "teach" an artificial agent, a cartoon character named Betty. They then test Betty's learning by posing questions and evaluating if the answers are correct. The software is not online and currently only works with a small set of maps of biological processes, like the food chain, photosynthesis, and the waste cycle. A third system, which uses what the authors call "construct on scaffold" and "construct by self" approaches (9), was developed in Taiwan and provides evaluation results and hints to students by comparing the student's map with that of an expert. This program is Chinese language only and limits students when creating their concept maps to a short list of concept and linking words identical to those used in the expert criterion map. Dr. Roy Clariana has develop several "Mapper" applications at Penn State University that can score student essays or concept maps by comparing them with expert texts/maps as well as using a variety of distance (network proximity) data. His software can turn a student essay into a rudimentary concept map and then evaluate it. This software works well if the students limit the words they use to those of the expert answers (44). The criterion-related validity of this system has been found to be good (11). Finally, faculty members of the National Center for Research on Evaluation, Standards, and Student Testing at the University of California developed the only other known online tool that automatically scores concept maps. Their "Concept Mapper" software is a web-based Java applet that has automated scoring via a match-to-expert algorithm using expert maps as templates. Like most others, this software limits the number of words that students can use to those on a list used by the expert. They examined the validity of concept mapping as a measure of elementary students' scientific understanding (24, 28).

Development of instruction: faculty and students at work. Our initial interest in using visual models stems from a repeated observation on multiple-choice exams that many of our best physiology students could remember numerous details about the replication of DNA while not even comprehending the basic hierarchy that chromosomes have genes made of DNA. As we examined different topic areas in biology, this type of problem persisted. We found that spending more time teaching about scale and hierarchy (talking and waving our hands) did not change the students' performance, but using concept maps appeared to be an effective approach. Models are one of the common ways of representing phenomena in science; they are "the main vehicle by which science actually produces its explanations and predictions" (8). Using visual models to represent a hypothesis and communicate ideas is common practice in science (10). As a result, we worked with biology, chemistry, physics, and geology faculty members and their students using design methodologies (14, 43) to create a number of concept map exercises designed to work well in large introductory science classrooms.

We found that while time is always limited in the lecture period, when students were given opportunities to build concept maps by hand with the guidance of the instructor, they then quickly grasped the idea. In addition, it only took a 5-min introduction to the software in lecture, best if done by calling on a student to demonstrate, to give a tour that was considered satisfactory by the students. Students can quickly gain an appreciation for some of the subtleties of making a good map as a result of their instructor taking a few minutes at the start of the lecture to review samples of the best hierarchy and connections done by peers on an assignment.

In our own work with even these very small concept maps, the variation in student creative approaches to mapping their ideas is great. Since our students tend to succeed best on exams by looking for linear paths to solutions, thinking about multiple

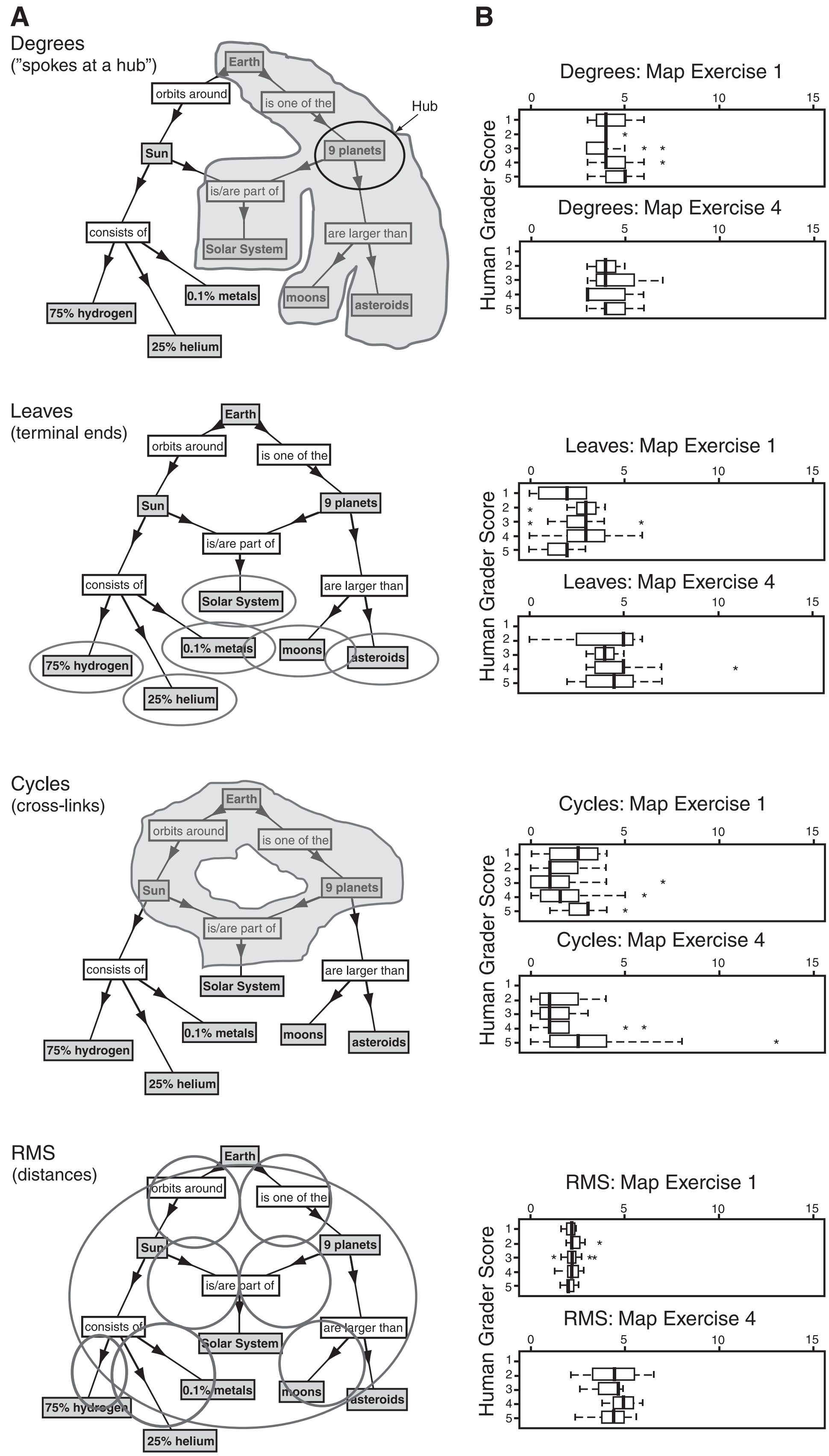

Fig. 3. Software-based measurement compared with human scoring of student maps. Human expert scoring of student maps from a biology course was evaluated in conjunction with automated analyses of trends in the map data. *A*: illustrations indicating what the degrees, leaves, cycles, and root mean square (RMS) topology approaches count when measuring a student concept map (see METHODS for more detail). *B*: respective charts that compare the human grader scores versus software-based measures of maps completed at the start of the course on *map 1* with those at the end on *map 4* (n = 75 and 41 students, respectively). Each chart separates and individually compares map sets that received scores of *1* (low) to *5* (high) from the instructor on the *y*-axis to the topology measure on the *x*-axis. The whiskers are drawn to data points within 1.5 × the length of the interquartile range away from the lower and upper quartile boundaries.

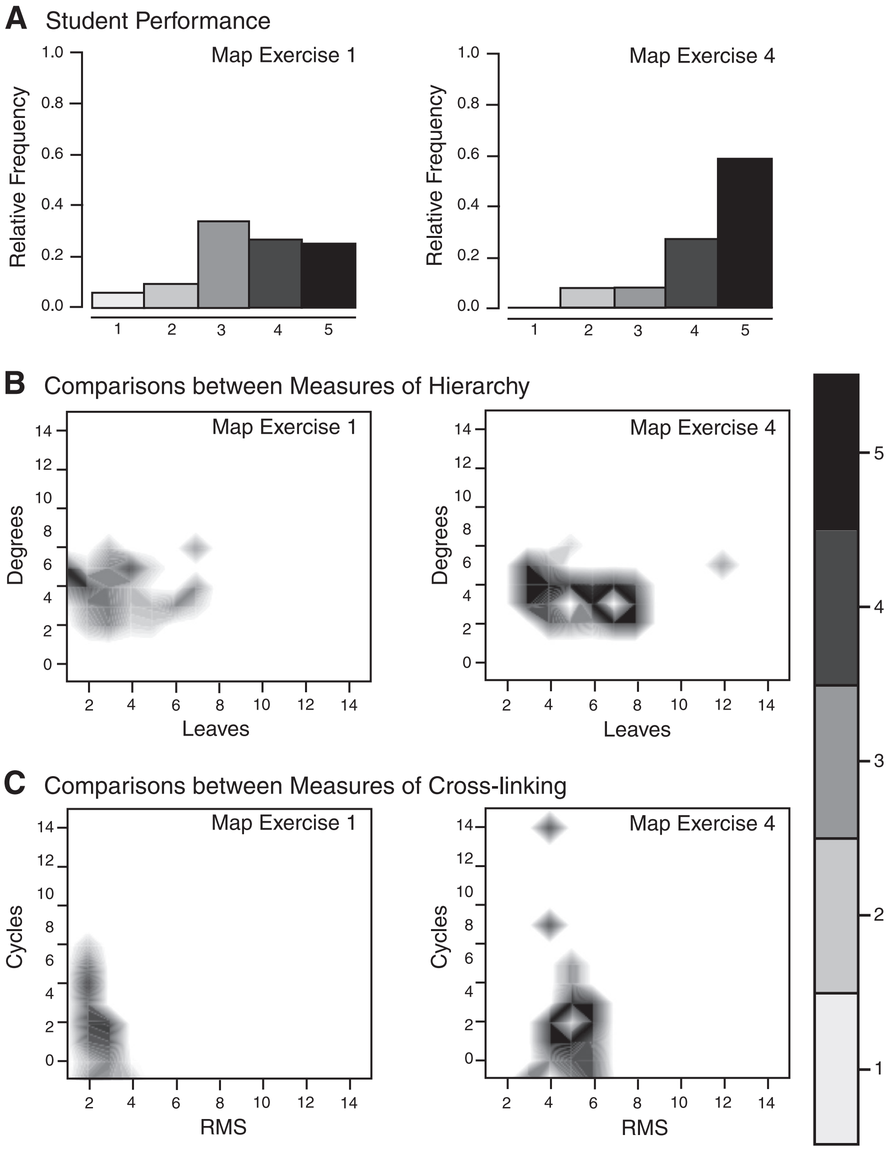

Fig. 4. Student performance compared with computer measures of hierarchy and cross-linking. *A*: student score distribution of *map 1* and *map 4* showing improvement in the frequency of higher grades awarded by the instructor (n = 75 and 41 students, respectively). *B*: two charts showing trends in topological computer scoring approaches designed to evaluate hierarchy when applied to the same maps evaluated in *A*. These are a comparison of the average values of two network topology measurements, degrees and leaves, for the all maps. *C*: two charts showing comparisons of measures associated with cross-linking, cycles and RMS, when applied to the same maps evaluated in *A*.

paths to an answer like those in physiology can challenges our "best" students most of all. We observed that even our graduate students and faculty colleagues sometimes found a simple 10-item concept map assignment designed for freshmen to be an engaging challenge and would debate the best organization and linking words/relationships. In addition, because the concept map as defined by Novak strongly values the element of hierarchy, the arrangement of top versus subordinate concept words can vary considerably depending on the context. One student might design a cell biology concept map from an anatomic context, yet another might design it from an energy context or the context of the biosynthesis of insulin. Even with the same set of 10 words, this allows for a variety of different student-generated maps to have the potential of being equally good. The question and context posed by the instructor for the assignment direct the student and can further refine what a good map should look like. For all student maps, the area we depend upon most to be the same is the validity of the connecting words chosen to characterize a relationship between two concepts. These linking words or phrases are more often clearly correct or incorrect. We occasionally found certain words that mean different things depending on which disciplinary context is involved; this generated enlightening discussions among faculty members about aligning the use of language in introductory courses across different science disciplines.

Development of automated scoring: network topology approaches enhance a grading matrix. Web-based concept mapping can enable students to save, revisit, reflect on, share, and explore complex problems in a seamless, fluid manner from any computer on the internet (38). As mentioned above, currently, instructor-provided grading matrixes (in the form of spreadsheets) have enabled Robograder's automated scoring of student concept maps to successfully grade 9,205 of the 35,404 student-made propositions (26%) in our database. We are seeking additional complementary approaches to increase that percentage as well as give feedback regarding map structure and hierarchy (35–37).

We are pursuing several strategies to improve Robograder: *1*) using approaches that are topological to recognize patterns in visual models that are frequently associated with good quality mapping; *2*) using pedagogy to include students in a peer review of concept maps, and perhaps contribute via wiki, and, as a result, build greater numbers of quality correct propositions in the grading matrix over time; and *3*) using online dictionaries and thesaurus software to immediately recognize acceptable misspellings as well as related synonyms and antonyms. In a separate study of WordNet, we found that the online electronic lexical database and thesaurus were very successful supplementing the finite set of answers provided by faculty instructors, a resource that has also been used in other investigations of concept map data (7, 18, 23).

The automated grading approaches presented in this report were based on the network structure of the student concept maps. Methodologies using map network patterns related to hierarchy (leaves and degrees) and cross-linking (cycles and RMS) were evaluated. In this case study, software-based scoring approaches that focused on topology measurements termed RMS and leaves correlated best with the human grader's own holistic approaches. This may either relate to how exposition (RMS) and detail (leaves) may drive either quality thinking of the part of the student or interpretation of that thinking on the part of the audience (e.g., the instructor).

The combination of multiple methods like those mentioned above could be a very powerful addition to our current assessment approaches. Our future efforts will likely be directed toward relating topological form to the potentially autogradable semantic interactions between concepts. The capacity to analyze and verify these predictions will grow in power with the accumulation of additional data and classroom-to-classroom comparisons. More studies of this sort may point to topological approaches that can be used automatically online to increase the number of propositions on a student concept map with which Robograder could give substantive positive and negative feedback. These approaches combined with answers provided by instructors, experts, and students in an online environment (perhaps driven by a wiki) has the potential to assist faculty members in grading student maps with efficiency and accuracy in large introductory courses.

A number of other researchers have pursued parallel studies into the development of a tool for automatic classification of concept maps based on a topological taxonomy that distinguishes novice from expert maps (47). The Novak and Cañas group developed what is likely the most sophisticated desktop-based concept mapping software, CmapTools. They also recently created a topological taxonomy of concept maps (6) and an associated software feature in CmapTools, the automatic topological classifier, designed to assess the quality of student-made concept maps based on their structural complexity (45). Their research supports and extends our findings with topological measures.

More studies of this sort may point to alternative approaches that can be used automatically online to increase the number of propositions on a student concept map with which Robograder could give that feedback. Our group is also currently studying approaches allow the concepts of structure, behavior, and function to be modeled in maps (26, 27). This strategy may be more flexible and revealing than traditional concept maps. In the near future, we hope visual modeling with software like the Concept Connector and its feature Robograder will aid our colleagues in large introductory science classes to use an additional tool in teaching and assessment. Beyond the multiple-choice exam and rote learning, challenging students to wrestle with the new ideas of science within the boundary of a concept map with responsive feedback may begin to help direct them toward expert knowledge and higher-level learning (33, 35, 50).

ACKNOWLEDGEMENTS

The authors thank Drs. James Smith, Janet Batzli, Duncan Sibley, John Merrill, and Everett Weber as well as Marija Krha and Michael Haenisch for helpful discussions.

GRANTS

This work was supported by National Science Foundation Research Grant 0206924 (to D. B. Luckie).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

1. **Ausubel D.** *The Psychology of Meaningful Verbal Learning*. New York: Grune/Stratton, 1963.
2. **Biggs J, Tang C.** *Teaching for Quality Learning at University: What the Student Does* (3rd ed.). Maidenhead, UK: Society of Research into Higher Education and Open University Press, 2007.
3. **Biswas G, Schwartz D, Bransford J.** Technology support for complex problem solving: from SAD environments to AI. In: *Smart Machines in Education: the Coming Revolution in Educational Technology*, edited by Forbus KD, Feltovich PJ. Menlo Park, CA: AAAI/MIT, 2001, p. 71–97.
4. **Bruner J.** *The Process of Education*. Cambridge, MA: Harvard University Press, 1960.
5. **Buzan T, Buzan B.** *The Mind Map Book*. London: BBC Books, 1995.
6. **Cañas AJ, Novak JD, Miller NL, Collado CM, Rodríguez M M, Concepción M, Santana C, Peña L.** *Confiabilidad de una Taxonomía Topológica Para Mapas Conceptuales* (online). http://cmc.ihmc.us/cmc2006Papers/cmc2006-p233.pdf [15 November 2010].
7. **Cañas AJ, Valerio A, Lalinde-Pulido J, Carvalho M, Arguedas M.** *Using WordNet for Word Sense Disambiguation to Support Concept Map Construction* (online). http://www.cs.indiana.edu/~avalerio/papers/2003spire.pdf [15 November 2011].
8. **Casti JL.** *Searching for Certainty: What Scientists Can Know*. New York: Morrow, 1990, p. 51.
9. **Chang K, Sung Y, Chen S.** Learning through computer-based concept mapping with scaffolding aid. *J Comp Assist Learn* 17: 21–33, 2001.
10. **Chi MT, De Leeuw N, Chiu MH, Lavancher C.** Eliciting self-explanations improves understanding. *Cogn Sci Multidiscipl J* 18: 439–477, 1994.
11. **Clariana RB, Koul R, Salehi R.** The criterion related validity of a computer-based approach for scoring concept maps. *Int J Instruct Media* 33: 317–325, 2006.
12. **Clement J.** Model based learning as a key research area for science education. *Int J Sci Educ* 22: 1041–1053, 2000.
13. **Cobb P.** Theories of knowledge and instructional design: a response to Colliver. *Teach Learn Med* 14: 52–55, 2002.
14. **Collins A, Joseph D, Bielaczyc K.** Design research: theoretical and methodological issues. *J Learn Sci* 13: 15–42, 2004.
15. **Conlon T.** Formative assessment of classroom concept maps: the reasonable fallible analyser. *J Interact Learn Res* 17: 15–36, 2006.
16. **DeHaan RL.** The impending revolution in undergraduate science education. *J Sci Educ Technol* 14: 253–269, 2005.
17. **DiCarlo SE.** Cell biology should be taught as science is practised. *Nat Rev Mol Cell Biol* 7: 290–296, 2006.
18. **Fellbaum C.** *WordNet–an Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.
19. **Fisher KM.** SemNet software as an assessment tool. In: *Assessing Science Understanding: a Human Constructivist View*, edited by Mintzes JJ, Wandersee JH, Novak JD. San Diego, CA: Academic, 2000.
20. **Gobert JD, Buckley BC.** Introduction to model-based teaching and learning in science education. *Int J Sci Educ* 22: 891–894, 2001.
21. **Goel AK, Chandrasekaran B.** Functional representation of designs and redesign problem solving. In: *Proceedings of the Eleventh International*

Joint Conference on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann, 1989, vol. 1.
22. **Grosslight L, Unger C, Jay E, Smith CL.** Understanding models and their use in science–conceptions of middle and high-school students and experts. *J Res Sci Teach* 28: 799–822, 1991.
23. **Harrison SH, Wallace JL, Ebert-May D, Luckie DB.** *C-TOOLS Automated Grading for Online Concept Maps Works Well With a Little Help From "WordNet"* (online). http://cmc.ihmc.us/papers/cmc2004–047.pdf [15 November 2010].
24. **Herl HE, O'Neil HF, Chung GKWK, Dennis RA, Lee JJ.** *Feasibility of an On-line Concept Mapping Construction and Scoring System* (online). http://www.eric.ed.gov/PDFS/ED424233.pdf [15 November 2010].
25. **Hibberd R, Jones A, Morris E.** *The use of Concept Mapping as a Means to Promote and Assess Knowledge Acquisition. CALRG Report No. 202*. Milton Keynes, UK: The Open University, 2002.
26. **Hmelo CE, Holton DL, Kolodner JL.** Designing to learn about complex systems. *J Learn Sci* 9: 247–298, 2000.
27. **Hmelo-Silver CE, Marathe S, Liu L.** Fish swim, rocks sit, and lungs breathe: expert-novice understanding of complex systems. *J Learn Sci* 16: 307–331, 2007.
28. **Klein DC, Chung GK, Osmundson E, Herl HE, O'Neil HF.** *Examining the Validity of Knowledge Mapping as a Measure of Elementary Students' Scientific Understanding. CSE Technical Report No. 557* (online). http://research.cse.ucla.edu/Reports/TR557.pdf [15 November 2010].
29. **Kuhn D.** Children and adults as intuitive scientists. *Psychol Rev* 96: 674–689, 1989.
30. **Kuhn D.** Thinking as argument. *Harvard Educ Rev* 62: 155–178, 1992.
31. **Luckie DB, Batzli JM, Harrison S, Ebert-May D.** C-TOOLS: concept-connector tools for online learning in science. *Int J Learn* 10: 332–338, 2003.
32. **Luckie D, Harrison S, Ebert-May D.** *Introduction to C-TOOLS: Concept Mapping Tools for Online Learning* (online). https://www.msu.edu/~luckie/CTOOLS.pdf [15 November 2010].
33. **Markham KM, Mintzes JJ, Jones MG.** The concept map as a research and evaluation tool: further evidence of validity. *J Res Sci Teach* 31: 91–101, 1994.
34. **National Research Council.** *Transforming Undergraduate Education in Science, Mathematics, Engineering, and Technology*. Washington, DC: National Academy, 1999.
35. **Novak J.** Concept mapping: a useful tool for science education. *J Res Sci Teach* 27: 937–949, 1990.
36. **Novak JD, Canas AJ.** *The Theory Underlying Concept Maps and How to Construct and Use Them* (online). http://cmap.ihmc.us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps.htm [15 November 2010].
37. **Novak JD, Gowin DD.** *Learning How to Learn*. New York: Cambridge, 1984.
38. **Pea R, Tinker R, Linn M, Means B, Bransford J, Roschelle J, His S, Brophy S, Songer N.** Toward a learning technologies knowledge network. *Educ Technol Res Dev* 47: 19–38, 1999.
39. **Ruiz-Primo MA, Schultz SE, Li M, Shavelson RJ.** Comparison of the reliability and validity of scores from two concept-mapping techniques. *J Res Sci Teach* 38: 260–278, 2001.
40. **Ruiz-Primo MA, Shavelson RJ.** Problems and issues in the use of concept maps in science assessment. *J Res Sci Teach* 33: 569–600, 1996.
41. **Safayeni F, Derbentseva N, Cañas A.** A theoretical note on concepts and the need for cyclic concept maps. *J Res Sci Teach* 42: 741–766, 2005.
42. **Seymour E, Gorogg DL, Burris G, Weiman C, Collins M.** *Undergraduate Science, Math and Engineering Education: What's Working? Hearing Before the Subcommittee on Research, Committee on Science, House of Representatives, One Hundred Ninth Congress, Second Session, March 15, 2006, Serial No. 109-40* (online). http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=109_house_hearings&docid=f:26481.pdf [15 November 2010].
43. **Suter L, Frechtling J.** *Guiding Principles for Mathematics and Science Education Research Methods. NSF Report 00-113* (online). http://www.nsf.gov/pubs/2000/nsf00113/nsf00113.pdf [15 November 2010].
44. **Taricani EM, Clariana RB.** Technique for automatically scoring open-ended concept maps. *Educ Technol Res Dev* 54: 65–82, 2006.
45. **Tarte G.** *Conéctate al Conocimiento: una Estrategia Nacional de Panamá Basada en Mapas Conceptuales* (online). http://cmc.ihmc.us/cmc2006Papers/cmc2006-p248.pdf [15 November 2010].
46. **Trigwell K, Prosser M, Ginns P.** Phenomenographic pedagogy and a revised approaches to teaching inventory. *High Educ Res Dev* 24: 349–360, 2005.
47. **Valerio A, Leake D, Cañas AJ.** *Automatic Classification of Concept Maps Based on a Topological Ontology* (online). http://cmc.ihmc.us/cmc2008papers/cmc2008-p086.pdf [15 November 2010].
48. **Van der Valk T, Van Driel JH, De Vos W.** Common characteristics of models in present-day scientific practice. *Res Sci Educ* 37: 469–488, 2007.
49. **Weimer M.** *Learner-Centered Teaching: Five Key Changes to Practice*. San Francisco, CA: Jossey-Bass, 2002.
50. **Wittrock MC.** Generative learning processes of the brain. *Educ Psychol* 27: 531–541, 1992.
51. **Wycoff S.** Changing the culture of undergraduate science teaching: shifting from lecture to interactive engagement and scientific reasoning. *J Coll Sci Teach* 30: 306–312, 2001.