Original Investigation JAMIA

Research Paper ■

An Ethnographic, Controlled Study of the Use of a Computer-based Histology Atlas during a Laboratory Course

HAROLD P. LEHMANN, MD, PHD, JOAN A. FREEDMAN, MS, JOHN MASSAD, PHD, RENEE Z. DINTZIS, PHD

Abstract **Objective:** To evaluate the use and effect of a computer-based histology atlas during required laboratory sessions in a medical school histology course.

Design: Ethnographic observation of students' interactions in a factorial, controlled setting.

Measurements: Ethnographer's observations; student and instructor self-report survey after each laboratory session with items rated from 1 (least) to 7 (best); microscope practicum scores at the end of the course.

Results: Between groups assigned the atlas and those not, the ethnographer found qualitative differences in the semantic categories used by students in communicating with each other and with the faculty. Differences were also found in the quality of the interactions and in the learning styles used with and without the computer present in the laboratory. The most interactive learning style was achieved when a pair of students shared a computer and a microscope. Practicum grades did not change with respect to historical controls. Students assigned the atlas, compared with those not assigned, reported higher overall satisfaction (a difference in score of 0.1, $P = 0.003$) and perceived their fellow students to be more helpful (a difference of 0.11, $P = 0.035$). They rated the usefulness of the microscope lower (a difference of 0.23, $P < 0.001$).

Conclusion: A computer-based histology atlas induces qualitative changes in the histology laboratory environment. Most students and faculty reacted positively. The authors did not measure the impact on learning, but they found that there are aspects of using the atlas that instructors must manipulate to make learning optimal. Ethnographic techniques can be helpful in delineating the context and defining what the interventions might be.

■ **JAMIA.** 1999;6:38–52.

Affiliation of the authors: The Johns Hopkins University, Baltimore, Maryland.

This work was supported in part by a grant from the Robert Wood Johnson Foundation and by an educational technology mini-grant from The Johns Hopkins University Provost through the Subcommittee on Electronic and Distance Education.

Some results reported in this paper were presented at the 1997 Annual Meeting of the Association of American Medical Colleges, November 1997, in Washington D.C.

Correspondence and reprints: Harold P. Lehmann, MD, PhD, 407 Blalock, Johns Hopkins, 600 North Wolfe Street, Baltimore, MD 21287-4461. e-mail: ⟨lehmann@welch.jhu.edu⟩.

Received for publication: 1/20/98; accepted for publication 9/18/98.

Educators have hoped for many years that computers would improve medical education.[1] The report on General Professional Education of the Physician (GPEP)[2] recommended their use, and the ACME-TRI report[3] assessed the implementation of the GPEP recommendations. In both reports, specific mention was made of the promise of multimedia programs to help in teaching image-intensive subjects like anatomy and histology.

Yet some observers have been circumspect about these promises. In an important review, Keane et al.[4] brought evaluators to task for asking the wrong questions in evaluating the use of computers in the medical curriculum. Specifically, he called for evaluators to examine the use of computer-based resources in the curriculum as a whole, rather than focusing on how computers could replace lectures, or measuring whether learning was the same from computers as from lectures. Friedman[5] made a similar plea, calling for studies that go beyond comparing whether two media are equally successful at supporting learning.

In following the GPEP recommendations, we developed Histology Imagebase,[6] a computer-based atlas of histologic images for use in our histology basic science course. To rise to the challenge suggested by Keane et al. and Friedman, we designed a study to evaluate the effects of integrating the atlas into our laboratory course. We performed a controlled, factorial design study. We hypothesized that using computers as part of a microscope laboratory would have the following effects:

- Increase satisfaction with the laboratory
- Increase student-to-student interactivity
- Increase attendance
- Increase instructor productivity
- Increase microscope skills

We anticipated that these factors would be important to educators who were deciding whether the educational gain would be worth the financial and technical investment in computers. The impact on learning was not a focus of the study.

Our study is unique in the use of a cultural anthropologist to provide the major data for the study. Our methodologic hypothesis was that an anthropologist's structured observation is appropriate for the type of evaluations recommended by Keane et al. and Friedman.

Methods

The Course

The teaching of physiology and histology has been combined into one course, entitled organ systems.[7] The two subjects are closely coordinated, so when students are learning the physiology of any particular organ system, they are also studying the histology of the system. Histology lectures address not only arrangement of cells and tissue but also such matters as how cell and tissue components carry out specific functions, how cells communicate and interact, and how the body regulates and coordinates its many functions.

Each of 12 histology lectures is followed by a two-hour laboratory session in which the cells and tissues discussed and viewed in the lecture are studied at the microscope. The laboratory reinforces and reviews the lecture material. Laboratories primarily involve the examination of slides through light microscopy, although there is discussion of new techniques of cell and molecular biology. Electron micrographs are also studied.

There are four separate laboratory rooms, each accommodating approximately 30 students. In each room, students are assigned individual desks and microscopes. Two students in adjacent desks share a box of 100 microscope slides. The pairs are not required to work together, but they usually end up doing so.

Each student is issued a paper laboratory manual describing the staining and prepartion of the tissue as well as its physiologic functions. No specific laboratory atlas is required, but most students have available at their desks one of several recommended atlases and histology texts.

Each laboratory room is equipped with one two-headed microscope. This is used by different pairs of students during the laboratory period. Occasionally, a laboratory instructor will mount a slide with a particularly interesting view on the two-headed scope for all the students to inspect during the period.

At the beginning of the laboratory period, a small number of photomicrographs of specific areas of microscope slides are projected, to orient students to key areas of interest. At the end of the period, a longer review is conducted using selected slide views of photomicrographs. During this review, students are free to ask questions and may be questioned about the material studied. The projected photomicrographs are the same as those in the computer database. The aim

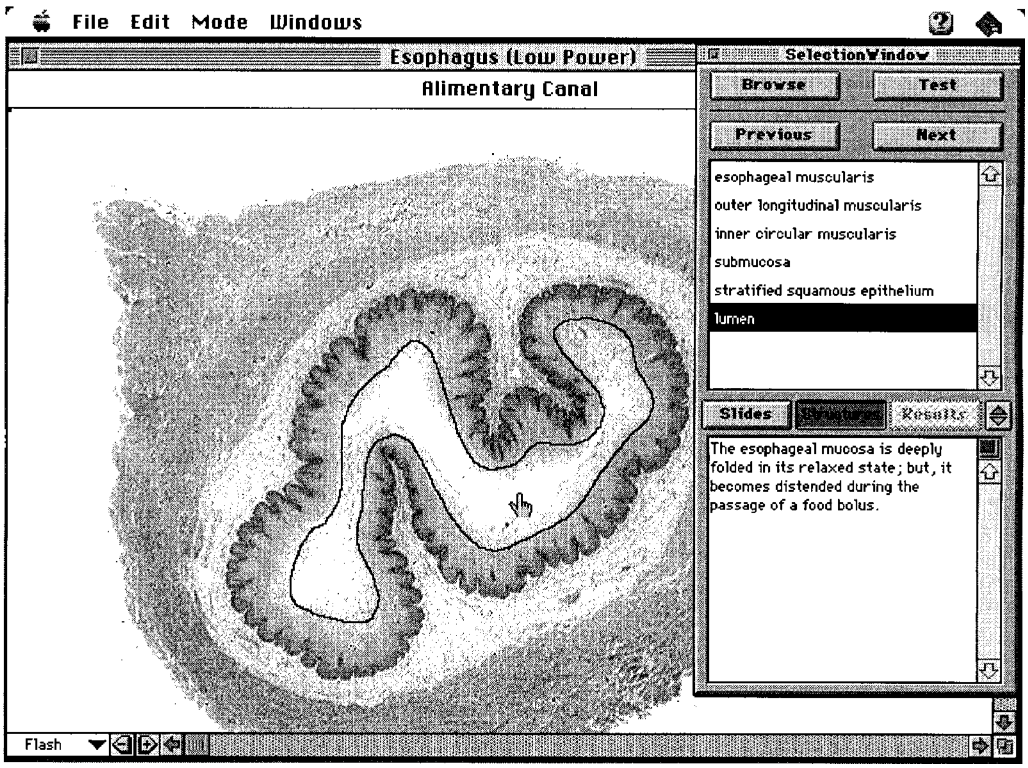

Figure 1 Screen shot from Histology Imagebase. When a user clicks on a region, indicated by the hand-shaped cursor on this image of an esophagus cross-section, the system highlights the name of the region (lumen, at right) and displays explanatory text about it.

is to ensure that every student has found and correctly identified the appropriate structures and understands the function of the organ system being studied.

Each laboratory session is manned by one faculty member and two teaching assistants. Teaching assistants are generally MD or PhD candidates who have already completed courses in histology and pathology. Faculty and teaching assistants do not rotate to different laboratories, so they have continuous contacts with their laboratory group for the 12 sessions of the course. To ensure unformity in teaching and in subject matter covered, faculty and teaching assistants meet before each laboratory session to review slides and to agree on key points of presentation.

The histology course is taken by the 120 medical students and a few graduate students (17 in 1996) from a range of departments (Art as Applied to Medicine, Cell Biology and Anatomy, and Human Genetics).

The Atlas

The atlas, Histology Imagebase,[8] contains 250 images, each at a resolution of 24 bits and a size of 640 × 480 pixels. The images are digitized photomicrographs that were photographed from the glass slides used in the course. The atlas has replaced the slide carousels used by faculty members to preview—and by students to review—the course's teaching slides. To create the atlas, significant structures were outlined on the computer by a medical student using a number of off-the-shelf tools. Content was created, reviewed, and edited by the course director (R.Z.D.). The files have been grouped into directories by organ system, just as the slide carousels had been. Figure 1 presents a screen shot from the atlas.

Images were viewed via the OverLayer[9] application, a Macintosh 680x0/PowerPC native PICT image browser that enables users to view images that have labeled overlays. Each overlay, or layer, contains predefined QuickDraw regions of discontiguous areas, a structure label, and a short paragraph description. An overlay may be located by clicking on it directly or by selecting its name from a scrolling list. A version for the World Wide Web is currently under development.[9]

In addition to presenting images and layers, OverLayer contains a testing mode for identifying or locating a randomly selected layer within a randomly selected slide. All students were given access outside the laboratories to a subset of the histology atlas in conjunction with three laboratories six months before the histology course (as part of the molecules and cells course) and with one laboratory three months before the course (as part of the immunology course).

For use in this study, freestanding computers (Macintosh Quadra 700s with 19-inch monitors) were placed on mobile carts that were moved between laboratory sessions. The carts were placed in areas where they did not obstruct microscope use but were adjacent to the microscope stations. To interact with the computer, students left their microscopes to sit or stand near the computer. Time at the computer was not rationed; the students determined their own use of the machines.

The atlas was also made available in the student computer laboratory, the Information Resource Center, two floors below the histology teaching laboratories.

Faculty members were instructed in the use of the atlas in a one-hour session before the course began. They were also given suggestions (by J.A.F.) on the various ways the computer could be integrated into the laboratory work and into the students' process of active learning. All instructors were supportive and enthusiastic about the addition of the computer atlas to the laboratory environment.

Study Design

The evaluation was designed as a fully factorial trial, controlling for three possible effects: computer, section, and session, the latter as two potential confounders. For each laboratory session, one section was designated the computer section; the other three sections were controls. Each selection had in-laboratory use of the atlas three times during the course. Each section was its own control, when comparing the responses of students in those sections when they were assigned and when they were not assigned the atlas. This comparison checked for the possibility that students in one section might have higher baseline preferences for or against computers. Similarly, each session was a control against the others, to check for the possibility that computer preferences may be related to the material covered or to secular trend, that is, to changes in students' preferences over time.

Students were told during the introductory session and in handouts to use the computers in tandem with the microscope: One student of the pair would stay at the microscope for a series of slides, and the other would be at the computer. Data were collected from March to June 1996.

Although an exemption from our institutional research review board was granted, we informed the students that the goal of this study was to evaluate the use of the computer in the laboratories and that their responses were confidential and not tied to their performance. We also told them that an anthropologist would be circulating and observing their behavior and that they should feel free to talk with him (as many subsequently did). The anthropologist made his presence explicit during observation.

Outcome Measures

An ethnographer (J.M.) circulated among all sections, spending twice as much time in the computer-designated section as in the control sections. He paid attention to the following activities: the physical layout of people and objects; the working arrangements of the students; the language used by students; the activities of the students; and their use of the major learning aids, microscope and computers. The observations were recorded by hand and by audiotape and were used to address the study hypotheses regarding students' interactivity, attendance, and satisfaction and instructor and student productivity.

At the end of each laboratory session and in each section, students and instructors were asked to complete a short survey. The student questionnaire used a seven-point scale (from 1, least, to 7, most) to assess the following: difficulty of the subject matter for the student; usefulness of the microscope; usefulness of the computer; helpfulness of the instructor; availability of the instructor; helpfulness of fellow students; overall satisfaction with the day's laboratory session. The instructors' questionnaire used a seven-point scale (from 1, least, to 7, most) to assess the following: difficulty of the subject matter for students; students' grasp of material by the end of the laboratory session; effectiveness of the computer in decreasing need for the instructor's help at the microscope; effectiveness of the computer in improving the quality of questions asked; amount of time spent responding to questions asked by students; amount of time students spent in the laboratory; and instructor's overall satisfaction with the day's laboratory session. (Copies of the questionnaires are available from the authors.)

The completed questionnaires were collected by staff members who were not part of the histology faculty. The results of these questionnaires were used for the study hypotheses related to student and instructor satisfaction and productivity.

Test results were used to address the issue of microscope skills. Four examinations are given in the course. Three of these are coordinated with the physiology examinations and are administered in the lecture hall. The histology part of these three consist of projected photomicrographs with specific structure, cells, or tissues, or a combination of these, to be identified, with occasional questions relating to function. The photomicrographs used in examinations are never those used in reviews or in the computer atlas.

The photomicrographs differ from studied slides not only in material but often in stains used. Thus, students are expected to generalize from the set they have learned.

The fourth examination is conducted in the laboratory at the microscope. During this examination 35 unknown slides are circulated around the room sequentially. Each student is allowed two minutes to identify the material on the slide. Because of the nature of this examination, no specific structures are highlighted to be identified. Again, the slides used in this examination have never been seen before by the students, either in the laboratory, on a previous examination, or in the atlas.

Analysis

Ethnographic data were analyzed in an iterative manner (see Fig 8.1 of Friedman and Wyatt[10]). After initial observations of students' and instructors' language, semantic categories were devised for use during the second phase, where a qualitative scale of social interaction was also derived. These two subjective scales were used to analyze patterns of learning aid use (atlas vs. microscope). In all cases, narratives were saved.

To look for bias in the return of questionnaires, differences in response rates were assessed by a chi-squared test. Questionnaire data were analyzed as continuous (parametric) data after verifying normal distributions. For statistical inference from bivariate models, t-tests were used. To take account of potential confounding, an analysis of variance (ANOVA) was performed, with the factors computer, section, and session and the interaction computer $\times$ section. Because grade scores were not significantly different across years, a power calculation was performed. The software package JMP for the Macintosh (SAS Institute, Inc.) was used for all analyses.

Results

We present the qualitative results—a summary of the ethnographer's observations—followed by the quantitative results from the students and instructors questionnaires, and then the results of the microscope practicum.

Ethnographer's Report

At least 15 minutes were spent observing each class that was not assigned a computer, and 30 to 45 minutes were spent observing the computer-assigned class. For administrative reasons, observations began in session 4.

To give a sense of the data collected, the following is a paragraph from the ethnographer's tape-recorded narrative:

> Session begins with approximately 15-minute slide presentation on the topical area of the lab session. This represents a highlighting of the focus of the lab visually and a verbal link to previous lab sessions. As each slide is presented, verbal cues are given to demonstrate that it "highlights characteristics" of each focus structure, hence it is a representative image, not necessarily matching the slide structural identification exercises that the students will undertake themselves. (04/23/96)

The process of observing the laboratories was iterative in that abstractions of students' and instructors' speech and actions were made in the course of the observations and the validity of those abstractions was assessed in further observations. We report our results in terms of the abstractions made. There were three sets of abstractions. The first was *semantic categories*, that is, the quality of the language used by students during their learning in the course of the laboratory sessions. The second set was *interaction*, which includes the social and spatial components of students' behaviors. The third set was *learning styles*, which deals primarily with the level of passivity or activity displayed by students in learning. The process of learning overall we call the *heuristic process*, suggesting that, although we cannot describe its internal process, we can describe how the students appeared to learn.

The semantic categories we identified, and refined over the course of observation, are Identification, Appearance, Structure, Proximity, and Function; definitions and examples are presented in Table 1. These categories ascend in order of complexity in the knowledge-building process over the course of the laboratory sessions. In terms of the heuristic process, they are cumulative. The Identification category is communicated initially in the presession slide presentation as a means of introduction and orientation. In the review slide presentation, it is often omitted until each slide is reviewed as a means of self-test for the student. Generally, during laboratory work itself, the process forms a feedback loop as the knowledge base builds: The slide is initially identified by name (Identification), then a probing process unfolds where the student moves from Appearance observations to Structure identification and through Proximity features, which often involves a discussion of Function attributes, concluding with the ability to identify the slide (using Identification) from these features without prompt.

Table 1 ■

Semantic Categories

Semantic Category	Definition	Example
Identification (ID)	Refers to the name of the overall structure	"This is an example of esophagus," where *esophagus* is the name of the slide.
Appearance (AP)	Refers to visible properties in a slide	"Do you see this smooth region?"
Structure (ST)	Refers to physiologic structures in a field	"This is ID.* Here is the ST, which is AP due to staining, so you can find it elsewhere."
Proximity (PR)	Refers to relative location of structures	"See these X's surrounding the Y? That tells us this is an example of ID."
Function (RN)	Refers to a wide variety of forms, to a summation of part or all the above categories	"Even if you were mistaken and thought that this was the small intestine, how would you know? Is every cell in the small intestine mucous secreting? No, so you wouldn't see this constant staining like you do when every cell is mucous secreting, so they are functioning as . . ..

*In later sessions, the ethnographer recorded only the semantic categories and not the actual words used.

When students are preparing for the slide examination, however, the semantic category of Identification is of primary concern, because that is what is tested. Even in this case, the use of other semantic categories is essential to the identification process. Indeed, as knowledge builds over time, the students' overt communicative behavior shifts toward the more cumulative semantic categories, but the semantic feedback loop is still fundamental, in the background, to the cognitive processes.

In Session 5, the ethnographer performed a semiquantitative comparison. The results showed that the computer group coded 16 of 24 linguistic cues as Structure-related, while the noncomputer group coded 16 of 28 linguistic cues as Appearance-related and 9 of 28 as Proximity-related. These frequencies suggest that the different learning environments may evoke different type of language and communication and may also function differently in the way the heuristic process of identification takes place.

With the microscope, students are not physically able to rely on gesturing to indicate position in the visual field, and, while the slide orientation or a paper-based atlas may suggest Structure features to seek, there is no assurance that a particular structure will appear on a particular field. Hence, students must rely heavily on Appearance linguistic cues to communicate placement in the visual field. In contrast, when using the OverLayer-based atlas, students can gesture toward structures on the screen. The language in the atlas is based on Structure names, and that becomes the language used by students and instructors.

Later observation suggested that students rely less on Appearance language as they accrue knowledge of common structural features, and shift toward the language of Function. Use of the microscope requires continued reliance on Appearance linguistic cues, whereas OverLayer, by displaying a shared visual image, permits students to use more "advanced" semantic categories.

Our second set of abstractions concern interactions. We observed students who worked alone ("singletons"), in pairs, or in larger groups; groups of three or more students generally behaved similarly. The presence of the computer was noticed immediately by the students and greeted with exclamations indicating the popularity of working in this learning environment. Far from merely creating a celebratory atmosphere, the presence of the computer equipment signals that a quite different set of rules of behavior is appropriate in that laboratory session. Because most students anticipated working at those computers, in lieu of the microscopes, very few proceeded to set up the microscope, slide case, and texts. The presence of the computers in the laboratories tended to de-emphasize the pairing usually associated with the microscope.

An unanticipated learning behavior was the students' use of the histology imagebase on computers located in the Information Resource Center as an integral part of the laboratory experience. By the middle of the

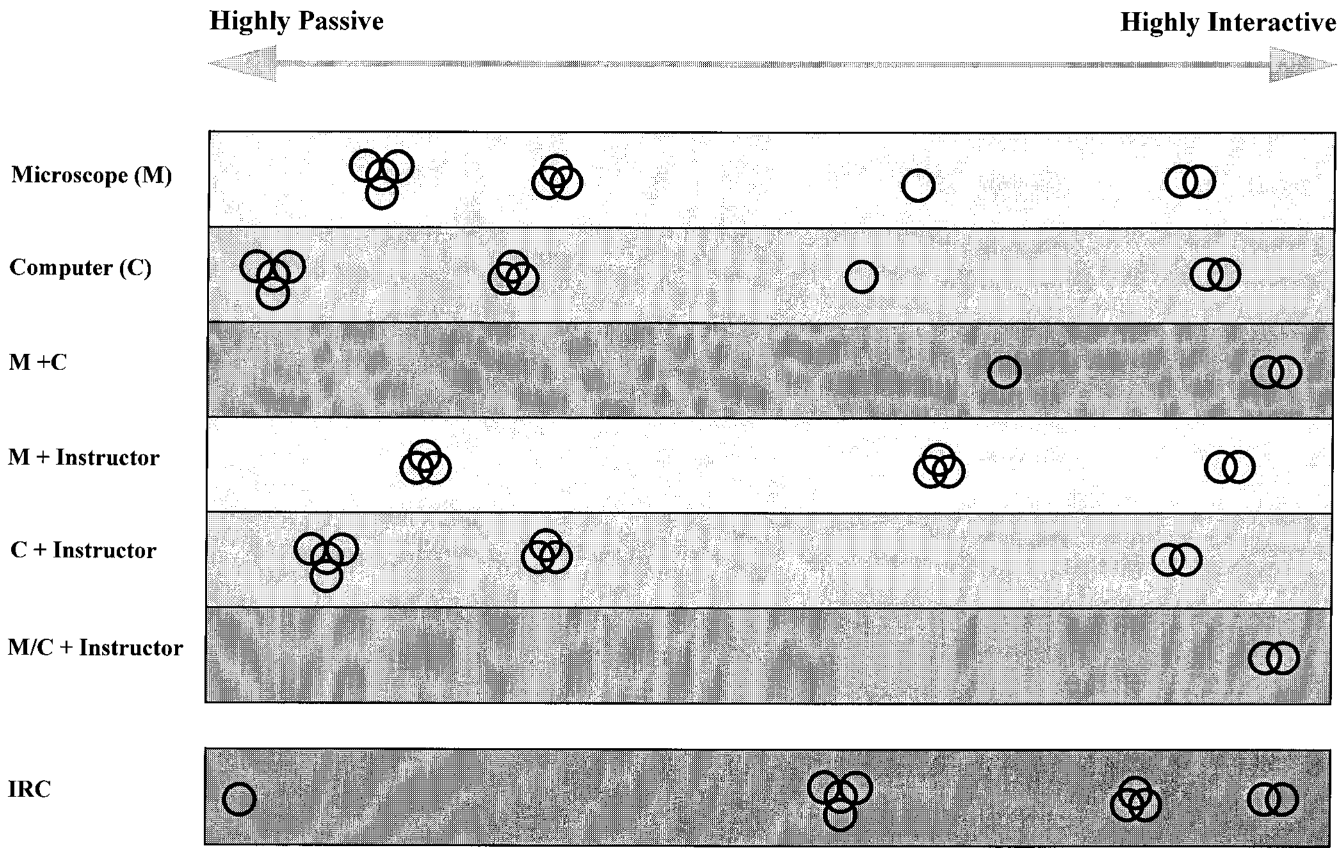

Figure 2 Qualitative scale of observed passivity and interactivity in different settings. A circle indicates a student working alone. Two overlapping circles indicate a pair of students; three, a trio; four, a group.

course, significant numbers of students who were not assigned the computer were going down the two flights to the center to use Histology Imagebase to prepare for the laboratory and were even shuttling back and forth during the laboratory session itself. While this behavior may have contaminated our non-computer group, it demonstrates that students appeared to find the Imagebase atlas useful.

Our third set of abstractions concerns *learning style*. The anthropologist used a subjective scale, with "highly passive" as one anchor and "highly interactive" as the other, to rate the interaction among students and with the instructor. By interactive we mean that the students working together or with an instructor were involved in discussing the visual field and were probing and challenging each other toward understanding. As students were observed working in the laboratory session, their behavior was coded using a scale ranging from passive to interactive and an indication of the social mode—singleton, pair, trio, and group. Figure 2 graphically represents our qualitative findings regarding interaction.

The most interactivity was observed in pairs. The interaction lessened on the addition of more students or an instructor. This increase in passive behavior is the result of other social processes and psychologic differences becoming more influential. Interestingly, students working alone were not necessarily more passive, because they still interacted with texts—laboratory notes, manuals, and a histology atlas—that require students to probe as they would via verbal communication.

As previously suggested, the computer promoted less interaction than the microscope because of both social and heuristic processes. Students working with the computer are less interactive than the same number of students working with the microscope. Pairs are the most interactive in both learning environments, although they are more so on the microscope than on computer. Students working alone on the computer may be similar to those at the microscope in their use of text; however, students working alone on the computer tended to be highly passive, observing the visual field and casually reviewing the textual material. Without social interaction, this learning environment became highly passive. In groups of four or more working at the computer, interactions become more like group discussions, other social processes (e.g., gender-related and ethnic) directly affected the heuristic process, and most of the students were highly passive. Similar patterns relating the social mode to learning environment on this scale were observed among students working in the Information Resource Center.

In later sessions, as students were preparing for the review and slide examination, fewer of them were using the computers when available in the laboratory, choosing instead to use the microscope. As a result, the computers remained mostly idle, although they were available as needed. They were used to compare visual images, to receive prompts from the atlas on the computer and answer basic questions, and to review through the self-test mode. Some students even covered the text on the screen with paper to see more histologic structure without receiving prompts as to the identity of the slide or those structures in the visual field. Students were the most interactive when working either as a pair or singleton.

When students ask instructors questions, the same types of dynamics were evident in terms of the passive-to-interactive scale. When students worked at the microscope, they were most likely to call questions individually. Even though they were working together, the student partner either continued to work or observed the interaction passively. Pairs were more apparently actively involved than singletons. When a pair became a trio (adding the helping instructor) the interactivity diminished. Students working at the computer are most often in pairs or groups and questions are asked as a group, usually by a group leader with the others observing passively or interjecting occasionally. At its extreme, interactivity turns into a passive, didactic group discussion or, with the instructor in larger groups, an informal review session.

In the Information Resource Center, students were observed working in pairs and using the same strategies in this formation as in the laboratory. Pairs were more prevalent in the laboratory than in the resource center because, in the center, students focused on review or on group discussion. Working alone, one student reported that he used the computers in the resource center for review before working at the microscope in the laboratory. Another reported that she used the computers in the resource center for review after working at the microscope in the laboratory. Both reported that they preferred the resource center, because the computers were more readily available and because the atmosphere was less socially active than in the laboratory. These two strategies support the statement by several students that they would like to have ready access to the computer-based information for a variety of review purposes and in a variety of locations, preferably through the World Wide Web or an on-campus network.

Students who worked in groups in the Information Resource Center specifically reported that they preferred working in there because it had a more casual atmosphere for them. They were observed to be highly interactive around the computer terminal, using a variety of linguistic cues, with no discernible pattern to their use. Later in the term, these same students were observed returning to work alone or in pairs for more systematic review of slides on the computer. Several expressed concern that they had not utilized the microscopes enough and may have placed themselves at a disadvantage for the final examination, which would require them to work in that environment. Anecdotally, these students also reported that they used the computer platform for examination review, covering identification and functional information that was provided on the screen to improve their ability to review.

As in the cases of students working alone in the Information Resource Center, the observer reported that the students used the computers both for pre-review and post-review before and after their work at the microscope. In pre-review, the students reported that they used the computer to familiarize themselves with terminology and with slide appearance before working at the microscope. They found this pattern of work helpful in overcoming their difficulties with the visual field in the microscope. In post-review, the students reported that the microscope provided another set of exposures to the visual data, and they found the self-test aspect of the computer platform particularly useful for general review. The self-test function is not available on the microscope directly, and they found this type of review a uniquely helpful aspect of the computer. Some students reported that they used the computer for both pre- and post-review.

Finally, the prospective nature of this study made it possible to follow some individual students over the course of the 12 sessions. Students who emphasized microscope use in the laboratory site in the early sessions were observed to increase their computer use in later sessions. They continued to use the microscopes throughout the term but gradually incorporated the computer into their work. Students who emphasized computer use solely in early sessions were observed emphasizing microscope use in later sessions at the expense of the computer. The latter group had difficulty working at the microscope even in later laboratory sessions, since they had minimized their experience with it as a ''tool'' in earlier laboratory

Table 2 ■

Comparison of Students' Responses Polled Across Sessions

Survey Item	Mean Response (SE)		
	Group Assigned Atlas, $n = 248$	Group Not Assigned Atlas, $n = 693$	P Value
Subject difficulty	4.10 (0.08)	4.15 (0.05)	NS
Microscope usefulness	4.44 (0.11)	4.87 (0.07)	0.0010
Instructor helpfulness	5.36 (0.08)	5.41 (0.04)	NS
Instructor availability	5.47 (0.08)	5.37 (0.05)	NS
Fellow-student helpfulness	5.14 (0.09)	4.91 (0.05)	0.027
Overall satisfaction	4.94 (0.06)	4.72 (0.04)	0.003

NOTE: Students' responses were coded on a scale from 1 (least) to 7 (most). P values were derived from t-tests. NS indicates that the value was not statistically significant.

sessions. Technically, they had difficulty manipulating the visual field in the microscope, but their interpretations of the visual data suffered as well. They appeared less advanced in the use of semantic categories in communicative acts even in later sessions. Furthermore, by limiting their computer use at the laboratory sites in later sessions, they had fewer opportunities to discuss functional attributes through social interaction there with the instructor and other students.

Those students who appeared to work solely at the microscope when the computer was available reported that they used the computers at remote sites either before or after laboratory sessions. In addition, many students reported that, during later sessions when the computers were not available at their particular laboratory site, they would go to the Information Resource Center after the pre-session slide presentation to review further on the computers before returning to work more carefully on the microscopes at the laboratory site. Using the computers in the resource center prepared them to work more easily at the microscope, after having been prompted by the interactive program as to "what to look for." As one student put it, "It helps me find things fast, then I come work on the microscope. If I start on the microscope, I can't find anything" (05/31/96). For another student, the use of the computers in the resource center was even more imperative:

> "[On the microscope there's] too much data in the field. . . . [I] might teach myself something wrong. If I could have a [teaching assistant] with me at all times, but even then, it's imprecise because we have to communicate what's in the field on the microscope, red thing you know, and if you bump it, even with the pointer, you can't be sure." (05/31/96)

Although it was made during a relatively late session, this comment expresses well the difficulty some students have using the microscope as a learning aid throughout the course. It also supports our semantic categories for language, interaction, and learning style abstractions.

Student and Faculty Surveys

Response rates from the students ranged from 100 percent at the beginning to about 50 percent at one point in the middle of the course, for an average of 70 percent. Of 941 responses, 248 were from students assigned to groups with Histology Imagebase and 693 were from groups without the computers; the difference in response rates is not statistically significant. Instructor response ranged from 58 to 92 percent, also with a mean of 70 percent. Of 77 total responses, 18 were from faculty in sessions assigned the atlas, and 59 from those in sessions not assigned the atlas. Again, the difference in rates is not statistically significant.

We examined the responses of students in sessions assigned the computer and those assigned to sessions without Histology Imagebase.

Table 2 presents descriptive summaries. The students assigned the atlas found fellow students to be more helpful (5.14 compared with 4.91, a difference of 0.23, $P = 0.027$) and had higher overall satisfaction (4.94 compared with 4.72, a difference of 0.22, $P = 0.003$). On the other hand, they found the microscope to be less useful (4.44 compared with 4.86, a difference of 0.42, $P = 0.014$).

In the ANOVA multivariate models that took all factors into account, the effects are expressed as differences rather than absolute values. Students assigned the atlas reported overall satisfaction 0.11 units higher than those not assigned the atlas ($P = 0.003$)* and reported their fellow students to be more helpful (a difference of 0.11, $P = 0.035$). The students in sections assigned the atlas rated the usefulness of the microscope lower by 0.23 units ($P < 0.001$), independent of section or session. Notice that these differences are smaller than corresponding differences mentioned in

*For the sake of brevity, the full details of these models are not listed in a table but are available from the authors on request.

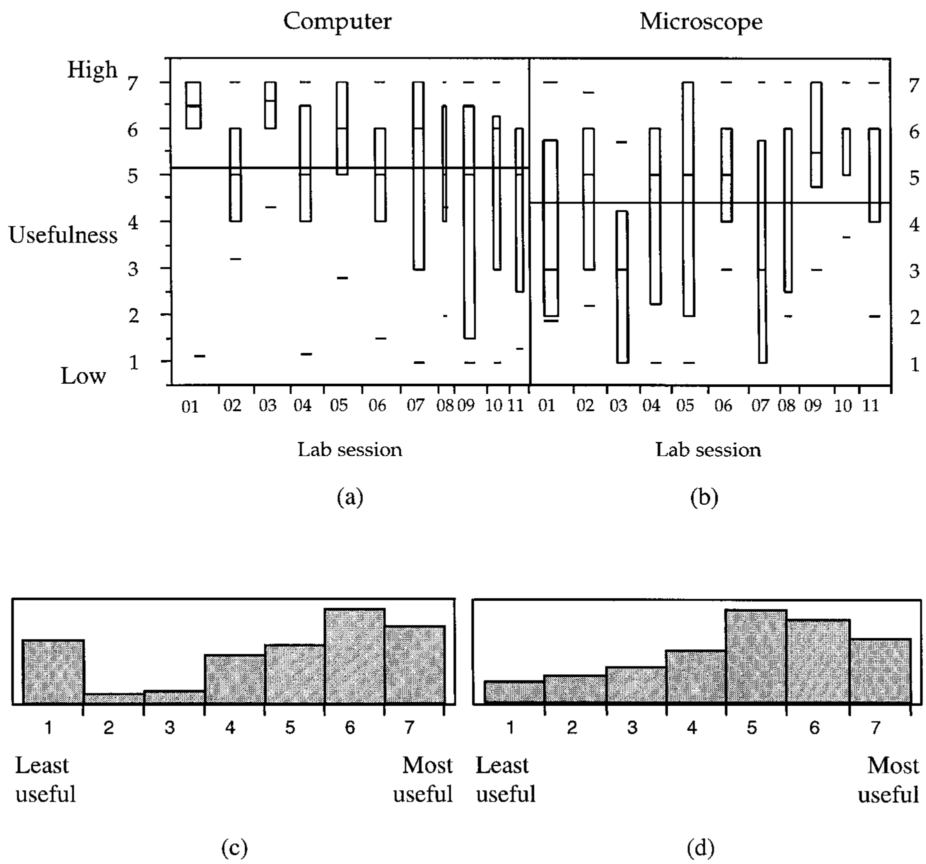

Figure 3 Comparison of responses of students in the computer-assigned sections regarding the usefulness of learning aids. *Top*, Box plots showing responses regarding the usefulness of (*a*) the computer, and (*b*) the microscope. The bottom line of each box plot indicates the 10th percentile; the box, the 25th to 75th percentile; the top line, the 90th percentile; and the horizontal line in the box, the median. Box plot widths are proportional to sample sizes. The horizontal line across each graph indicates the grand mean. *Bottom*, Histograms showing responses across all sessions regarding the usefulness of (*c*) the computer, and (*d*) the microscope.

the previous paragraph, because the ANOVA model accounts for the confounding due to section and session.

Confounding due to section was found in the responses of students in two laboratory sections that were consistently, and statistically significantly, different from those of the other two sections, one 0.3 units higher, the other 0.3 units lower.

Further confounding was found due to session, with statistically significant differences based on session alone that were of even greater size than those based on atlas assignment. For instance, the maximum difference in perceived subject-matter difficulty was 1.0 unit, independent of other factors. This effect was probably due to secular trend rather than to subject matter, because there was no correlation of responses and students' assessments of the subject matter difficulty and because instructors' assessment of subject matter difficulty decreased over time (see below).

The microscope was increasingly perceived as more useful over the course of the sessions (increasing by 1.2 units; $P < 0.001$).

To give a sense of how perceptions of the atlas and of the microscope fared during the course, Figure 3, *top*, compares the self-reports of students assigned the computer (across sections and sessions). There is no statistically significant trend of computer-usefulness scores across sessions ($P = 0.052$). However, it appears from the graph that students perceived the atlas as less useful after Session 8; this appearance is confirmed by a *t*-test of earlier sessions (before Session 9) vs. later sessions (average drops from 5.3 to 4.5, $P = 0.005$). Microscope usefulness also has not statistically significant overall trend, but the early perception of usefulness, at 4.1, rises to 5.4 in the later session ($P < 0.0001$). This improvement in regard for microscopes is mirrored in the self-reports of students in groups not assigned a computer (not shown). Either students appreciated the microscope more, or their realization that the final practicum would be based on the microscope changed their opinion, or there was a novelty effect. In other words, after the initial flush, students found the computer to be as helpful as the microscope.

Finally, the bimodal distribution of the computer ratings in the raw histograms in Figure 3, *bottom*, suggest that there was a small group of students who just did not find the computer useful.

Instructors perceived the student's difficulty with the material to decrease during the course, from a mean of 3.73 to 4.8 ($P < 0.0011$). Faculty perception of students' grasp of the material and the instructor's own satisfaction with the laboratory did not change during the course. There was no statistically significant difference in faculty perception of the amount of time students spent in the laboratory between groups assigned the atlas and those not assigned.

Finally, instructors who had been assigned the atlas rated two questions not asked of the other groups. Instructors rated the effectiveness of the computer in decreasing the need for their help at the microscope at a median of 5 (interquartile range, 4 to 6), where 1 was "increased need," and 7 was "decreased need." They rated the effectiveness of the computer in improving the quality of questions also at a median of 5 (interquartile range, 4 to 6), where 1 was "no change," and 7 was "improved questions." The ratings suggest a modest positive effect of Histology Imagebase in improving the learning environment.

Examination Results

Table 3 shows the average examination scores on the microscope practicum over four years. Although the trend across all four years shows a statistically significant difference, the last three years have, statistically speaking, the same score. The study had the power of about 50 percent to detect the observed difference in the last three years, suggesting that there might still be an effect that we could not detect.

Discussion

We have gained a rounded picture of the impact of a computer-based atlas on student behavior and learning during histology laboratories through a combination of structured observation and self-reporting in a factorial, controlled study. What emerges is that, after the initial enthusiasm and interest in the computer program has worn off, many students develop their own optimal balance of traditional and novel learning methods. This balance involves working with the microscope, fellow students, and instructors and even using external resources, like the Information Resource Center, in addition to the laboratory computer.

We can address our initial hypotheses on the effects of the computer:

- *Increase satisfaction with the laboratory*: We found that availability of Histology Imagebase raised satisfaction self-report by students about 0.1 units of a 7-point scale, even taking sections and sessions into account. Qualitatively, students expressed pleasure at having the machines available. The computer was reported as more useful than the microscope (about 0.2 units), although this perception varied a great deal over time, and there was a core of students opposed to the machines.
- *Increase student-to-student interactivity*: Students assigned the atlas rated their colleagues as more helpful, by about 0.1 units. Qualitatively, we documented a complex relationship among the language necessary for communication, group size, interactivity, and learning environment. The most interactive configuration was a pair of students using both the atlas and a microscope. The microscope tended to force students to use language related to the appearance of structures on the slide, whereas the atlas encouraged discussion about identification labels and function.
- *Increase attendance*: The instructors' self-reports suggest that students assigned the computers spent less time in the laboratory. However, availability of the Histology Imagebase in the Information Resource Center made the resource center part of the laboratory experience, forcing an enlarged definition of the term "attendance."
- *Increase instructor productivity*: Instructors' self-reports suggested that the atlas led to a mild decrease in the amount of time instructors spent helping at the microscope and a mild increase in the quality of questions students asked. Both these effects were desired.
- *Increase microscope skills*: There were no statistically significant differences in student examination scores on the practicum (although the study had only a 50 percent power to detect the observed difference). Given the direction of the scores, we can conclude that the electronic atlas probably did not worsen students' microscope scores.

The effects we observed were in the direction we expected. As the ethnographer pointed out, many students wanted the study to show a positive effect, and our experience outside this study has been that the machines hosting the atlas get a tremendous amount of use between laboratory sessions. This use pattern recurred the year following this study, yet none of the effects we documented were large. How are we to reconcile the small effects seen with our prior beliefs that the effect would be large?

First, students differ in their preferred styles of learning. There was a minority of about 10 percent of students who consistently rated the computer negatively. On the other hand, there were students who preferred the atlas, running downstairs to use it in the Information Resource Center before and after the labora-

Table 3 ■

Comparison of Average Scores on Microscope Practicum over Four Years

Year	Computer Resource	Total Possible Score	Average Score* (SD)	Percentage of Maximum Score	
				Average	95% Confidence Interval
1993	None	127	99.3 (9.3)	78	64–100
1994	Database of images and captions outside laboratory	135	111.2 (11)	82	66–100
1995	Histology Imagebase outside laboratory	125	105.7 (7)	85	73–100
1996	Histology Imagebase in laboratory	126	103.5 (9.5)	82	67–100

*Sample size of 120 students in each year.

tory session. Second, the laboratory course has two goals: students should learn to identify structures, and students should learn to use the microscope. Obviously, with regard to the second goal, a computer-based tool will always come out wanting. It may be argued, however, that in the near future, basic-science teaching will focus on teaching students to use computers instead of microscopes, because physicians will in fact be viewing all clinical images through computer-based technologies. In this future, computer-based atlases will probably have a stronger place in medical student's armamentarium during laboratory sessions.

We did not study the atlas's impact on learning. To do so requires a larger study, given the high baseline achievement of the students. Furthermore, the study would require more intense observation of the students, since the gains would be related to cognitive style and style of learning, and therefore affect only a subgroup of students.

The use of ethnography in informatics research has attracted much interest in recent years.[11] To our knowledge, this is the first study to employ ethnographic methods in the evaluation of computer-aided medical learning. We believe that this methodology speaks precisely to the questions raised by recent critics. Keane et al.[4] pointed out that we need to go beyond test scores as the primary outcome measure in evaluating computer-aided learning. Friedman[5] echoed the need to go beyond media-comparison studies. In our study, we attempted to provide a holistic understanding of the impact of an electronic atlas.

It is important that the research methods complemented each other. Without the ethnographic observations, we would have lost the key qualitative results of this study. Yet without the self-reports, we would have lost an important sense of the changes over time and among laboratory sections. For instance, we found that indeed, student responses were confounded by session and section, and we were able to tease out these effects by our design.

Most previous studies have relied on tests of knowledge as the outcome measure. Mars and McLean, in their study of a histology program used during a course,[12] focused on pre- and post-test scores, demonstrating improvement with their electronic atlas. Walsh and Bohn[13] demonstrated positive perception by medical students using a radiology program but no significant changes in test scores. Broader outcome measures were used by Mangione et al.,[14] who demonstrated some qualitative differences but no differences in test scores in an auscultation tutorial used by an individual compared with small-group learning, and Garrett et al.[15] used self-report scores of 23 third-year students to demonstrate perceived usefulness of a blood-smear teaching program.

We call our program a computer-based *atlas* because there is no educational design beyond the sequence of images and the choice of structures to be highlighted. Other "multimedia" teaching programs emphasize their tutorial aspect or their ability to replace lectures. Electronic atlases are increasingly available, both for general use—like those derived from the Visible Human Project[16]—and for specific purposes, like neuroanatomy[17] and histology.[18] They all use the Identification and Structure approach to labeling.

Stephenson[19] reviews prior research on the use of computers in small-group learning and concludes that the optimal group size is two or three students, that achievement is not necessarily improved, and that same-gender groups yield the best learning. Our optimal group size was two. This difference is probably due to the content area. The computer programs in the 35 studies reviewed by Stephenson were all tutorial; none was an adjunct to a laboratory, where other sources of information would be available.

There are a few limitations to our study. As in any trial, issues of contamination—control subjects inadvertently receive the experimental treatment, or vice versa—are of concern. In our case, many students not assigned to the atlas used computers in the resource center. Although this increased use of the atlas affects the inferential validity of the study, it does not affect our other major aim—to examine the impact of having the computer-based atlas in the laboratory itself and to examine its use during a session.

Ideally, we would have had multiple, cross-checking observers in each class, but there is a rich tradition in the anthropology literature of having a single, well-trained observer.[20–22] Ethnographers have, over the last two decades, integrated language-use analysis with consideration of the interactional context of human cognition.[23–25] This technique has been applied to the observation of scientists working in a research laboratory.[26]

Because we used a single observer, we do not claim that the frequencies and numbers reported from those observations are anything more than a first-order approximation. A formally precise quantification would require the recording, verbatim transcription, and coding of multiple series of communicated events on different learning environments throughout the term of the course. Ours was a pilot experiment to determine what types of findings could be systematically shown with more exhaustive techniques and by reproducing the experiment with multiple students across time. This project establishes the basic techniques for a more formal quantitative and qualitative analysis. For instance, our scale of learning styles falls within the set of styles laid out by researchers in problem-based learning.[27] We presumed that interactive learning was "better" than passive on the basis of conclusions of those researchers, but we did not verify this presumption in our study sample. We recommend returning to the subject using Werner's systematic methods[25] for unpacking the depth of semantic categories and relating them to students' learning strategies.

The declining number of students answering the self-report surveys may diminish our certainty in the values of the responses. Although the lowest response rate was 55 percent, however, the sessions with the lowest response rates did not have responses out of line with sessions before or after them, suggesting that the respondents were not self-selected on dimensions that mattered. If anything, it was students who were "pro-computer" who had the greatest incentive to answer the survey, since they would perceive that better responses for computers would influence the administration to purchase them. Yet approval of computers actually decreased. We do not think that the decrease is due to a novelty factor, because the students all had access to these programs in the Information Resource Center between laboratory sessions and even during laboratory sessions. Rather, it was probably due to their realizing they they would be tested on the microscope.

To what degree are our conclusions limited because of the nonmasked nature of this study? One might argue that the students would interpret that the purpose of this study was to justify the purchase of more computers, and so those who supported the use of computers would be motivated to misrepresent the truth. In fact, the ethnographer was told by some students that they were inflating their responses for just that reason. Yet, overall, the ratings varied and went lower than initially, strongly suggesting that the students' numeric responses were honest, despite a minority. The students had the real task of learning histology and of passing their examinations; they did not have the time to alter their behavior for the observer. The strength of having a participant observer is just to detect such cheating.

Our results have interesting implications for teachers. The computers appeared valuable to students and would optimally be available in all laboratory rooms, at each laboratory station, as at Cornell University School of Medicine, for instance (S. Stensaas, personal communication). Active efforts are needed to help students use the software in the optimal social configuration—i.e., with another student. Recall that, although we told students that that configuration would be optimal, many students did not settle into that routine until halfway through the course. Perhaps having computers available continuously and as part of the workbench would speed that process. Yet, it may be appropriate to limit their use in the initial laboratory sessions until students become proficient with microscopes and with their abilities to name things using the Appearance semantic category. A tactic to achieve this limitation might be to have only a single computer available, and at the control of the instructors. Another tactic would be to place a machine in another room, not a laboratory room, with limited access early in the course but with increased access later. Of course, if microscope skills are not the focus of the course, then these recommendations may be obviated.

Our observations regarding semantic categories have implications for the developers of electronic atlases. The Web has opened immense opportunities for instructors to make their material available to a wide variety of users. How should these atlases be constructed? Generally, an image (photomicrograph, say) is presented with a caption.[28] Occasionally, images are presented with labels either printed on the image itself,[29] displayed in a second image linked via an image map, or provided through client feedback.[30] In each case, the labels are based on the Identification and Structure semantic categories.

We recommend that developers provide multiple labeling schemes based on semantic categories. For instance, a second labeling scheme for a slide would be a list of words dealing with Appearance, so a student could click on the label ''eosinophilic'' and all red-stained areas would be outlined. Or the student would click on a patch of cells on a slide and the label ''sheet of cells'' would be highlighted. This labeling scheme would make the atlas more cognitively comparable to the microscope, since the students would be using the language of Appearance in referring to the images. This arrangement should also help students earlier in the course, just when they are learning Appearance descriptors, according to the heuristic loop we have described.

This multischeme approach would make sense in other domain areas, such as endoscopy or radiology, where the atlas must teach both the language of Appearance (''sessile,'' ''radio-opaque'') and the language of Identification (''villous adenoma,'' ''renal calculus'').

Proximity labeling would require labeling relationships among pairs or groups of structures. Our development laboratory has attempted to address issues of the Functional semantic category by creating a quiz-based application based on the same images as those in Histology Imagebase.

In summary, we have performed an ethnographic, controlled study on the use of an electronic atlas in histology laboratories. Our conclusions are that computer-based atlases are helpful to students during laboratory learning and that they complement the use of the microscope. The use of computers during laboratory sessions changes the character of the laboratory, and the proper use of computer-based atlases and aids must be considered in light of the total laboratory experience. Attention to issues such as language, social interactions, spatial arrangements, and learning styles can help students gain the maximum benefit from the computers and from their laboratory as a whole. These issues will become more important as computers become the primary means of viewing clinical images.

The authors thank Marty Wachter for programming Histology Imagebase, Bonnie Cosner for project support and data management, and Bruce Hornick for technical help. They appreciate John Shatzer's suggestions on project design and manuscript content. Luke Ling helped construct Figure 2. The authors also thank the reviewers for helpful comments.

References ■

1. Piemme TE. Computer-assisted learning and evaluation in medicine. JAMA. 1988;260(3):367–72.
2. Muller S (chairman). Physicians for the twenty-first century: report of the project panel on the general professional education of the physician and college preparation for medicine. J M Educ. Nov. 1984;59(part 2).
3. Swanson AG (principal investigator), Anderson MB (project director). Educating medical students. Assessing Change in Medical Education: The Road to Implementation (The ACME-TRI Report). Acad Med. June 1993;68(supp):S1–S48.
4. Keane, DR, Norman GR, Vickers J. The inadequacy of recent research on computer-assisted instruction. Acad Med. 1991; 66(8):444–8.
5. Friedman CP. The research we should be doing. Acad Med. 1994;69(6):455–7.
6. Lehmann HP, Wachter MR. Delivering structured educational images over a network. Proc 19th Annu Symp Comput Appl Med Care. 1995:989.
7. DeAngelis CD (ed). The Johns Hopkins Medical Curriculum for the Twenty-first Century. Baltimore, Md.: The Johns Hopkins University Press, 1998.
8. Lehmann HP. Medical informatics, educational technology, and the new curriculum. In: DeAngelis CD (ed). The Johns Hopkins Medical Curriculum for the Twenty-first Century. Baltimore, MD.: The Johns Hopkins University Press, 1998.
9. Lehmann HP, Nguyen B, Freedman JA. OverLayer. Baltimore, Md.: The Johns Hopkins University Press, 1998.
10. Friedman CP, Wyatt J. Evaluation Methods in Medical Informatics. New York: Springer, 1997.
11. Forsythe DE. Using ethnography to build a working system: rethinking basic design assumptions. Proc 16th Annu Symp Comput Appl Med Care. 1992:505–9.
12. Mars M, McLean M. Students' perceptions of a multimedia computer-aided instruction resource in histology. S Afr Med J. 1996;86(9):1098–102.
13. Walsh R, Bohn R. Computer-assisted instructions: a role in teaching human gross anatomy. Med Educ. 1990;24(6): 499–506.
14. Mangione S, Nieman LZ, Greenspon LW, Margulies H. A comparison of computer-assisted instruction and small-group teaching of cardiac auscultation to medical students. Med Educ. 1991;25:389–95.
15. Garrett TJ, Savage DG, Hendrickson G. Assessment of an interactive microcomputer-videodisc programme for teaching medical students to evaluate the peripheral blood smear. Med Teacher. 1990;12(3/4):349–51.
16. National Library of Medicine. The Visible Human Project. Bethesda, Md.: NLM, 1997.
17. Nowinski W, Fang A, Nguyen B, et al. Multiple brain atlas database and atlas-based neuroimaging system. Comput

Aided Surg. 1997;2(1):42–66.
18. Ogilvie R. An interactive histology image–barcode manual for a videodisc image library. Medinfo. 1995;8(part 2):1698.
19. Stephenson SD. The use of small groups in computer-based training: a review of recent literature. Comput Hum Behav. 1994;10(3):243–59.
20. Fishman JA. Sociolinguistics: A Brief Introduction. Rowley, Mass.: Newbury House, 1970.
21. Hymes D. Foundations of Sociolinguistics: An Ethnographic Approach. Philadelphia, Pa.: University of Pennsylvania Press, 1974.
22. Fishman JA (ed). Readings in the Sociology of Language. The Hague, The Netherlands: Mouton & Company, 1968.
23. Tyler SA (ed). Cognitive Anthropology. Prospect Heights, Ill.: Waveland, 1987.
24. Spradley JP (ed). Culture and Cognition: Rules, Maps, and Plans. Prospect Heights, Ill: Waveland, 1987.
25. Werner O, Schoepfle GM,. Ahern J. Systematic Fieldwork. Newbury Park, Calif.: Sage Publications, 1987.
26. Latour B, Woolgar S. Laboratory Life: The Construction of Scientific Facts. Princeton, N.J.: Princeton University Press, 1986.
27. Norman G, Schmidt H. The psychological basis of problem-based learning: a review of the evidence. Acad Med. 1992; 67(9):557–65.
28. Swanson JR. Anatomy and Histology of Normal Skin. Chicago, Ill.: Loyola University Press, 1996.
29. Johnson KA, Becker JA. The Whole Brain Atlas. Cambridge, Mass.: Harvard University Press, 1997.
30. Hagler HK, Kumar V (eds). University of Texas Southwestern Pathology Case Studies. Stanford, Conn.: Thomson Science, 1997.