9-20-2006

Characteristics of a cognitive tool that helps students learn diagnostic problem solving

Jared A. Danielson
Iowa State University, jadaniel@iastate.edu

Eric M. Mills
Iowa State University

Pamela J. Vermeer
Iowa State University

Vanessa A. Preast
Iowa State University

Karen M. Young
University of Wisconsin

See next page for additional authors

Characteristics of a cognitive tool that helps students learn diagnostic problem solving

Abstract
Three related studies replicated and extended previous work (J.A. Danielson et al. (2003),*Educational Technology Research and Development, 51*(3), 63–81) involving the Diagnostic Pathfinder (dP) (previously Problem List Generator [PLG]), a cognitive tool for learning diagnostic problem solving. In studies 1 and 2, groups of 126 and 113 veterinary students, respectively, used the dP to complete case-based homework; groups of 120 and 199, respectively, used an alternative method. Students in the dP groups scored significantly higher (p = .000 and .003, respectively) on final exams than those in control groups. In the third study, 552 veterinary students responding to a questionnaire indicated that the dP's gating and data synthesis activities aided learning. The dP's feedback and requirement of completeness appear to aid learning most.

Keywords
Cognitive Load Theory, Cognitive Tools, Diagnostic Problem Solving, Feedback, Gating

Disciplines
Higher Education | Science and Mathematics Education | Veterinary Medicine | Veterinary Pathology and Pathobiology

Comments
This accepted manuscript is an article from *Educational Technology Research and Development*, October 2007, 55(5); 499-520. *The final publication is available at Springer via http://dx.doi.org/10.1007/s11423-006-9003-8.*"

Authors
Jared A. Danielson, Eric M. Mills, Pamela J. Vermeer, Vanessa A. Preast, Karen M. Young, Mary M. Christopher, Jeanne W. George, R. Darren Wood, and Holly Bender

Characteristics of a Cognitive Tool that Helps Students Learn Diagnostic Problem Solving

Jared A. Danielson, Eric M. Mills, Pamela J. Vermeer, and Vanessa A. Preast

Iowa State University

Karen M. Young

University of Wisconsin, Madison

Mary M. Christopher and Jeanne W. George

University of California, Davis

R. Darren Wood

University of Guelph

Holly S. Bender

Iowa State University

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Abstract

Three related studies replicated and extended previous work (J.A. Danielson, et al, 2003) involving the Diagnostic Pathfinder (dP) (previously Problem List Generator [PLG]), a cognitive tool for learning diagnostic problem solving. In studies 1 and 2, groups of 126 and 113 veterinary students respectively used the dP to complete case based homework; groups of 120 and 199 respectively used an alternative method. Students in the dP groups scored significantly higher ($p = .000$ and .003, respectively) on final exams than those in control groups. In the third study, 552 veterinary students responding to a questionnaire, indicated that the dP's gating and data synthesis activities aided learning. The dP's feedback and requirement of completeness appear to aid learning most.

Cognitive Load Theory

Cognitive Tools

Diagnostic Problem Solving

Feedback

Gating

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Characteristics of a Cognitive Tool that Helps Students Learn Diagnostic Problem Solving

Problem solving is “the ability to combine previously learned principles, procedures, declarative knowledge, and cognitive strategies in a unique way within a domain of content to solve previously unencountered problems” (Smith & Ragan, 1999 p. 132). *Medical diagnostic problem solving* involves identifying disruptions of normal physiology in a human or animal patient; this is difficult because the biological systems involved are complex, with highly interactive sub-systems. Consistent with many domains (see for example, the review in (Bransford, Brown, & Cocking, 2000 pp. 31-50)), medical practitioners who have more elaborated structures of underlying rules and concepts out-perform those who have less elaborated structures (Bordage, 1994; Bordage & Lemieux, 1991). If the best problem solvers have coherent elaborated structures of rules/concepts, what approaches best help learners create such structures? One possibility is to use *cognitive tools* (Kozma, 1987; Salomon, 1988).

Cognitive tools use or model an expert approach to a given process. Just as traditional tools extend human muscle or sensory functions, cognitive tools extend cognitive, or symbolic function (Salomon, 1988). Drawing on research illustrating the connection between how individuals represent problems and their ability to solve them (e.g. Mayer, 1976; McGuinness, 1986; Pape & Tchoshanov, 2001; Zhang, 1997), Jonassen (2003) argued that cognitive tools (such as semantic network tools, expert systems shells, and systems modeling software) can promote learning of problem solving. The underlying premise is that people learn by constructing models/explanations of systems. However, there has been little exploration of how specific characteristics of cognitive tools might address specific learning requirements.

Cognitive load theory (CLT) (Sweller, 1988) is seen by a growing number of researchers as a tool to help designers conceptualize instructional design for complex learning (van

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Merriënboer, Kirschner, & Kester, 2003). CLT has been applied to a variety of complex instructional problems (Paas, Renkl, & Sweller, 2003; van Merriënboer & Ayres, 2005), and could apply to the design of cognitive tools. Applying CLT means designing instruction to account for the cognitive load imposed by the learning task. Effective instructional design respects *intrinsic cognitive load* (inherently irreducible cognitive activities), reduces e*xtraneous cognitive load* (activities that are irrelevant to schema acquisition) and maximizes *germane cognitive load*, (activities that facilitate schema acquisition and automation) (Paas et al., 2003). CLT can provide a lens to explore the theoretical merit of cognitive tools. First, germane cognitive load should be produced by software tools that embody relevant processes that are readily adoptable to the mind. Second, effectively harnessing the computer's ability to remember and organize and providing an environment in which learners can construct explanations of phenomena should reduce extraneous cognitive load.

The Diagnostic Pathfinder

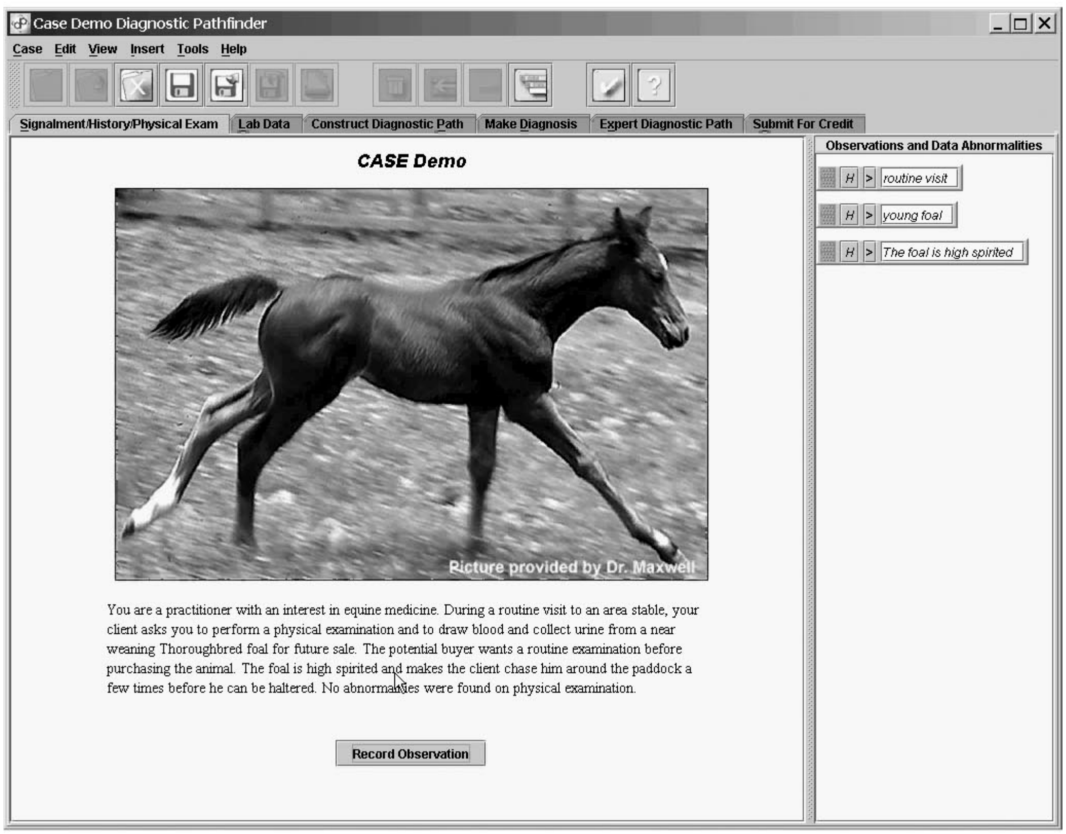
The Diagnostic Pathfinder (dP) was designed to help clinical pathology students learn to effectively synthesize laboratory data (such as blood chemistry, urinalysis, etc.) The dP is a cognitive tool in that its purpose is to model and facilitate a process for considering large amounts of interconnected information, with the hopes that the process might appeal to and be adopted by the mind. It is, however, unlike the tools described by Jonassen (2003) in that it imposes more structure upon the way students approach problems. The dP has been described in detail elsewhere (Danielson, 1999; Danielson, Bender, Mills, Vermeer, & Lockee, 2003), though a brief description follows. The dP presents veterinary cases, and allows learners to explain the data for a given case; it involves four core learning interactions: 1. collect historical data, 2.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

collect laboratory data, 3. synthesize data into a solution called a diagnostic path, and 4. compare the learner solution to an expert solution.

The interaction for collecting historical data (Figure 1) presents an image and a brief textual description of the patient's history. Observations gleaned here join the list in the right

Characteristics of a Cognitive Tool: Figure 1



hand column of the interaction. This list remains through the next several interactions, ultimately containing all data identified for consideration in the diagnostic rationale. After examining the history, learners identify abnormal data; for each laboratory test, they determine whether the

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

patient finding is normal or not, and name the abnormality if the finding is abnormal. If learners type an abnormality name incorrectly, they are prompted to try again, with help offered after the third try. Learners must name all abnormal data correctly before constructing their *diagnostic path* (the diagnostic rationale communicated in outline format). Data abnormalities join the growing data list initiated in the first interaction.

Next learners create a diagnostic path -- their diagnostic rationale in outline format. As opposed to the simple task of identifying data abnormalities, constructing a tenable diagnostic path is complex, requiring a knowledge of disease processes (referred to as *mechanisms*) and how those mechanisms disrupt normal physiology. Learners insert mechanisms and arrange history items, data abnormalities, and mechanisms using a drag and drop interface. Relative placement of items in the path communicates relationships among items; that which is above and to the left causes, or is supported by, that which is below and to the right. Many diagnostic paths become long and involved, such as the path represented by the fragment shown in Figure 2.

Characteristics of a Cognitive Tool: Figure 2

[M] DIABETES MELLITUS
NOTES:
a deficiency or resistance to insulin
[H] 10 year-old, spayed- female, Chihuahua
[H] progressive weight loss over the past 2 months
[M] INCREASED GLUCOSE IN THE BLOOD
  [D] hyperglycemia
    [M] WILL CAUSE FORMATION OF CATERACTS
      [H] cataracts in both lenses
  [M] GLUCOSE WILL EXCEED THE RENAL THRESHOLD
    [D] 3+ glucosuria
    [M] OSMOTIC DIURESIS
      [H] polydipsia, polyuria
      [D] range of minimal concentration
      [M] LOSS OF ELECTROLYTES (SIMPLE DEHYDRATION)
        [D] hyponatremia
        [D] hypochloridemia
          [D] hyperkalemia
        [D] ???triple phosphate crystalluria???

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

This path was created by the student who received the highest final grade in a his/her section of over 80 students, and demonstrates a mature understanding of relevant pathophysiology. Explicitly involving 6 body systems, the full path contained 35 mechanisms, 43 data abnormalities, 14 observations from the history, 91 direct causal/supportive relationships, and 359 relationships total (spanning multiple generations). In contrast, the diagnostic path represented by the fragment in Figure 3 was completed for the same case, section, and course as

Characteristics of a Cognitive Tool: Figure 3

- [M] CUSHINGS
  - [H] polydipsia, polyuria
    - [D] hyperproteinemia
  - [H] polyphagia
  - [H] progressive weight loss over the past 2 months
  - [H] now anorectic and lethargic
  - [H] depressed
  - [H] febrile
  - [H] polypneic
  - [H] thin
  - [H] moderately dehydrated.
  - [H] cataracts in both lenses
  - [H] moist rales over most of the lung fields.
  - [D] polycythemia
  - [D] lipemic plasma

Figure 2, but by a low-scoring student. This full path contained 39 data abnormalities, and 11 history observations–comparable to the path in Figure 2. However, in contrast to the previous one, this path contained only one causal mechanism, 50 direct causal/supportive relationships, and 51 direct or spanning relationships. Apparently, rather than engage with the case, the student guessed at the "Cushings" diagnosis, established a conceptual relationship among the first three data items, and moved the remaining history and data abnormality items into the path with little more thought. This path shows that the dP allows, but does not require, thoughtful synthesis of the content.

Once learners have created and submitted a diagnostic path, they receive feedback in which the learner's diagnostic path is displayed alongside the instructor's diagnostic path.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

If the dP promotes learning of diagnostic problem solving, there are a finite number of potential reasons. Because the dP uses cases, implied within the following discussion is the assumption of a case-based context. Danielson et al (2003) hypothesized that two dP characteristics promoted its success—first, the dP requires that the sub-elements of the domain be identified prior to the data synthesis activity, and second, both the learner and instructor approach identical problems, produce solutions independently using the same process, and compare the results in an identical format. Here we extend those hypotheses. We will consider the instructional strategies embodied in the dP in terms of two broad categories: 1. gating activities, and 2. data synthesis activities.

In the context of this paper, *gating* refers to the characteristics of the dP that impose certain performance criteria as the conditions that permit learners to move from one interaction to another. The dP provides gating for three criteria: accuracy, sequence, and completeness. Accuracy is enforced with data abnormality naming and spelling. Sequence is enforced twice; learners must identify all abnormal laboratory data before creating a diagnostic path, and they must complete their path before viewing the expert path. Completeness is enforced once; learners must include all data abnormalities in the diagnostic path.

Data synthesis might occur as a cognitive process at any point as the learners engage with the dP. However, we posit that most synthesis occurs while manipulating the diagnostic path, and later while viewing feedback. Several characteristics of the data synthesis activities might affect learning. First, a specific format (outline) is used to communicate causal and supportive relationships. Second, a multiple-select, drag and drop mechanism is used for manipulating the data. Third, feedback is presented immediately, and in a format identical to that used by learners to communicate their own diagnostic rationale.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

The Experiments

We intended the following experiments first, to validate a prior study (Danielson et al., 2003) which showed benefit in dP-use, and second, to clarify relationships among specific characteristics of the software and learning outcomes. A brief discussion will accompany each of the first two experiments. The discussion following the third experiment will synthesize the findings of all three experiments. All three studies employed the same curricular intervention–the dP. Learning impact is the dependent variable of interest for the first two studies. Questionnaire data collected during the first two studies are presented as part of the third study.

Experiment 1

Method

Participants. All sophomore veterinary students who were enrolled at the University of California, Davis College of Veterinary Medicine during 2002 and 2004 participated. All students were admitted based on identical criteria. The 2002 and 2004 classes comprised 120 and 126 students, respectively, and were approximately 95% California residents and 82% women.

Procedures. During their second year in veterinary school, all students participated in the UC-Davis core clinical pathology course. This 3.3 credit hour course met for 11 weeks and involved 2 hours of lecture and 3 hours of case discussion or laboratory each week. In 2002 students were assigned individually to complete a total of 50 cases (4-6 per week, representing the course's core concepts) using a paper-based method that was similar to creating a diagnostic path. Specifically, students were to identify laboratory abnormalities, group them according to mechanism and severity, develop a prioritized list of most likely diagnoses, and develop a list of additional tests needed for a final diagnosis. Homework was not collected or graded, but in the weekly case discussions, randomly selected groups of students were required, in front of the

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

class, to give spontaneous graded presentations involving cases that had been assigned as homework. The course instructor reported that this requirement led most students to complete the homework. All students completed a final exam.

In 2004 the same instructor taught the same course, using the same overall format, with three differences. First, students were assigned 53 cases instead of 50. Of those cases, 43 were assigned for homework, 36 to be completed individually (using the dP), and 7 to be completed in small groups. The ten additional cases were not required for course credit. Second, whereas in 2002, the grade was distributed as 40% midterm, 50% final exam, 5% case presentations, and 5% group project/cases, in 2004 the grade was distributed evenly among cases/quizzes, midterm, and final. Third, students did assigned cases using the dP, not the previous paper-based approach.

Instruments. The core pathology course final exam was used to measure learning impact. The exam consisted of four clinical cases, each containing a brief history of the patient and a list of laboratory results. A series of short-answer, multiple choice, or true-false questions accompanied each case. To answer these questions, the students had to analyze the case data as practiced during class. Example items are found in Figure 4. Final exams were administered on paper in the student laboratory, were equivalent from year to year, and were collected to prevent students from passing them on to subsequent classes.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

CASE 1.

You are called to a ranch in the California foothills to examine a 10-day-old Hereford calf that is severely depressed and unable to rise. The calf has had runny diarrhea for at least 2 days and its perianal area is matted with watery feces. Dehydration is moderate to marked. You obtained a blood sample for serum chemistry (Day 1). You treat the calf with an IM injection of pen/strep, and fluids (sodium chloride with added bicarbonate and 5% dextrose). A second blood sample is obtained and analyzed 24 hours later (Day 2).

CHEMISTRY (SERUM)

Test Name	**Day 1**	**Day 2**	**Ref. Int.**	**Units**
T. PROT	7.3	6.8	6.8-8.6	g/dl
ALB	3.8	3.6	3.0-4.3	g/dl
GLOB	3.5	3.2	3.0-3.5	g/dl
TBIL	0.1	0.1	0.0-0.1	mg/dl
AST	96	100	0.0-150.0	mU/ml
CK	184	508	0.0-200.0	mU/ml
CA	9.2	9.8	8.9-11.9	mg/dl
P	15.1	10.7	5.0-8.0	mg/dl
GLU	28	266	33.0-66.0	mg/dl
ALP	409	377	27-107	mU/ml
GGT	119	100	15.0-39.0	mU/ml
BUN	113	52	8.0-23.0	mg/dl
CREA	4.7	2.5	0.5-1.3	mg/dl
NA	135	129	133.0-143.0	mEq/l
CL	108	96	93.0-103.0	mEq/l
K	7.1	3.0	3.9-5.2	mEq/l
TC02	11	17	19.0-34.0	mEq/l
AG	23.1	19	12.0-22.0	mEq/l

1. What is the **primary** pathophysiologic mechanism of the metabolic acidosis? (2 pts)
2. What are two (2) possible pathophysiologic mechanisms of the hypoglycemia on Day 1? (4 pts)
   a.
   b.
3. Explain the pathophysiologic mechanisms leading to the abnormal potassium values on Days 1 and 2. (4 pts)
4. Which laboratory value(s) support previous colostral ingestion by this calf? (2 pts)
5. What is your interpretation of the mild increase in CK activity on Day 2? (2 pts)

Results

We used a two-tailed independent samples *t*-test to compare final exam scores, with a specified significance level of .05. Mean final exam scores for 2002 and 2004 were 85.0 (n=119)

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

and 90.1 (*n*=126) respectively. Scores for 2004 were significantly higher than scores for 2002, $t = 5.479$, $p = .000$, effect size (Cohen's d) = .70.

Discussion

Students using the dP to complete their homework outperformed students who used a similar paper-based process. The effect size is respectable by conventional standards (e.g. Pedhazur & Schmelkin, 1991) and is of a magnitude that would be meaningful to students (5 full percentage points). Because the study was not experimental, factors other than dP-use might have contributed to the difference in final exam scores. However, because students were recruited to the school using identical procedures and policies, and from the same populations, it is unlikely that differences in score were due to systematic pre-existing differences between groups. Also, whereas factors other than dP-use changed slightly between the years (i.e., a slightly different number of cases, and slight group work differences), the most notable difference between the implementations was dP-use. Therefore the most likely cause of the improvement in final exam scores was dP implementation.

Experiment 2

Method

Participants. All students enrolled in the sophomore curriculum at the University of Wisconsin - Madison School of Veterinary Medicine from 2001 to 2004 participated. The veterinary school admitted all students using identical entrance requirements, and drawing from the same populations; 75% were female, and 81-88% were recruited from Wisconsin (depending on the year). These classes (2001-2004) comprised 79, 79, 74, and 80 students, respectively.

Procedures. During their second year in veterinary school, all students took UW-Madison's core clinical pathology course. The course involved 45 hours of lecture (mixed

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

didactic material and case presentation), 45 hours of case-based laboratory discussion, 8 case-based assignments, and 3 large exams. Beginning in 2002, one of the course instructors began to offer a clinical pathology elective that was limited to students concurrently enrolled in the core course. Each week, participants in this pass/fail elective used the dP to complete several cases and met for one hour with the instructor to discuss the week's cases. Students received credit for attending and completing all assigned cases. Solution quality was not evaluated. In 2002, enrollment in this course was limited to six students who were selected for varying levels of academic standing and technology experience. Beginning in 2003, the elective was made available to all students enrolled in the core course. Forty two students (57% of the core course) took the elective in 2003, and 65 (81% of the core course) took it in 2004. Students in the elective completed 10 cases in 2002, 45 cases in 2003, and 30 cases in 2004. There were few assigned cases in 2002 because that year was intended to be a limited pilot of the dP. The 45 cases in 2003 were meant to provide a rigorous and comprehensive reinforcement of the material covered in the core course. In 2004 the number of cases was decreased to 30 to include a representative set of species and disease conditions without overwhelming students.

Instruments. The core pathology course final exam at UW-Madison, used to measure learning impact, has two parts. First, students view two glass slides, and draw conclusions regarding the related case(s). Second, students answer multiple choice questions based on data from 22 cases. Figure 5 shows representative final exam items. Final exams were equivalent from year to year and were retained so students would not pass them on to subsequent classes.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Characteristics of a Cognitive Tool: Figure 5

Answer the following questions using the clinical pathology data from dogs A through D. Letters may be used more than once.

	A	B	C	D	Ref Int
PCV %	26	31	15	58	**37-55**
MCV fL	57	66	85	72	**60-77**
MCHC g/dL	32	34	28	33	**32-36**
TP g/dL	4.8	7.1	7.4	7.9	**6.0-7.5**
Plasma color	straw	icteric	icteric hemolyzed	icteric	**straw**
RBC Morph	normal	normal	poly 3+	normal	
Chemistry:					
ALT U/L	46	1840	124	76	**< 79**
AST U/L	32	1324	136	204	**< 44**
ALP U/L	115	431	355	977	**< 166**
T Bili mg/dL	0.4	5.8	5.4	15.2	**<0.6**
CK U/L	735	129	146	130	**7-203**
Albumin g/dL	2.0	3.7	3.9	4.3	**2.5-4.0**
Chol mg/dL	79	213	288	512	**111-290**
BUN mg/dL	6	21	18	42	**8-25**
NH3 ug/dL	172	75	43	64	**20-90**
Urinalysis:					
Urine color	light yellow	dark gold	red-gold	dark gold	
Spec Grav	1.005	1.020	1.025	1.050	
Heme	negative	negative	++	negative	
Bilirubin	negative	++	++	+++	
Sediment	ammonium biurate crystals	negative	negative	negative	

(1) 14. Which dog has laboratory data indicative of prehepatic jaundice? _____

(1) 15. Which dog has laboratory data indicative of hepatic jaundice? ____

(1) 16. Which dog has laboratory data indicative of posthepatic jaundice? _____

(1) 17. Which dog has liver disease characterized by decreased hepatic synthetic activity? ____

(1) 18. The serum chemistries in this dog indicate that a large number of hepatocytes are damaged but that hepatic synthetic and clearance functions are adequate. ____

Results

We used a two-tailed independent samples *t*-test with an alpha level of .05 to compare mean final exam scores. Across all years, scores for students participating in the dP elective were combined and compared to scores for all students not participating in the dP elective. The mean final exam score for students taking the dP elective was 87.0 (n = 113), and the mean score for

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

those not taking the elective was 84.7 (n = 199). This difference was significant, t = 2.96, p = .003, effect size (Cohen's d) = .36.

Discussion

Students who participated in the dP elective outperformed those who did not on the final exam. However, this study has several weaknesses, the largest being that students in the experimental groups (dP elective) spent more formal class time engaged in tasks relevant to the final exam than students in the control groups. Hence, students taking the elective should have out-performed the other students, regardless of the instructional strategy employed. An additional weakness of this study was that students were self-assigned to groups. Perhaps above-average and/or over-achieving students systematically selected the elective. Conversely, students who were performing poorly and wanted to help their odds at success in the core course might have systematically selected the elective. In short, there may be preexisting systematic differences between treatment groups. The effect size, by conventional interpretations of Cohen's d, (Pedhazur & Schmelkin, 1991) is modest, but because it represents slightly more than two full percentage points on the final exam, students would likely find it meaningful. Despite the study's limitations, it supports the idea that time spent working cases using the dP benefited learning of diagnostic problem solving. It does not, however, permit us to discuss how dP-use might compare to other instructional strategies requiring similar time and effort.

Experiment 3

Experiment 3 provided an indirect measure of the dP's effectiveness as indicated by student responses and permitted us to link students' perceptions of the dP with specific characteristics of the software.

Method

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Subjects. Between the Spring of 2002 and the Fall of 2004, the dP was used to teach eight semesters of clinical pathology at five colleges of veterinary medicine, including Iowa State University, The University of Wisconsin - Madison, The University of California - Davis, The University of Guelph (Ontario College of Veterinary Medicine), and Virginia Tech. A total of 640 students participated in these classes, 70% female and 30% male. Other than Guelph's small pilot course of 6 students, the smallest class contained 42 students and the largest contained 126. The average class size was 80. All students were asked to complete a questionnaire regarding their experience; 552 complied. By class, this response rate varied from 54% to 100%, with the overall response rate being 86% (calculated by taking the total number of respondents across all institutions [552] and dividing by the total number of learners who used the dP [640]). Table 1 shows the number of learners at each participating site (by year), the number of cases assigned per student, and the percentage of students at each site who responded to the questionnaire.

Table 1

Learner, case use, and questionnaire response information across participating sites

Institution/Year	Number of Assigned Cases	Number of Students	Questionnaire Response Rate
Virginia Tech 2002	91	89	99%
Iowa State 2002	91	96	100%
Iowa State 2003	91	97	70%
Wisconsin 2003	45	42	100%
Wisconsin 2004	30	65	100%
California Davis 2003	6	120	54%
California Davis 2004	36	126	98%

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Guelph 2003	12	6	83%

Procedures. Students at the participating institutions used the dP to complete case-based homework assignments and prepare for exams. Curricular approaches at the institutions varied, though all the institutions except for Guelph employed a traditional medical curriculum in which students received lectures in clinical pathology interspersed with case-discussion laboratories. Guelph used a curriculum that mixed several approaches, including lecture, collaborative learning, and problem based learning. As seen in Table 1, the number of cases assigned to each student using the dP varied from institution to institution, with the smallest number being 6 at UC-Davis in 2003, and the greatest number being 91 at two schools. Instructors chose cases and case quantities to accommodate their particular curricular needs and contexts. In 2003 the dP was piloted at UC-Davis and at Guelph, explaining the lower numbers of cases and student participants, respectively. Although those students had less exposure to the dP than the students in the other groups, their data have been included in the analysis because both represent reasonable implementations of the tool; it is likely that an instructor might ask students to work only a dozen or fewer cases during a semester. Also, it takes between 30 and 90 minutes to complete one dP case, so even students who did only six cases had several hours of exposure to the tool. Students completed the questionnaire near the end of the semester in which they used the dP. To reduce potential for positive response bias, surveys were administered anonymously.

Instruments. The full questionnaire used in the study is found elsewhere (Danielson et al., 2003), and was designed to determine the students' perceptions of the dP's clarity (or usability), feasibility, and learning impact. We will focus on the items having to do with learning impact --

This is a manuscript from an article from *Educational Technology Research and Development,* 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

items 6, 7, 12, 16, and 17 (see Table 2). Items measuring enjoyment or usability might indirectly indicate learning gains, but are closely tied to software feasibility factors, such as computer bugs and network problems, so we will not discuss them here. We will also explore responses to open-ended questions, which clarify the Likert responses. The open-ended questions are: 23. For questions above that you ranked particularly negatively, please indicate why here, 24. What are the things you like most about using the dP?, 25. What are the things you like least about using the dP?, 26. What would you change about the dP if you could?, 27. Any additional comments you'd like to make about the dP:, and 28. If you used the dP for less than 20% of your cases, why did you choose not to use it?

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Table 2

Mean responses by item number, institution, and year

Questionnaire item:	VT 02 *n* = 86	ISU 02 *n* = 96	ISU 03 *n* = 68	UW 03 *n* = 42	UW 04 *n* = 65	UC 03 n=55	UC 04 *n* = 124	UG 03 *n* = 5
6. Using the dP me account for more lab data than I otherwise would have accounted for. less same more 0 1 2 3 4 5 6 7 8 9 10 NA	8.7	8.5	8.4	8.1	8.2	7.7	8.6	8.6
7. Using the dP made my problem lists more precise than they would have been otherwise. less same more 0 1 2 3 4 5 6 7 8 9 10 NA	8.7	8.1	8.3	7.8	8.3	7.2	7.8	8.0

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

12. The dP makes doing my Clinical Pathology homework more worthwhile than similar paper-based assignments. definitely not absolutely 0 1 2 3 4 5 6 7 8 9 10 NA	8.6	7.9	8.6	6.8	7.4	6.1	7.3	8.8
16. Using the dP helps me to organize my thoughts about a case. definitely not absolutely 0 1 2 3 4 5 6 7 8 9 10 NA	8.4	8.0	8.6	8.3	8.2	6.8	7.8	8.8
17. Using the dP makes understanding clinical pathology…. harder easier 0 1 2 3 4 5 6 7 8 9 10 NA	8.3	8.2	8.9	8.5	8.4	6.9	7.9	8.3

Note. Copyright 2002, 2005 by Jared A. Danielson, Holly S. Bender, Pamela J. Vermeer, Eric M. Mills

VT = Virginia Tech; ISU = Iowa State University; UW = University of Wisconsin, Madison; UC = University of California, Davis; UG = University of Guelph

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Results

Likert Items. We calculated descriptive statistics for responses to survey Likert items across all respondents by institution. Table 2 reports mean student responses to the Likert items by institution and year. Students at all institutions generally indicated that dP-use enhanced learning. Item 6 was meant to measure perceived completeness of the solution. Students felt that using the dP led them to account for more data. Item 7 was meant to measure whether or not students felt that dP-use affected the precision of their diagnostic rationale. Students indicated that dP-use made their rationale more precise. Items 12 and 17 were meant to show the dP's overall perceived value to the students as a learning tool. Students generally found the learning value to be high. Finally, item 16 was meant to indicate the dP's perceived effect on the students' ability to organize data. Students indicated that the dP enhanced this ability.

Open-Ended Responses. Two raters analyzed open-ended responses. Though many of the responses were easily interpreted, the analysis was complicated because some comments could be meaningfully interpreted only by someone having a thorough familiarity with the software. For example, in response to question 24 (What are the things you liked most . . . ?) one student wrote "Repeats, memorizing". To interpret this response raters had to know how/when the software repeats, and what aspects of the software encourage/require memorizing. Because of these context-specific complexities, we used a consensus approach for data analysis. Initially, one of the primary researchers (Rater 1) used a generative, or open-coding process to reveal broad trends in the data. A research assistant (Rater 2) then codified the responses according to those trends, and recorded them in a database. Throughout the process, the raters frequently met informally to discuss ambiguous responses. During these meetings, the raters identified new

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

categories, and merged several existing categories. Rater 1 then reviewed each response as coded by Rater 2. Because we used a consensus approach, we did not calculate inter-rater agreement.

The resulting categories, in order of most responses to fewest, were as follows: (a) general response, (b) ease/efficiency of thinking, (c) ease of use/convenience, (d) requirement that all data abnormalities be typed by hand and spelled correctly, (e) requirement of completeness, (f) process of manipulating data in the diagnostic path/diagnostic path format, and (g) expert feedback. The categories that emerged from this analysis included all but two of the dP's gating and data synthesis characteristics identified previously. Those two, sequence and diagnostic path format, did not emerge through the open-coding process as strong themes in the data. However, to aid in the discussion of these factors, we analyzed the responses post hoc to determine response frequency in these categories. We coded all responses as (a) positive, (b) mixed (both positive and negative, or in some way unclear), or (c) negative and/or suggested improvements. We then counted each response-type by category. Because the Likert responses were largely positive, we might expect positive responses to open-ended questions. It is also reasonable, however, to expect a disproportionately high number of critical comments because more open-ended questions were designed to elicit critical responses (items 23, 25, 26 & 28) than positive ones (item 24). Table 3 shows response frequency and percent of total respondents for each category, whether the category emerged during open coding or was included in the post hoc analysis, and the number of semester/institution implementations (called "# of Groups") in which at least one student agreed with the comment.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Table 3

Results of Open-Coding Analysis: Number of Responses by Category and Number of Groups (Institution/Semester) Providing the Response at Least Once

		Positive		Mixed		Negative/Suggestions	
Category of Response	Analysis Used	Number (%)	# of Groups	Number (%)	# of Groups	Number (%)	# of Groups
1. General Response	Open Coding	242 (44%)	8	153 (28%)	8	19 (3%)	6
2. Ease/Efficiency of Thinking	Open Coding	178 (32%)	8	2 (0%)	1	4 (1%)	2
3. Ease of Use/Convenience	Open Coding	115 (21%)	8	13 (2%)	5	21 (4%)	7
4. Requirement of Correct Spelling	Open Coding	43 (8%)	7	10 (2%)	5	71 (13%)	8
5. Requirement of Completeness	Open Coding	70 (13%)	6	10 (2%)	3	9 (2%)	4
6. Diagnostic Path Manipulation	Open Coding	46 (8%)	6	4 (1%)	2	39 (7%)	7
7. Expert Feedback	Open Coding	68 (12%)	7	15 (3%)	5	0 (0%)	0

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

8. Sequence	Post Hoc	8 (1%)	4	1 (0%)	1	1 (0%)	1
9. Diagnostic Path Format (Outline)	Post Hoc	5 (1%)	2	1 (0%)	1	10 (2%)	3

Note. Percents were rounded to the nearest whole number. The Number heading refers to the number of comments total, in the category. The percent refers to the percent of total respondents (552) represented by the number of respondents in the category.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

1. General Response Category. Of the 552 total respondents, 414 made comments related to the learning effects of the software (242 positive, 153 mixed, and 19 negative). In general, this pattern of responses mirrors the responses to Likert items.

2. Ease/Efficiency of Thinking. Many respondents referred to the software's general effect on the way they thought about the cases (178 positive, 2 mixed, and 4 negative). Positive comments ranged from vague (e.g. "a great way to get us to think clinically. . .") to more specific (e.g. ". . . I could organize my thoughts in a logical manner.") Mixed comments suggested that the dP was useful, but did not change the way learners thought about pathology, or that it was inconsistent with their way of thinking. Negative responses were that the dP did not help thinking, or that learning results did not justify the effort required to use the dP.

3. Ease of Use/Convenience. Comments in this category were 115 positive, 13 mixed and 21 negative. Positive comments included specific aspects of the software that made study more convenient, as well as generic statements such as "saves time", or "easy." Mixed comments often showed a positive change in the students' attitude toward the software over time, as their ability to use it increased. The negative comments usually involved factors inherent in working with computers, or inadequacies in the specific computer a given learner was trying to use.

4. Requirement that all data abnormalities be typed by hand and spelled correctly. Comments about this requirement were 43 positive, 10 mixed, and 71 negative. Most positive comments emphasized the benefit of learning the vocabulary. Most negative comments, as well as the negative aspects of the mixed comments, had to do with the repetitive nature of typing the same abnormality name many times (once for each of several cases), and not with the basic requirement of getting the names right.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

5. Requirement of completeness/sequence. Eighty eight students commented on the aspects of the software that require the learner to consider all laboratory data when constructing a diagnostic path, or to classify all data as being normal or not (70 positive, 10 mixed, 9 negative). One characteristic positive comment was, “It made me analyze each and every piece of data, something I probably normally would not have done.”

6. Manipulating data in the diagnostic path/format of diagnostic path. Eighty nine students commented on the process of manipulating data in the diagnostic path and/or the diagnostic path format -- 46 positive, 4 mixed, and 39 negative. Many of the positive comments had to do with the ease of manipulating the data in the interface as opposed to using paper. Many negative comments stated that learners could not see all the relevant data at once.

7. Expert feedback. Eighty three students commented on the feedback they received regarding their rationale in the form of the expert diagnostic path. Sixty eight of these comments were strictly positive. The fifteen students giving mixed comments all found the expert feedback useful, but wanted that feedback altered or expanded in some way. Several students also requested access to the expert rationale without having to complete the case first. None of the comments suggested that students did not want expert feedback.

Post Hoc Analysis - Sequence and Diagnostic Path Format. During the open coding process, we included comments that had to do with sequence in the “requirement of completeness” category, so “sequence” is a sub-section of 5, above. Ten learners specifically mentioned the requirement of considering data in a certain sequence (8 positive, 1 mixed, 1 negative.) During the initial coding we included comments regarding diagnostic path format as a sub-part of section 6, above. Of those, sixteen had to do with the outline format (5 positive, 1

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

mixed, 10 negative). Negative comments often requested some other format, such as a concept map or flow-chart.

Discussion

Prior to presenting the experiments, we identified six characteristics of the dP that might contribute to its effectiveness. To what extent are those hypothesized reasons supported by the data? We will explore that question in terms of the empirical studies and student responses to the surveys. Where applicable, we will also relate the characteristic to CLT and/or cognitive tools.

Support in Empirical Studies – General Observations

If the dP was more effective than the control strategies, it is reasonable to conclude that this was due to one or more features of the software that were not present in the control strategies. All of the dP features mentioned above were unique to the dP, except for the specific format used to communicate the diagnostic rationale (outline), which was also used in the paper-based control groups. Therefore, the empirical studies provide no evidence that using an outline format to communicate diagnostic rationale improves diagnostic problem solving ability. (Note that this does not constitute evidence that the outline format does not work; it simply did not contribute to the observed differences in this case.) The empirical studies supported all other identified software characteristics equally.

Support by Characteristic

Table 3 shows that student support for each strategy varied. In Table 4, we have categorized the level of student support for each strategy as High, Medium, Low, Unclear, or Mixed.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Table 4

Instructional Intervention by Treatment and Student Perception

	Treatment		
Instructional Intervention	Control	dP	Student Perceived Benefit
Feedback	Delayed	Immediate	High
Completeness	Not enforced	Enforced	High
Diagnostic Path Manipulation	Paper/pencil	Drag/drop	Medium
Sequence	Not enforced	Enforced	Unclear
Accuracy of Spelling	Not enforced	Enforced	Mixed
Diagnostic Path Format	Outline	Outline	Low

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Feedback. The dP offers immediate feedback in the form of a static expert diagnostic path. The alternative treatments typically offered feedback or some other form of explanation as well, but in the form of delayed case discussions. (Note that dP courses also frequently used case discussions.) The immediate feedback produced the most unqualified support from the students. Twelve percent of all respondents specifically mentioned this feature as being beneficial, with no negative comments. This led us to describe this feature's perceived benefit as "high."

Requirement of Completeness. The dP requires not only that all data abnormalities be identified, but that they appear in the diagnostic path. Although students can fail to explain the data correctly, or can omit relevant history findings, they have to use all the data abnormalities they have identified in their diagnostic path. Students overwhelmingly identified this aspect of the dP as being impactful, both through Likert item 6 and through the open-ended responses. In terms of open-ended responses, we have categorized this feature's perceived impact as being "high" because 13% of all respondents identified it as being positive, and because positive comments outnumbered negative ones by a ratio of nearly 8 to 1. This requirement, in light of CLT, seems quite defensible. A task with high intrinsic cognitive load can be made doable only by first automating (accommodating to schema) the related sub-tasks/knowledge. Explaining the relationships among data items while simultaneously attempting to ensure that the items are identified and properly classified and named, would increase overall cognitive load. Conversely, dividing the task such that learners identify most relevant data elements prior to data synthesis might reduce the cognitive load imposed by the synthesis task.

Manipulation of Diagnostic Path. The dP allows students to arrange and rearrange data elements, either individually, or in contiguous or non-contiguous groups, using drag and drop functionality. This functionality is inherent to the dP's computer-based environment and was not

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

available to students in the control groups of the quasi-experimental studies. Note that students in control groups could have used software tools that allowed for drag and drop, etc., for constructing their diagnostic rationales but most used pencil/pen and paper. Furthermore, other computer-based tools would not have automatically produced the column of data abnormalities and history findings as the dP does. Student comments (8% of all respondents) generally support the idea that this aspect of the dP was beneficial. Additionally, many comments endorsing the software's ease of use/convenience referred to data manipulation. Nonetheless, many students also indicated that this process was difficult or problematic (7% of all respondents). As indicated earlier, learners often attributed this difficulty to not being able to see/manage all the data at once. Given the inherent limitations of computer-based data presentation, and the amount of data being considered, it is not clear that some other mechanism for manipulating the data would be easier. Therefore, because many indicated that this feature was beneficial, and this coincides with the overwhelmingly positive response to the software's overall convenience/ease of use (see Table 3), and because the data synthesis task is likely to be considered difficult by many students regardless of the mechanism used for manipulating the data, we indicated that the contribution of this element to the software's perceived success was "medium." Clearly, the tool would not be very useful without some way to manipulate the data, so excluding the data synthesis process altogether would certainly not improve it. However, other formalisms for representing or manipulating data, such as concept maps (Novak, 1990) might work better for some students.

Although details of representation are in question, theory supports the general characteristic of the dP that allows students to manipulate and organize data in terms of relationships among concepts. In the framework of CLT, a tool that facilitates the mechanics of data manipulation would reasonably be expected to reduce the extraneous cognitive load

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

associated with the data synthesis task. (Clicking and dragging involves less cognitive overhead than erasing, writing between the lines, or re-writing.) Furthermore, this interaction most closely epitomizes the essence of cognitive tools; here learners construct their own models of the system.

Sequence. Van Merriënboer, et al (2003) indicate that requiring novices to complete one problem solving phase before moving to another could aid learning by reducing extraneous cognitive load, though they knew of no empirical research that explicitly supported such a strategy. Students did not mention this aspect of the dP much, as seen in Table 4, though the mention it did receive was positive. Maybe students took this feature for granted, thus not mentioning it, or maybe "completeness" comments implied "sequence" as well. It is also possible that students did not feel that sequencing activities contributed to their learning success.

Naming/Spelling Accuracy. Unlike the other features, the spelling requirement received nearly twice as many negative as positive comments. The many comments indicating that this feature was simply "too much", as opposed to being entirely wrong, suggest a potential manifestation of the *expertise reversal effect* wherein requiring students to actively engage with information that has already been accommodated to schema hinders learning, presumably because of the extraneous cognitive load imposed by elaborating information that has already been automated (Kalyuga, Ayres, Chandler, & Sweller, 2003). In our context, one who knows that the name "anemia" is associated with a low red blood cell count might experience less cognitive load while synthesizing anemia with other data than one who does not. However, making a learner spell "anemia" after it is learned may simply impose extraneous cognitive load.

Implications for the Design of Software Tools

The fact that these studies involved authentic learning problems and settings argues for their external validity. Nonetheless, unlike controlled laboratory settings, ours did not allow us to

This is a manuscript from an article from *Educational Technology Research and Development,* 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

isolate variables. Some combination of the dP characteristics described above made the software effective, but we cannot say definitively what that combination is. What, therefore, can we infer?

1. Harnessing internal motivation through gating strategies.

Cognitive tools literature emphasizes the process of manipulating data and the creation of the resulting representation of the problem space. Our experience suggests that additional gating features can be useful when combined with the concept of cognitive tools. Specifically, requiring that learners' explanations or models meet some standard of "completeness" appears to have been useful, with a possible additional contribution from the sequential gating of activities. Of potential concern, such a feature might simply create dependency on the tool. However, this seems not to have been the case here because improvement was detected in a different environment, a case-based final exam administered on paper.

The completeness requirement appeared to motivate effort. Students strongly agreed with the statement, "using the dP made me account for more laboratory data than I otherwise would have accounted for." They specifically mentioned that this aspect of the software made them do things like "account for", "analyze" and "not ignore" relevant data. In fact, however, the dP did not make students think about data other than very superficially, much less analyze it, as illustrated by Figure 3. Students were only required to include all identified data in their paths. Perhaps discrepancy reduction (Reeve, 2005) motivated the learners–they knew that a complete solution would include all data organized in some coherent fashion (even if that meant explicitly identifying a given datum as irrelevant). Applying a minimal standard of completeness may have inspired learners to self - impose a higher standard of coherence.

2. Data Manipulation and Feedback

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

While a great deal of research and theory already favor the use of feedback, our experience supports the use of feedback specifically in the context of cognitive tools. If cognitive tools are based on the premise that "humans learn more from constructing and justifying their own models of systems than from studying someone else's" (Jonassen, 2003 p. 372), we would additionally suggest that our participants appeared to learn more from constructing and justifying their own models of systems, and then studying someone else's, than from simply constructing/justifying their own models. It might be argued that the process of comparing ones model to an expert model and then self-assessing is a form of justification; that being so, it appears to be an effective form of justification.

Limitations and Future Directions

The primary limitation of the research described herein is that although the overall benefit of the software was tested empirically, the evidence supporting (or not) various software characteristics or combinations of characteristics is based on theory and student report rather than on empirical evidence. Future studies should involve experimental designs that isolate and test the effectiveness of individual instructional characteristics or combinations thereof.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

References

Bordage, G. (1994). Elaborated knowledge: A key to successful diagnostic thinking. *Acad Med, 69*(11), 883-885.

Bordage, G., & Lemieux, M. (1991). Semantic structures and diagnostic thinking of experts and novices. *Acad Med, 66*(9 Suppl), S70-72.

Bransford, J. D. E., Brown, A. L. E., & Cocking, R. R. E. (2000). *How People Learn: Brain, Mind, Experience, and School. Expanded Edition*. District of Columbia: National Academies Press, 2102 Constitution Avenue N.W., Washington DC 20055

National Academy of Sciences - National Research Council, Washington, DC. Commission on Behavioral and Social Sciences and Education.

Danielson, J. A. (1999). *The Design, Development and Evaluation of a Web-based Tool for Helping Veterinary Students Learn How to Classify Clinical Laboratory Data.* Unpublished doctoral dissertation, Virginia Tech, Blacksburg.

Danielson, J. A., Bender, H. S., Mills, E. M., Vermeer, P. J., & Lockee, B. B. (2003). A Tool for Helping Veterinary Students Learn Diagnostic Problem Solving. *Educational Technology Research and Development, 51*(3), 63-81.

Jonassen, D. H. (2003). Using Cognitive Tools to Represent Problems. *Journal of Research on Technology in Education, 35*(3), 362-379.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist, 38*(1), 23-31.

Kozma, R. B. (1987). The Implications of Cognitive Psychology for computer-based learning Tools. *Educational Technology, 40*(11), 20-25.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Mayer, R. E. (1976). Comprehension as affected by structure of problem representation. *Memory and Cognition*, *4*(3), 249-255.

McGuinness, C. (1986). Problem Representation: The Effects of Spatial Arrays. *Memory and Cognition, 14*(3), 270-280.

Novak, J. D. (1990). Concept Maps and Vee Diagrams: Two Metacognitive Tools to Facilitate Meaningful Learning. *Instructional Science, 19*(1), 29-52.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments [Special issue]. *Educational Psychologist*, *38*(1), 1-4.

Pape, S. J., & Tchoshanov, M. A. (2001). The Role of Representation(s) in Developing Mathematical Understanding. *Theory Into Practice, 40*(2), 118-127.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Reeve, J. (2005). *Understanding Motivation and Emotion* (4th ed.). New Jersey: John Wiley & Sons.

Salomon, G. (1988). AI in reverse: Computer tools that turn cognitive. *J. Educational Computing Research, 4*(2), 123-139.

Smith, P. L., & Ragan, T. J. (1999). *Instructional Design* (2 ed.). Upper Saddle River, New Jersey: Merrill.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

van Merriënboer, J. J. G., & Ayres, P. (2005). Research on Cognitive Load Theory and Its Design Implications for E-Learning [Special issue]. *ETR&D, 53*(3).

This is a manuscript from an article from *Educational Technology Research and Development,* 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist*, *38*(1), 5-13.

Zhang, J. (1997). The Nature of Representations in Problem Solving. *Cognitive Science*, *21*(2), 179-217.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.

Figure Captions

Figure 1. Interaction for viewing and recording information from the history.

Figure 2. Fragment of a selected diagnostic path of a high scoring student.

Figure 3. Fragment of a selected diagnostic path of a low scoring student.

Figure 4. Excerpt from the final exam used in Experiment 1 at UC, Davis.

Figure 5. Excerpt from the final exam used in Experiment 2 at UW, Madison.

This is a manuscript from an article from *Educational Technology Research and Development*, 2007, 55; 499-520. Doi: http://dx.doi.org/10.1007/s11423-006-9003-8. Posted with permission.