CHAPTER
10

Adaptive Computerized Educational Systems: A Case Study

ROGER D. RAY
Rollins College

UNDERGRADUATE TEACHING IN THE MODERN UNIVERSITY

Consider a very typical teaching and learning scenario in higher education today. The instructor for a large-enrollment Introductory Psychology (substitute any other scientific discipline you would like) university course begins the first day of classes by distributing a syllabus with textbook reading assignments spread across the semester. Other assignments may also be included, such as research projects or practical volunteer credits, but foremost are the textbook assignments. After all, the textbook has 300+ pages and cost the student more than any other course resource; thus, the textbook plays the most important role in the course outside of the instructor's classes themselves. But, it is interesting to note which assignments gain the students points for grading. The textbook readings almost never generate points directly, while other forms of activities do. How is an instructor expected to track and give credit to 200 or more individuals per class for completing each chapter assigned? Certainly, no partial credits are given for gradations of understanding the readings. Instead students are duly warned that readings will be covered by in-class tests (but typically not more than two to three such exams plus a final are given within a semester). Tests are few in number because they are likely to take the whole class period, and class time typically is precious to lecturers. So in-class exams must also cover lecture materials to make sure students attend class, listen, and learn from what the instructor has to say. But, if 10 to 15 chapters are covered in the semester, then 5 to 8 chapters are covered by each in-class test composed of approximately 60 to 80 items total. That is about 5 questions per chapter, with 5 questions for lectures on the same chapter materials. This means the density of any specific content's sampling is typically quite small, at least bringing into question the reliability of such sparse sampling. No wonder students complain that, despite their feeling that they have substantial knowledge, few of the right questions are ever asked to prove their mastery of the material!

Add to this scenario the actual lecture activities. Instructors hope, with little realistic expectation, that students have read the material in time for the in-class explanations, supplementations, or expansions of the readings assigned. Thus, for example, today's lecture starts coverage of the physiological foundations of behavior with the instructor in high hopes that students have read all about neurons, nerves, and their composite construction of a nervous system, including both peripheral and central components. Given such high expectations, the lecture starts by focusing on one specific central component of significant practical interest—the hypothalamus as a specialized body of neurons and its role in the fundamentals of food intake, body weight maintenance, and sexual behaviors. Just as the instructor is about to transit from neural to hormonal control, some student sheepishly asks if the instructor could please restate what a "neural" is . . . .

It probably is not presumptuous to think that instructors reading this might have had more than one carefully prepared lecture doomed by the sudden realization that few, if any, in the class actually read the assignment prior to class. Students may have been assigned to read and study the material as foundations for today's planned lecture, but it is perhaps one of the few ubiquitous experiences shared by all instructors of introductory courses to find such assignments largely ignored. Consider just one published example to highlight this point.

Because I personally have over 35 years of classroom experience, the following example from the recent literature did not surprise me in the slightest. Sikorski *et al*. (2002) reported on a two-university survey of student use of introductory texts that found as high as 91%, *but as few as 31%*, of students in introductory psychology classes actually even *purchased* the required text, much less read it or studied it. In fact, the majority of students at both universities surveyed reported " . . . that taking notes and studying them (without reading the text) was the single most important contributor to doing well" (p. 313). So much for lecturing to prepared learners!

UNDERGRADUATE TEACHING IN SMALL LIBERAL ARTS COLLEGES

Most instructors in today's large universities have no direct experience with the existence of educational institutions that pretty much ensure that their students in introductory courses learn through preparatory reading. Instructors with doctorate degrees actually teach individual students, at least as a class supplement and commonly as independent studies, in the rarified atmosphere called the small liberal arts college. Small classes foster early identification of ill-prepared students, and instructors often make individualized help readily available for those having difficulty with class preparations. For example, the last introductory psychology class I personally taught had nine students enrolled. I will admit that this is not typical even for offerings of Introductory Psychology in my school, although an enrollment limit of 25 to 30 is standard. It does, however, stand as an illustrative example of the educational environment being discussed. Thus, the relatively few instructors teaching in the smaller liberal arts colleges across the United States will easily recognize this scenario and find it quite familiar.

Of course students in this environment may still attempt to avoid buying or reading assigned text materials, but class activities tend to focus more on didactic exchanges that make faulty preparations by students more apparent and even personally embarrassing. So many students tend to read assignments, if for no other reason than fear they will be called upon to expound upon this material in class. Or at least they *try* to read assignments. Which brings me to the second element of my small college scenario. This second element is a problem that exists in most of higher education, although it is not as serious at elite institutions that screen applicants with highly selective criteria for admission. The problem? Students are more and more frequently coming to college with poorly developed skills for reading and comprehending textbooks, even if they try. Thus, it is not uncommon to have students who attempt to prepare for class, but who also find that if they are required to discuss or explain the material in class, it is highly difficult for them. When a student has a problem understanding the textbook in my course, either the student seeks me out or I ask that student to seek me out to obtain help. When asked to do so, that student is likely to come to my office for that help. And, because my own efforts to help such students follow many well-known and highly sound behavioral principles, it may be instructive to review these principles.

I assign readings in my introductory class with the goal of establishing a common working vocabulary of the principles, variables, and conceptual foundations for the discipline, and that is where I typically focus my first probes when I work with a problem student. That is, I will ask the student a few questions to get some idea of where to start in remediation (what behavior analysts call "establishing a baseline"). If I determine the student truly has had significant problems getting these fundamentals from reading the chapter, I will ask the student to show me how he or she reads the material and what kinds of note taking and rehearsal activities the student does after reading the material.

If the student demonstrates serious reading skill deficiencies, I start remediation by having the student first read a selected paragraph aloud to me. Then I have the student point out the central concept being introduced or elaborated in that paragraph, sentence by sentence. If the concept cannot be correctly identified, we will spend review time on what a concept is and how to recognize one when it is presented by textual discussion. After the student has successfully identified the primary concept, I will further ask the student to point to the attending or defining properties in the paragraph that elaborate that concept. If the student cannot do this readily, I will have the student reread the paragraph aloud, and together we will attempt to isolate the sentence that contains the most primary or important concept.

We will then consider what in that sentence and subsequent sentences elaborates on that concept. This continues perhaps by my reading one of the sentences but leaving blank one property so the student can practice filling in the missing words as properties of the concept. For example, I might ask a question to verify that the student understands how the concept and properties relate in a form such as: "*Neurons* (main concept) taken as a collective make up a ______." (*Nerve* would be the desired answer.) Then, I might follow with: "All *nerves* [note the shift to a related concept] considered as an integrated whole makeup the entire ______." (*Nervous system* would be the desired answer). We will typically do this until the student can fill in at least three to four such properties.

Once we have moved through a few paragraphs at this level and the student has shown mastery of this lower skill, we typically move on to consider how the multiples of paragraphs we have covered relate to each other. For example, how do the concepts of synapse and neuron contrast or compare to each other? What is the relation of one to the other? Do drugs tend to influence synapses or neurons?

It is perhaps time to reflect on the behavioral principles being followed here. Cognitive psychologists would likely say that I am working to find what Vygotsky (Berk & Winsler, 1995) called the "zone of proximal development." Having found this zone where the student can work successfully only if I help, I then begin to scaffold the student's learning by focusing on concept and property identification. This is an interesting restatement (apparently unknowingly so) of principles articulated a bit differently by B. F. Skinner (1968) that considered the shaping of behavior through use of a series of "successive behavioral approximations" for making the transition from what a student can already do (*baseline*) to what the teacher aspires for the student to be able to do (the *teaching objective* or *target behaviors*). I believe the behavior analytic articulation afforded by Skinner is more complete because it tells us precisely what we should do and when we should do it to move the student progressively through this process. Thus, how one scaffolds is less clear to me in regard to precise variables, behaviors, and timing than is the articulation of successive approximation strategies.

For example, behavioral principles articulate three concurrent efforts that one should make that could be described as successive approximations during the shaping process. These emphasize what is done regarding (1) behavioral *antecedents* in the environment, (2) stages of *behavioral development* themselves (what is sometimes called a *task analysis*), and (3) stages of how behavioral consequences, in the form of *reinforcement densities*, are managed. I will briefly elaborate on each component of this antecedent–behavior–consequence analysis.

First, the behavior analytic model points to the significant contributions of attendant antecedent stimuli that precede the behavior. These include instructions, the way text is broken down into segments, and many other forms of what may be described as *prompts* to help generate, guide, and sustain desired behaviors. Of course, one does not want to have to use antecedent prompts forever, so one gradually (step by step through a series of successive approximations from high-density prompting to no prompting at all) *fades* the use or presence of such antecedents.

The second use of successive approximation is called *response shaping*, and it focuses not on the antecedents to behavior but rather directly on the behavior being taught. Skills are learned not as full-blown activities but in gradual stages of development, as when a baby learns to crawl through a series of various activities that are foundational components of crawling. Walking likewise starts with standing, then shuffling while holding onto something (an antecedent prompt), then gradually taking what looks more and more like steps, with stepping relying on less and less support until full-fledged walking is occurring. This is sometimes referred to as taking "baby steps" to go from standing to full-scale walking. In any case, the effort is to change behavior in successive and sufficiently small stages from what exists prior to shaping to the desired goal behavior that will terminate shaping.

The third successive approximation procedure focuses on reinforcing consequences in the environment and how frequently they are used. The goal is to decrease the density of reinforcements through a process that might be called *leaning*. Think of this as a metaphor for rich versus lean meats or diets. It all has to do with density of some element (*e.g.*, fat) in the meat or diet. Behavior analytic principles stress, beyond almost anything else, the important role of reinforcing consequences in determining whether behavior will increase or decrease in likelihood.

E. L. Thorndike (1898) discovered the importance of behavioral consequences and formally articulated how he felt they worked in his Law of Effect. But B. F. Skinner made an even more important discovery while working with reinforcements (Skinner, 1956; Ferster & Skinner, 1957). He discovered that one could move, in graded (successive) steps, from high-density use of reinforcement (*i.e.*, reinforcing *every* response occurrence) to extremely low-density reinforcement (lean or intermittent schedules where a very small proportion of responses are reinforced) and thereby actually increase the likelihood of the behavior! Of course if one attempts to "lean" reinforcement density too quickly, disuse of the behavior is most likely because of the process of extinction (Skinner, 1956). So another successive approximation from rich (high density) to lean (low density) delivery of behavioral consequences is also desirable. Thus, we have initial prompting and the gradual fading of these environmental antecedents to behavior, a gradual shaping of the form and function of the behavior itself, and continuous reinforcement being gradually leaned even to rare occurrences as highly specific recipes of variable manipulation. I find this more precise a specification than the usually vague suggestions offered in most discussions of scaffolding the learning process.

Teaching content conveyed by text through successive approximation techniques was partly "automated" very early in the development of scientifically inspired approaches to improved teaching (Skinner, 1968). The technique was then (and still is) called *programmed* instruction (Vargas & Vargas, 1992). In this approach, the tutor is removed from the process by breaking down the text into small units (called a *frame*) to be read and mastered before going to the next frame. Each frame of material presented stays relatively constant in size and complexity so even those with the poorest of reading skills can learn without experiencing failure. One problem with traditional programmed instruction is that it uses no form of successive approximation, except in its formation of the learner's facility with the *content* being taught. It does nothing to fade a student's reliance on programmed forms of text materials, to lean the density of reinforcing feedback, nor to shape better reading skills.

With my side review of behavioral principles complete, let me return to my tutorial scenario where we last left the student as having only succeeded in learning how to read for, and to verbalize, the critical concepts and concept properties paragraph by paragraph. It certainly is not my intent as a teacher to have to do supportive tutoring with the same student for every chapter throughout the course. So, I quickly establish a set of step-wise goals (successive approximations) for shaping the student's reading comprehension and study skills beyond this early stage of development, while also fading my own prompting and questioning as a part of the student's study activity. I also want to get much larger units of behavior from the student before I give reinforcing feedback (that is, I want to lean reliance on high-density reinforcement). To accomplish this, I gradually begin working with larger and larger units of the text, begin to fade my use of the author's own language in favor of a more abbreviated and paraphrased use of terms and properties being presented, and begin to probe more for the student's understanding of what goes with what—eventually asking for the student to verbally outline what is covered in major sections of the chapter without assistance.

This multiple successive approximations of prompting then fading, shaping of concept/term and association selection ability, and leaning out the density of my consequential feedback for accuracy and even fluency (speed of responding) is what most effective tutors would likely do. And, the results are such that before long tutorial help is less frequently needed, can be focused on larger and larger units of study, and can be assessed more and more abstractly by asking who, what, where, when, and why questions regarding major concepts. Eventually the student starts reading and taking notes and rehearsing this material independently.

It is very important to note that in the above scenario I end up teaching two different things simultaneously: one is the teaching of content, which is why the student came to me in the first place; the other is the acquisition of skills required to read with accuracy and fluency regarding the student's reflection of the content being read (what most educators would call *reading comprehension skills*). This process of dynamic and multidomain successive approximation is the epitome of what I will refer to hereafter as *adaptive instruction*. Unfortunately, such effective tutorials must be individually and adaptively focused and thereby can be afforded only to students in the most expensive educational institutions or most highly funded programs (such as remedial tutoring programs for university football and basketball players). The alternative to tutoring the less skilled students in low-tuition and high-enrollment environments is failing students out of the institution altogether, but technology may change that.

COMPUTERS AND ADAPTIVE INSTRUCTION

Efforts in programmed text instruction were quickly translated into mechanized forms for automated delivery via Skinner's elaboration of the "teaching machine" (Skinner, 1963; Vargas & Vargas, 1992) in the 1950s and 1960s. But, by the late 1980s, apparent potentials for a convergence with several additional technologies (behavioral/educational, artificial intelligence, and digital communications/computing) prompted the author to begin exploring ways of addressing shortcomings in this traditional approach. Computer programming was seen as one potential means for creating not only automated but intelligent delivery of the *entire* process summarized in the small college tutorial scenario, not just its first stage. Software programs written from the perspective of artificial intelligence and expert knowledge systems allow one to build a more dynamic and adaptive responsiveness to learner actions which automates many of those same stages of successive approximations to expert reading skills described. The phenomenal explosion of Internet connectivity now allows such computer programs to communicate with more centralized sources of content, so that these expert systems and personalized mirrorings of learning histories may be accessed by students from almost any physical location at any time.

But, why are such adaptive programs required? Well, even though I understand the power and empirical successes of both programmed instruction (Vargas & Vargas, 1992) and personalized systems of instruction (Keller, 1968), I have sufficient reservations about each, as typically practiced, to prompt me to attempt to improve upon both. Programmed instruction tends to present material only in micro-frames that can be mastered by even the slowest learner. Pity the poor student who can read large units of text and quickly assimilate the important concepts and properties being articulated but who now has to read only small units at a time before even being presented the next frame. Thus, relatively skilled learners often find themselves constantly frustrated by the unnaturally slow pace of material presentation.

Also, of course, pity the poor student who has learned to rely exclusively on the small frames as prompts for acquiring verbal facility with the material but who now has to read from a textbook that has not been programmed with lots of framing and other prompts! What are learners who have come to rely upon programmed instruction to do if they have to master material that has never been programmed for their pace of learning? I find traditional programmed instruction is designed almost to enable a dependency on poor reading comprehension skills much like families who unwittingly support a loved one enable a drug dependency! Until the advent of more adaptive computer-based programming, the only alternative seemed to be the "tough love" of the more traditional textbook. There was nothing in between to help those who originally needed programmed instruction to gradually wean them from such programming. Programmed instruction eventually should be faded as the primary prompt for successful learning.

Alternatively, personalized systems of instruction appear to me as favoring only the well-skilled reader and as failing to help readers with poor reading or study skills. True personalized instruction incorporates peer tutors to help students practice their poor reading skills over and over because such tutors typically are not trained to work on comprehension skill building. Thus, students typically end up "discussing" the material with their peer tutor until pure repetition finally allows them to pass a mastery test. It seems to me that these strategies do not address the root causes of not having mastered the material on first testing, if that was what occurred (as it most frequently does).

Prior work by my laboratory on the convergence of control systems and complex behavioral analysis (Ray & Delprato, 1989) inspired a new strategy for addressing this problem. I agreed with Kantor's (1970) assessment of the Experimental Analysis of Behavior movement, in that it was too singularly focused on only one element of a student's development—especially as it was expressed through the mechanics of programmed instructional design. Especially relevant, I believe, are dynamics modeled by adaptive control systems and their implications for computerized educational processes that allow computers to aid in the development not only of a student's facility with the content being presented but also skills that eventually transcend the need for supportive tutorial help in learning such content. These processes are best described as being guided by adaptive educational programming, or simply *adaptive instruction*.

ADAPTIVE CONTROL, TEACHING, AND LEARNING

The computers-in-education literature already reflects at least two quite distinct uses of the term *adaptive instruction*. Both uses of the term *adaptive instruction* include fluctuating goals, processes, and/or strategies that adapt to individual learner differences. However, in some of the literature, adaptive instruction describes mostly mechanical accommodations made only for individuals with physical or mental challenges and includes such solutions as alternative input or output devices for the blind or paralyzed users of computers. Alternatively, adaptive instruction describes how traditional content-oriented education is adjusted to address normal individual differences in learning styles, skills, or rates. My work focuses exclusively on this latter meaning and intent of the term *adaptive instruction*.

As the term is used currently, *adaptive instruction* describes adjustments typical of one-on-one tutoring as discussed in the college tutorial scenario. So computerized adaptive instruction refers to the use of computer software—almost always incorporating artificially intelligent services—which has been designed to adjust both the presentation of information and the form of questioning to meet the current needs of an individual learner in a fashion similar to how I would adjust both of these in direct personal tutorials. Traditional information adjustment ranges from simple, such as on-line help or navigational guidance systems, to more complex, such as intelligent agents or "find" systems for collecting and delivering pre-selected types of information or highly sophisticated tutoring systems designed to adjust such things as content presentation complexity or even appropriately difficult assessment materials to meet the needs of a given individual learner. I will focus almost exclusively on this latter use where adaptive tutorial and testing services are rendered. To help the reader understand how such a system works, a brief description follows concerning how adaptive instruction and adaptive testing differ and what they have in common.

ADAPTIVE INSTRUCTION

Adaptive instruction focuses on textual presentation and support (prompting) services that adapt to meet the needs of the user in the best way possible; however, even within this meaning the term often describes at least two different instructional service strategies: strategies that are *homeostatic* (Brusilovsky *et al*., 1996) and those that are truly *adaptive* in the same sense that control systems engineers use the term (Jagacinski & Flach, 2003). General systems theory, cybernetics, and especially adaptive control systems theory views the world as being composed of hierarchically arranged systems (Powers, 1973). These hierarchical systems are defined by unique organizational and operational/process characteristics. Thus, cybernetic systems are those that incorporate feedback in the control and maintenance of a system's structure and/or operational dynamics. Understanding the role of cybernetic feedback helps to differentiate, for example, between homeostatic systems vs. truly adaptive control systems.

Homeostatic characteristics common to home air-conditioning systems serve as a model for almost all modern "adaptive" instructional software systems. Air-conditioning systems incorporate inputs (filtration of outside heat into a room), feedback (current air temperature), a goal (thermostatic setting for desired room temperature), a sensor (thermostat element which is sensitive to the feedback), a comparator (the thermostatic dynamic which allows for differencing between goal setting and current temperature), and a processor (air compressor) controlled by the comparator (thermostat). In this example, the adaptivity is seen when outside temperatures overheat the room, thus causing the current room temperature to exceed the desired setting sufficiently to cause the controller (thermostatic mechanism) to turn the cooling compressor on (and eventually back off), thereby causing a new supply of cooled air to circulate into the room. That is, the air conditioner adapts to the heat by cooling the room, thereby maintaining a homeostatic balance in temperature. Like this example, most so-called "adaptive education systems" are designed to adapt to errors made by a student (analogous to a room being hotter than desired) by helping the student meet the static instructional goal (analogous to the thermostat setting) that has been predetermined by the instructional designer.

Just as a thermostat monitors the room air temperature, such adaptive instruction systems are designed to mirror the current knowledge of the learner—usually through the application of an "automatic knowledge generation" engine—but only to adjust for student failings by adjusting services (analogous to turning a compressor on and off) to meet a singular and pre-established educational content mastery goal (again, analogous to reaching the desired thermostat setting). It is this ability to adjust services that prompts most designers to refer to such instructional design elements as *adaptive*.

Upon closer inspection, this is a somewhat misguided use of the term *adaptive*. It is certainly not consistent with how cybernetic and systems researchers would describe the feedback-driven, disturbance-control dynamics for maintaining stability in homeostatic systems like our air-conditioning example. Systems researchers reserve the term *adaptive* to describe quite a different type of control system—a system that incorporates the capability of adjusting its own homeostatic goals when needed. Purely homeostatic systems incorporate only the previously mentioned metaphorical capabilities of sensing current states, of comparing those states to desired or goal settings via feedback, and then controlling adjustments in the operations of the system to decrease discrepancies between desired and current states. Truly adaptive systems also include the metaphorical ability to learn or *adjust* (another word for adapting) by self-modifying the goal or desired state. This increases dramatically the long-term maintenance or even enhanced development of the system's integrity.

Darwin saw such mechanics of adaptation at work in processes that expanded the variability in existing genetic pools (such as mutations), thereby allowing for evolutionary transformation of the structural and functional capacities and characteristics of entire existing species when changes in environments occurred. But, applying such distinctions and definitions to instructional software systems requires that we understand the limitations of most educational goals as they are built into typical adaptive instruction systems.

As noted, typical "adaptive" educational systems almost always include static goals with dynamic adjustments designed to accomplish those goals. But, to be truly adaptive, educational systems need to incorporate the ability to adaptively teach a student the immediate content as well as the ability to teach higher-level skills that transform the learner in fundamental ways. Learners need to develop reading, listening, and viewing comprehension skills. Better yet, we should add the goal of establishing or improving the student's ability to create self-reflective "mappings" or diagrams of the semantic networks that reflect true understanding (Novak & Gowin, 1984). Eventually, we also need to work on advanced problem-solving or generative behaviors (Epstein, 1993).

Such a system should not only adjust to the momentary (homeostatic) needs of the learner but should also recognize when that learner is becoming more adept at learning the material and should respond by challenging the learner to more advanced forms or levels of learning. That is, adaptive instructional systems should not only improve the students' knowledge base but also their learning skills. To do so requires that such a system be capable of shifting its educational goals as well as its services for helping a student accomplish those goals.

ADAPTIVE TESTING

Many readers will already be familiar with at least one definition of adaptive testing (van der Linden & Glas, 2000; Wainter *et al.*, 2000) as it has already been incorporated into many standardized assessment instruments such as those offered by the Educational Testing Services. Again, I use the term a little differently from the traditional literature. It is not just the content focus, but also the *format* of a question that should change in order for a question not only to assess but also to shape comprehension skills via its role in the tutorial process. Questions offer the reinforcing feedback that reflects successful progress of the student, but questions also help to establish for the artificial intelligence engine which kinds of services (as well as which specific content) are needed to help the student progress. Because this is not intended to be a treatise on adaptive testing, perhaps the best way to convey the subtleties is by concrete illustration.

MEDIAMATRIX AND ITS CURRENT USE IN HIGHER EDUCATION

As an attempt to integrate the various elements presented thus far, let me briefly overview the design and function of a software system called *MediaMatrix*. MediaMatrix was expressly designed to deliver highly integrative adaptive instructional and adaptive testing services, and early versions have been described in detail in prior publications (Ray, 1995a,b; Ray *et al*., 1995). It may thus suffice to give only a cursory summary of the various levels of service and how they manifest themselves to the student user to aid in learning new content while attempting also to improve comprehension skills.

As we have already seen, *prompting/fading*, *shaping*, and *leaning* principles tell us that good teaching is an *adaptive* affair. It requires active and sensitive adjustments as we move a student through the various and concurrent successive approximations that lead to the goal of an effective, independent, and knowledgeable learner. The concept of managing the learning process suggests that we need to be sensitive to where the student is at all times in terms of the student's need for prompting, segmenting content, and reinforcing through testing results. Such principles guide us to begin with the size of content segment best suited to the individual student. *And, that segment may, in fact, vary as a function of the character of the content*. Simple *concepts* built around common-sense terms that denote objects—technically known as *tacts* and *notates* (Skinner, 1957)—will be learned in large segments by most students. *Abstractions* require a much smaller segmenting if the terms are new and somewhat conceptual. Nevertheless, better educated students will find even difficult *abstractions* easy to acquire if their learning history has grounded them in the underlying *concepts* (Catania, 1998).

Thus, what is required is a software system that begins with highly prompted presentations, then follows the learner's responses to questions and, based on a running diagnostic history, has enough intelligence to adapt the programmed instructional material to individual needs. Such adaptations present text in larger or smaller *segments*, *fade* or return concept-related *prompting* as needed, and alter the *questions* between more or less highly *prompted* formats. In addition, gradually shaped responses from the student need to be developed to move the student from simple *recognition/selection* levels of competence to the less prompted and more *generatively* demanding conditions of *full recall or response production* and eventually to demonstrations of *problem-solving* and *generative* skills (Epstein, 1993). MediaMatrix was designed to allow such adaptive content presentation and interactivity.

MediaMatrix begins simply with the presentation of a standard textbook that is published electronically via the Internet through a custom browser interface designed to place all the artificial intelligence programming on client computers and all content distributed from remote server databases. Content includes text, graphics, questions, and student history, as well as data for supplemental tools such as a personal notebook system with printable notes. Free, full-service evaluation copies of the Introductory Psychology textbook and tutoring system, as well as a completely illustrated user's guide showing all features, are available on the Internet (http://www.psychi-ai.com), so no graphic illustrations will be provided here.

The MediaMatrix browser has several alternative presentation modes that alter the user's experience of the content and tutorial services. The simplest of these presentations is the Browse mode for reading with all tutoring services turned off. In this mode, the reader has access to full electronic text and graphics topically organized for unit-based reading where such units are defined by typically outlined topics of the textbook's chapters. The alternative to this simple Browse mode is Tutor mode, which turns on the system to offer adaptive instructional/tutorial services that dynamically adjust through two to five alternatives.

MediaMatrix encourages content programmers to design programmed instruction at a minimum of four concurrent levels, although the number of levels is adjustable by the developer as well as by the instructor. Most levels of Tutor mode function somewhat like any other level from the software's point of view, so each level is managed as an alternative software object. It is through content segmenting (how much prompting is given and the varying formats of probe questions) that levels become functionally distinct. This means, of course, that MediaMatrix can make good content programming possible, but certainly does not guarantee it in and of itself. Effective content programming requires a very deliberate effort to incorporate sound behavioral systems and operant learning principles, and poor adherence to those principles should not lay blame to the principles themselves as too often occurs (Erlwanger, 1973).

Tutor Level One: Fully Supported Shaping of Reading Comprehension Skills

In the Tutor Level One mode of studying textual resources, the text that was previously available for unaided reading in Browse mode is now presented in small successive segments by keeping the target segment in full black text, but dimming out all other surrounding text via use of a light gray font. The size of each targeted segment is determined by content developers according to their anticipation of the comprehension level of the lowest portion of the student population likely to study this material. It represents the segment size students will read before being asked a question on its content and meaning and is usually only one paragraph of text.

In addition to segmenting the text into small units and thereby affording the student an opportunity to learn in small steps through frequent testing, Level One mode aids the student even further by incorporating *stimulus prompts* that assist in determining which words within the text serve as the key *associated terms*, as *defining relations* among terms (intraverbals/notants), and as *concept presentations*. Such prompts are presented through underlining the respective words or phrases, which causes them to stand out within the context of all other words. As within the unprompted Browse mode, many of these underlined terms or phrases may be mouse-clicked to find their definition in a glossary of such terms.

To pass beyond any given segment of presented text, the student must click a Continue Tutor button and is then presented with an appropriately formatted (for that level) question on that segment's content. In addition to the Level One mode being designed for presenting small segments of highly prompted content, it is also designed for the use of highly prompted formats of questioning—in this case, multiple-choice questions that target primary-concept and single-property association development. This level is for the learner who needs a lot of *prompting* and *shaping* to learn the content, and astute readers will have already noted the high-density use of reinforcing feedback by having question-answering results available for every paragraph read. If a question is answered incorrectly, the student is shown the correct answer and is also prompted to reread the segment again. The student is subsequently presented with alternative questions (usually there are 4 to 12 alternative questions for every one-paragraph segment throughout the entire textbook) until a question is successfully answered and the student is moved to the next subsequent text segment.

Tutor Level Two: Successive Approximations to Less Prompted Learning

At the second level of tutoring, text segments typically involve two paragraphs of content rather than one. To move beyond this segment of text the student must answer correctly a fill-blank form of question that mirrors the question items existing in the multiple-choice format. By moving to the fill-blank format, the same concept or property probe is made but without the prompt of having the correct answer available. On all levels, if the student answers a question incorrectly, alternative questions appear for that segment of text until the student answers a question correctly. In eventual certification testing to demonstrate mastery of an entire chapter, the student is given a test constructed of a mixture of previously missed questions (to assess for error corrections), previously unasked questions (to improve assessment reliability), and some questions the student has answered correctly during tutoring (to assure some degree of progress reinforcement).

MediaMatrix incorporates an artificially intelligent inference engine that gives the system its ability to acquire data on the concepts, their associates, and the strength of the associate connection based on the developing history of the individual student's performance. Thus, a mirroring of each student's developing verbal associate network is created from the knowledge engine's data on the responses that a student gives to all tutorial and assessment questions on all levels. Such a system also incorporates an expert's image, or *concept map*, of all primary concepts and their verbal associates for comparative purposes. Overlapping areas between the student's and the expert's maps are reinforced while mismatches are used to select corrective questions.

Tutor Level Three: Further Development of Verbal Associate Networks

Tutor Level Three mode really begins advanced utilization of the special system of artificial intelligence that allows for adaptive test construction, as well as individually targeted advising as supplemental feedback. Prior levels have been presenting the student with questions written with specific associations in mind—associations between primary concept terms and elaborative property terms. In Tutor Level Three mode, this system is able to use the accumulated model-of-the-student information to construct exactly the paired-associate practice questions a given individual student needs on any given topic. Such questions may take the form of a single paired-associate item, or a *word associates recognition* testing form as illustrated:

Is *Thorndike* associated with the *Law of Effect*?

Other terms one might offer as potential associates to Thorndike include the following:

Cats
Puzzle boxes
Operant conditioning chambers
Trial and error
Respondent behaviors
Shaping
Law of Effect
Successive approximations

Students knowledgeable about the differences between the work of B. F. Skinner and E. L. Thorndike will quickly isolate cats, puzzle boxes, trial and error (procedures), and Law of Effect as associated with Thorndike, while operant conditioning chambers, respondent behaviors, shaping, and successive approximations originate with Skinner's work. The student's progress in revising misconstrued connections or strengthening weak associations can also be used as a diagnostic for advising the student about what areas require further study or how to better use the system or even to suggest that the student explore undiscovered areas of relevance.

The text a student reads prior to this type of question will by now include four to six paragraphs (thereby shaping advanced approximations to full-text reading and leaned reinforcement, as feedback has moved from one question per paragraph, a 1:1 ratio, to one question per four to six paragraphs, a 4–6:1 ratio). Further, the previous underlining as prompts will have been replaced and now the student will see only primary concepts and their verbal associates as a notes version of terms and properties. This *terms and properties* list appears much like the list in the above illustration of Thorndike as a paired associate question but is now organized correctly with, for example, Skinner and his associated contributions being one group while Thorndike and his associated contributions form another grouping. These prompts disappear, of course, when questions are presented.

Tutor Level Four: Full Development of Verbal Associate Networks

Tutor Level Four mode presents text in the same form and amount as if the student were in Browse mode, and thus is analogous to any printed textbook. Only the primary concept terms appear in the prompting notes area to help the student identify what this text is attempting to teach, and questions presented for tutoring assessment are in the form of multiple-fill blank associates questions, as pioneered by Verplanck (1992a,b). Such a question presents, for example, the name "E. L. Thorndike" with four subsequent blank field areas where a student is expected to supply associated terms that illustrate the student's familiarity with Thorndike's work (terms or phrases like those presented above in the paired-association item illustration).

By now the student has been shaped to read large units of text with minimal prompts, has acquired the skill to isolate the primary concepts being taught by that text, and has learned to identify (and remember) several appropriate descriptive terms or properties associated with those primary concepts. Note that reinforcing feedback has been leaned to where it now only appears after having read and mastered large amounts of material. Prompts have been almost totally faded away, and if the student cannot maintain this level of behavioral study skill and answers a series of questions incorrectly, the system will quickly move that student back down the tutor-level scale until successful performance is once again established. Establishing successful performance at any level only moves the student to a more challenging level for more practice there, just as if a human tutor were tracking progress and determining what current support needs and challenges should be presented. Success at Level

Four mode moves the student into a fifth and final tutorial mode known internally within the system as the Probe mode.

How the Probe Mode Works

In the Probe mode, students are left to read totally unprompted text very much on their own as they do in Browse mode or when reading any other standard textbook; however, in Probe mode, a variable schedule is at work for presenting the student questions. The question bank used in this mode is a collection of all questions improperly answered during study at all lower tutor levels. Any question answered appropriately is removed from the dynamic Probe Test Bank until the student has exhausted all possible questions, whereupon the student is praised for having graduated to unassisted learning levels and is offered the alternative of continuing to work in Browse mode or going directly to Assess mode for certification.

MORE ON ADAPTIVE PROGRAMMED INSTRUCTION: PARAMETRICS OF HOW MEDIAMATRIX WORKS

As noted, students actually begin by default at tutorial Level One, which incorporates smaller chunks of information and the more highly prompted selection/recognition form of question. As a student succeeds with Level One, Level Two is introduced. But, a student who begins either to falter or to excel at Level Two is automatically moved to either Level One or to Level Three, respectively, by the artificially intelligent inference engine of MediaMatrix. The definitions of *excel* and *falter* are *algorithmic*, and the parameters are fully adjustable by instructional designers to be either more or less sensitive. The system defaults to a combination of passing a series of questions with a specified average *accuracy* score plus a minimal *fluency* rate (Binder, 1993). A running average for six successive questions that falls above 90% accuracy and less than the maximum (30 seconds) time allotted to answer a question *(fluency)* moves the student to the next higher level. Alternatively, a running average for six successive questions that falls below 60% accuracy and 125% fluency moves the student to the next lower level. Again, these parameters (including the number of successive questions used as criteria) are fully adjustable to allow for system adaptation to alternative student populations or content difficulty.

ASSESS AND CERTIFICATION MODES

At any time a student may self-assess by directly choosing the user-mode selection button to access an Assess panel that allows students to construct their own (not-for-credit) quiz made up of any number and any types of questions on a given topic of text. Unlike tutorial questions that supply a lot of feedback and allow reviews of the related material for each question, the Assess mode questions give no immediate feedback. Only when the test has been completed is a summary of the test performance and a question-by-question diagnostic offered to the student. Finally, a student who feels sufficiently confident of progress in studying all topics within a chapter of the text may select the Certification mode. This screen is included as a part of the textbook application and offers the student a list of all chapters (with associated completion deadlines) currently still available for "mastery certification" testing. Such tests are adaptively composed for each individual student and graded for accuracy. Accuracy scores are submitted to the server for instructor use in determining course grades. Typically, such adaptively constructed, and thus unique, tests may be retaken any number of times prior to each chapter's deadline in order to encourage student mastery.

INSTRUCTOR OPTIONS FOR MANAGING STUDENT CONTACT WITH THE TUTORING SYSTEM

MediaMatrix was designed around a server-residing relational database. This database incorporates records for each individual instructor using MediaMatrix-based textbooks with one record for each separate course and, in the case of multiple sections, each section of that course. Student records also are created which relate each student to the appropriate instructor, course, and section. From the student perspective, the primary portal designed for use of this database is the metaphorical electronic textbook with its various modes of text presentation, including tutorials, self-assessments, and mastery certification testing. The mastery certification testing interface allows the student to take tests (supervised or unsupervised, depending on instructor strategies of use) on each full chapter of the text.

Instructors have their own portal into the database that allows them both to set various parameters and requirements of student use and to view both individual and class progress in certification testing. Thus, instructors have control over how many questions will be used in certification testing, what types of question formats should be used, and how long the student has to answer each question. Optionally, the entire test may be timed rather than just the individual questions. The instructor also has control over deadlines for taking chapter certification tests via this administrative system.

Typical use of a MediaMatrix textbook includes allowing students to retake alternative (adaptively constructed) chapter certification tests as many times as necessary for each student to reach a performance level with which that student is satisfied. In such cases, only the highest grade is typically counted, with all chapter grades being combined to account for, say, 50% of the final course grade. The remaining percentage typically comes from in-class testing and projects.

Among the important parameters an instructor may manage are the offerings of gradient-based *bonus* points for tutoring and self-assessing on topics within a chapter *prior* to certification testing on that chapter. Further, the instructor may implement a policy of allowing students to demonstrate that they can achieve satisfactory scores on certification testing (where "satisfactory" is defined by the instructor setting a minimum percent correct, such as 80) merely by reading printed versions of the text (which are made available as downloadable pdf files) and studying on their own. An instructor might allow any number of such diagnostic demonstrations but will typically limit them to two or three tests (via an optional setting) before having the system require a student to tutor on all topics the system has diagnosed as giving that student problems. Such requirements are expressed by showing the student a list of topics that must be successfully tutored to reopen certification. When all topics in the list have been eliminated, the certification testing is again available for the specified number of attempts to exceed the minimum requirement. Both the "bonus" and the "required tutoring after diagnostics" features were designed to encourage students to use the tutorial system to its full advantage.

Finally, an instructor has the option of actively tracking (for research purposes) all interactions students have with the systems assessment activities. There is a research software system that can query the database and offers summaries of how much time students use tutorial services and assessment services, student successes and failures with various levels of tutoring and types of questions, etc. And, while such research has only begun quite recently, it is already illustrating some interesting aspects of this adaptive tutorial and mastery testing/certification system. So, let me conclude with a very brief summary of some of these very early research efforts regarding the system's use and effectiveness.

EMPIRICAL RESEARCH ON MEDIAMATRIX DELIVERED ADAPTIVE INSTRUCTION

It is truly interesting to observe how students use, fail to use, and abuse systems designed to help them improve their learning skills and knowledge of a given academic subject. MediaMatrix was designed primarily as an electronic replacement for textbooks, not as a replacement of instructors. Of course, MediaMatrix *may* be used as an alternative to web-browser distribution of educational content in distance-based educational settings which minimize instructors in the educational process. My personal views of current Internet-delivered instruction typical of most schools is that the Internet has given us an interesting alternative form for delivering information, but it has yet to really accomplish much in the area of delivering alternative forms of education (*i.e.*, highly interactive and personally adapted services that teach, not just inform). But, getting students in contact with electronic forms of education (where they actually exist) is, itself, turning out to be an interesting challenge.

It is often informally reported among those who implement personalized systems of instruction that student drop rates are typically higher than in traditionally managed classes. In some cases, this is attributed to the loss of contact with scheduled classes, but it may also be largely due to the implicitly increased work loads involved in having to master material at a high level prior to moving forward in the course. Students also frequently find early and frequent testing is highly informative about their probable lack of success without putting in far more effort, and they use that feedback for electing early withdrawal rather than eventual failure in the course.

Like personalized instruction, a MediaMatrix delivered textbook is designed not only to tutor but also to certify mastery of materials published through that system. This mixture of education versus certification makes the electronic system a two-edged sword for students. It both helps them learn and shows them the degree to which they are failing or succeeding in that process. Students who are attempting simply to skim the text before testing find through early feedback that this strategy will not suffice for the course. When such students discover that an instructor expects them to tutor paragraph-by-paragraph and to acquire some mastery level—that often includes actually remembering how terms are used relative to one another—many are dismayed by the implicit expectations. Thus, some students will drop the course quite early, finding it too much work to actually master all the material. Others will complain about the required "memorization" in the course. We have not yet established relative percentages for this, but informally shared experiences find these to be common themes among instructors using the system.

A recent symposium at the meetings of the Association for Behavior Analysis (ABA) included two reports that were data-based evaluations of student use and subsequent performance in courses using a MediaMatrix-delivered textbook on Introductory Psychology (Kasschau, 2000). One study reported by Belden *et al*. (2003) investigated alternative contingency settings regarding (1) the use of bonus point offerings for self-quizzing and tutoring, as well as (2) whether tutoring was required following limited numbers of unsuccessful certification tests. Five different instructors were involved in the comparative study, each with different use settings, as illustrated in Table 1. Table rows may be viewed as qualitatively ranked in terms of what might be considered the most stringent use contingencies and requirements (instructor A) to the least stringent (instructor E). Figure 1 illustrates the corresponding average amount of time students in these respective instructors' courses spent tutoring.

Clearly, the amount of tutor time systematically follows the stringency of the contingency settings for student use. Belden *et al*. reported that all of these differences, except those between instructor D and E, were statistically significant (with $P<.01$). The authors then compared the average certification

TABLE 1
Listing of Instructors and Their Parametric Settings for the Use of MediaMatrix Based Introduction to Psychology

Instructor	Tutoring levels/types of questions	Required tutoring settings	Bonus status
A	Tutoring: 2 levels Certification: MC/FB	Two attempts to reach 80%	Tutor bonus: on Quiz bonus: on
B	Tutoring: 4 levels Certification: MC/FB	Two attempts to reach 60%	Tutor bonus: on Quiz bonus: off
C	Tutoring: 2 levels Certification: MC/FB	Required tutoring off, PSI format	Tutor bonus: on Quiz bonus: on
D	Tutoring: 2 levels Certification: MC/FB	Required tutoring off, lecture format	Tutor bonus: on Quiz bonus: on
E	Tutoring: 3 levels Certification: MC/FB	Required tutoring off, lecture format	Tutor bonus: on Quiz bonus: on

Note: MC = multiple choice; FB = fill in the blanks; PSI = Personalized System of Instruction.

au: spell out?

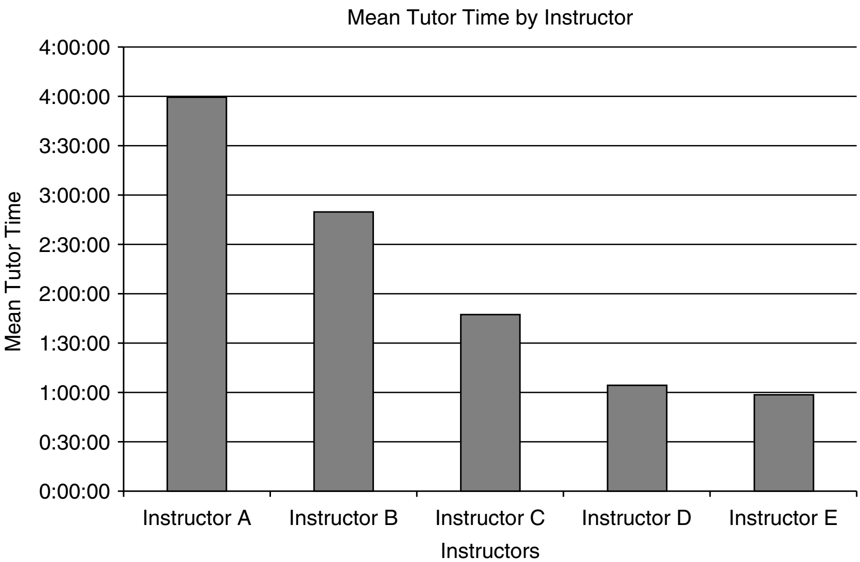

FIGURE 1
Mean tutor time for each of five different instructors using MediaMatrix delivered introductory psychology textbook. (From Kasschau, R. A., *Psychology: Exploring Behavior*. Winter Park, FL: (AI)^2, Inc., 2000. With permission.)

mastery scores for these same instructors and found that each class was approximately the same, ranging from 80% to approximately 83% accuracy in certification testing (this range is noteworthy in and of itself, as this entire range of scores is within a B grade for all five instructors). However, the instructor with the most stringent criteria (instructor A) had students achieving this maximum score within an average of only two attempts, while other instructors' students on average required from four to five attempts.

When the distribution for total student tutor time for all students across all instructors was separated into quartile groupings, there was a systematic and significant increase in maximum certification mastery scores, quartile by quartile. The lowest quartile achieved an average maximum test score of 78% and grades increased systematically quartile by quartile, with 84% as the mean certification score for the highest quartile grouping. Shifting from statistical to practical significance, these results represent just slightly over half a grade-range improvement in test scores for those who tutored most versus those who tutored least. This suggests that getting students in contact with the tutoring system plays a significant role both in the grade they ultimately achieve and how quickly they reach that grade level with respect to retakes of tests. Regardless, instructors found themselves lecturing and/or leading discussions in classes where the average student preparation level prior to attending that class was within the course's B grading range. That also has practical significance.

Finally, a second study reported in the ABA symposium by Butterfield and Houmanfar (2003) compared the use of the Kasschau (2000) text delivered via the MediaMatrix adaptive instructional system to the use of a commercially popular printed textbook. Both types of textbooks were used in a Personalized System of Instruction (PSI) environment where mastery certification testing within MediaMatrix was time limited for each question but not supervised. Certification testing for the printed text was taken via a WebCT based computerized administration of tests supervised by course tutors. Students in both sections also took the same pre-and post-tests covering psychological content common to both textbooks.

Butterfield and Houmanfar reported on a sample of 41 students as representative of their large introductory classes (from 200 to 600 students in each condition) across two different semesters (Spring and Fall of 2002), with their MediaMatrix user settings being those described for instructor C in Table 1. In both semesters studied, the adaptive instruction group had higher increases in average pre-post test scores than the traditional-text/WebCT sections. Almost no students were within the highest performance range (*i.e.*, answering in excess of 21 questions correctly of the 40 questions asked in the pre-and post-testing) for either group during the Spring or Fall term pre-test. But, approximately twice as many adaptive instruction students ended up in this category on post-testing than students from the WebCT group for Spring, and the Fall term had nearly 30% more adaptive instruction students in the high-performance range than students from WebCT. The Fall semester also had approximately twice as many adaptive instruction students both in the A and the B grade range for the final exam compared to the WebCT group. Although not as dramatic in size, a similar difference was reported for the Spring groups as well (*i.e.*, more adaptive instruction students had A's and B's, respectively, on the final exam than the WebCT group). Final course grades earned were more mixed in comparison because they also included points earned from assignments, discussion group attendance, etc.

In fact, truly comparative research on such systems is very difficult to design and even harder to execute. This type of research often comes much closer to representing program evaluation research than controlled experimentation. Over 3000 students were involved in the Belden *et al*. (2003) study, and this makes statistical differences easy to obtain. More important is to consider the practical significance. Their report of over half a grade range difference between those with the least time tutoring vs. those with the most time tutoring is probably much more important to a student than a statistically significant difference between final grades that are all in the B range. Likewise, an instructor is much more likely to be satisfied with a B average for the students' preparation for the lecture, should that instructor still decide to give lectures! It *does* seem that technology *can* address the problem of not having prepared learners for even our largest of university classes. But, there is a downside to this story as well as this obvious upside. It is to this somewhat more cautious element that I now turn.

HIGHER EDUCATION'S FOUR HORSEMEN OF ITS APOCALYPSE

The title of this section says it all. There are four major and significant impediments to accomplishing long-term improvements in our higher education system, and, unlike most pundits, a lack of funds is not really on the list. As is often the case, these four obstacles are often interconnected but are nevertheless worthy of independent as well as interdependent identification and consideration. The four include: (1) limits imposed by current hardware and software technologies; (2) difficulties in changing instructor behaviors (teaching); (3) difficulties in changing student behaviors (learning); and, finally, (4) limits imposed by how little we yet know about the intraverbal dynamics of an expert's verbal associate networks and their effective representation as artificial intelligence.

Let me consider hardware and software first. The system for Internet-delivered adaptive instruction described in this chapter took well over 10 years and in excess of $1 million to develop, debug, and bring to market. Because it is an integrated software and hardware system, it suffers from the current shortcomings of both of these technologies combined. Hardware (servers, Internet backbones, etc.) is still limited and not quite perfected. This translates into occasional server failures, dropped connections, lost data during transmission, and thus frustrations on the part of students, instructors, and technical support personnel alike. But, with the rapid evolution in hardware, this is changing—and that would appear to be a positive. However, hardware evolves so quickly that it also requires a full-time software maintenance program to stay compatible and, thus, to stay in the market. This means development of software is never complete. Worse yet, in systems this complex, the continuity of personnel sufficiently trained to continue software upgrades and to progress to new desirable features is difficult without significant scaling to larger and larger organizations with sustainable financial foundations (translate: larger companies). For such large companies to become involved, the numbers (both in terms of adoptions and its direct associate, revenues) would have to increase substantially. Adoptions and revenues segue into the problem of changing instructor and student behaviors.

Changing instructor behaviors is a first step in changing student behaviors, for instructors set the environmental conditions that shape and maintain not only how students study but also how students establish their expectations for what will be required of them for course credits. It is common that students in a university course expect to do little more than merely race through a long list of multiple-choice questions attempting to recognize answers they have mostly heard before. Remember the Sikorski *et al.* (2002) report that students felt the most important contributor to doing well in courses was to take notes without reading the text? Such students will almost never spend the time required for developing verbal reproductive fluencies sufficient for sustaining an educated conversation with anyone on the topic or for summarizing major concepts covered in a course along with the properties that make such concepts what they are.

Why? Well, what university instructor has the time, resources, or even the inclination to attempt to develop such behavioral repertoires, as verification typically means grading essay tests or something akin to them? To get instructors to change both their educational goals and their teaching behaviors by changing the form of their assignments, their testing formats, or their adoption of new technologies is difficult. I have demonstrated and explained the adaptive educational system described above to a great many instructors, and few have been sufficiently interested to attempt its use. And, even when instructors adopt it as an alternative to their traditional textbooks, they rarely set it for more than two levels of tutorial services or two alternative forms of certification questions. Explorations of why this is so is another essay altogether, but it is demonstrated in Table 1 when the reader explores the column "Tutoring Levels/Types of Questions" (used in certification testing).

From my own experience in using all tutorial levels and thereby asserting very high-performance expectations, students are very quick to attempt to reciprocally manage my expectations by drawing comparisons between my own requirements and those of my colleagues teaching the same numbered and titled course for the same number of credits. Thus, changing student behaviors also turns out to be very difficult because they are given lower-expectation alternatives and have histories of finding paths of least resistance to their own certification (read: "degree") with the minimum of education (read: "knowledge").

This is an institutional weakness emphasized by administrative efforts to meet consumer demands and thus to populate classrooms with more people, not higher-quality learning processes. So, I personally find limited promise for many adoptions of such technologically enabled tutoring systems that require significantly increased student effort, despite measurably superior outcomes. I am afraid that superior outcome measures are almost at the bottom of many instructors' or administrators' list of concerns in today's institutions of higher education.

Finally, let me offer a parting note on what I have learned about the frontiers of behavioral technologies. Beginning with Skinner's treatise on verbal behavior (Skinner, 1957) and continuing with Kantor's linguistics alternative (Kantor, 1977), the field of behavioral analysis has seemingly assumed far more than it has proven relative to human use of language—especially as this activity is related to educational objectives. Few instructors have attempted to specify for students how their course objectives relate to changing the student's working conversational and written vocabulary. Neither do course objectives attempt to specify the range of circumstances in which this vocabulary might be tested or even used. But, even if we wanted to specify such objectives, our collective failure in understanding how an "educated" student might converse with us is substantial. We know a student educated by our course when we experience one, but we have a very difficult time giving a detailed operational definition of our criteria for one! To specify the requirements for knowing one would be to specify how an expert uses our technical vocabularies.

One of my favorite examples learned from experience illustrates the feebleness of our efforts to build representations of the associations such experts make among concepts. Somewhere in the chapter on psychophysiology in most Introductory Psychology textbooks, a student is most likely to read about rapid eye movements (REM) and stages of sleeping as reflected in the scoring of electroencephalogram (EEG) records, and common is the establishment of a direct association between REM sleep and a sleeper's verbal report of concurrent dreaming upon being awakened from such a state. Further elaboration of this REM stage often reflects on the associated paralysis-like state of the muscles, ensuring that vivid dreams of, say, running, do not find us leaving our beds. So, let us assume you were concept mapping (Novak & Gowin, 1984) this paragraph.

To test a student's own efforts to concept map the paragraph in question, you select to ask a verbal associate question based on *dreaming* as the prompting term. You might well expect the student to respond with *REM*, *stages* *of sleep*, *EEG recordings*, and *paralysis-like muscle states*. But, how would you grade a student who also answered to the term *dreaming* with the associated response *Sigmund Freud*? Suddenly, you find a student who is actually better educated than your educational goals (and its related expert system) anticipated. This student just made a totally appropriate, albeit unanticipated, connection between one chapter's discussion of dreaming and another chapter's discussion of the same topic, but from a totally different contextual perspective.

This "contextualistic" interpretation and modification of the associated properties of a term goes on almost infinitely. And, despite contemporary attempts to deal with such phenomena as relational frames of what is technically referred to as *higher-order stimulus equivalences* (Hayes *et al.*, 2001), the practical translation of such efforts is far from sufficient for allowing us to map all potential associations represented in an introductory psychology course. Students quickly discover just how artificial the embedded "intelligence" of an expert mapping system really is, and I have serious doubts that behavioral principles will be up to this task within my own lifetime. So, the best alternatives will likely remain as "what is good enough" for keeping a student contextually focused and limited, but this is not always convincing for students. As such, even the best of artificially intelligent tutoring systems designed to teach conceptual materials and associated vocabulary are likely to garner a lot of student complaints about how "dumb" (and thus unfair) such systems are in grading student test responses.

CONCLUSION

I have attempted to demonstrate a sharp difference between the teaching that takes place outside of the classroom in universities and the kind that is at least afforded, if not taken advantage of by many, students in a more personalized educational setting such as those in the small liberal arts colleges. I have also described a computer-based technology that allows us to bridge that gap with the advantage of at least having more highly prepared learners sitting in our classrooms. I have cited a limited range of emerging research that supports that proposition as well.

Unfortunately, my enthusiasm for the very technology I have spent over a decade developing and testing is damped somewhat by practical factors limiting the likelihood that instructors will widely adopt it, as well as factors leading to student resistance to using it. Our own surveys on this count find variations in responses among students. When asked if they like the electronic testing outside of class, approximately 50% of students in a large class like it moderately or strongly, about 32% like it somewhat, and the remaining 18% dislike it moderately or strongly. Most (90%) like the opportunity to repeatedly test. A different survey was cited recently in the Butterfield and Houmanfar (2003) presentation at the ABA symposium discussed above. These authors reported that 68% (Spring 2002) and 60% (Fall 2002) of the class surveyed would recommend the traditional-text/WebCT version of the course to a friend versus 41% (Spring) and 45% (Fall) of the adaptive class making the same recommendation to a friend, despite the fact that the adaptive class had higher achievement.

Of course, some instructors value MediaMatrix's adaptive system, continue to use it, and even recommend its use to other instructors. We continue to pursue both software and content improvements just as the capability and reliability of the Internet also improves. Student computer literacy may also improve as more and more primary and secondary systems incorporate technology, and this may ultimately have a positive impact on how students view the system as well as serving to decrease the technical support required of instructors who adopt the system. Time will tell as to whether such electronic innovations that improve education will actually be adopted by a critical mass of instructors and whether this will be tolerated by administrators who seem more sensitive to public relations issues than educational ones. One always hopes . . . .

References

Belden, N., Miraglia, K., & Ray, R. D. (2003). *Getting Students To Use Adaptive TUTORIAL services: Strategies, Issues and Outcomes*. Paper presented at the meeting of the Association for Behavior Analysis, San Francisco, CA.

Berk L. E. & Winsler, A. (1995). *Scaffolding Children's Learning: Vygotsky and Early Childhood Education*. Washington, D.C.: National Association for the Education of Young Children.

Binder, C. (1993). Behavioral fluency: a new paradigm. *Educational Technology*, XXXIII(10), 8–14.

Brusilovsky, P., Schwarz, E., & Weber, G. (1996). A tool for developing adaptive electronic textbooks on WWW, in *Proceedings of webNet '96: World Conference of the Web Society*. San Francisco, CA: Association for the Advancement of Computers in Education, pp. 64–69.

Butterfield, S. & Houmanfar, R. (2003). *Self-Paced Interactive System of Instruction (SPIN) and Psych-ai Adaptive Instruction: A Systematic Comparison*. Paper presented at the meeting of the Association for Behavior Analysis, San Francisco, CA.

Catania, A. C. (1998). *Learning*, 4th ed. Upper Saddle River, NJ: Prentice Hall.

Epstein, R. (1993). Generative theory and education. *Educational Technology*, XXXIII(10), 40–45.

Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI mathematics. *JCMB*, 1(2), 7–26. au: pages?

Ferster, C. B. & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*. New York: Kluwer Academic/Plenum Publishers.

Jagacinski, R. J. & Flach, J. M. (2003). *Control Theory for Humans: Quantitative Approaches to Modeling Performance*. Mahwah, NJ: Erlbaum.

Kantor, J. R. (1970). An analysis of the experimental analysis of behavior (TEAB). *Journal of the Experimental Analysis of Behavior*, 13, 101–108.

Kantor, J. R. (1977). *Psychological Linguistics*. Chicago: Principia Press.

Kasschau, R. A. (2000). *Psychology: Exploring Behavior*. Winter Park, FL: $(AI)^2$, Inc.

Keller, F. S. (1968). "Goodbye, teacher . . . ". *Journal of Applied Behavior Analysis*, 1(1), 79–89.

Novak, J. D. & Gowin, D. B. (1984). *Learning How To Learn*. New York: Cambridge University.

Powers, W. S. (1973). *Behavior: The Control of Perception*. Hawthorne, NY: Aldine DeGruyter.

Ray, R. D. (1995a). MediaMatrix: an authoring system for adaptive hypermedia teaching–learning resource libraries. *Journal of Computing in Higher Education*, 7(1), 44–68.

Ray, R. D. (1995b). A behavioral systems approach to adaptive computerized instructional design. *Behavior Research Methods, Instruments, and Computers*, 27(2), 293–296.

Ray, R. D. & Delprato, D. J. (1989). Behavioral systems analysis: methodological strategies and tactics. *Behavioral Science*, 34, 81–27.

Ray, R. D., Gogoberidze, T., & Begiashvili, V. (1995) Adaptive computerized instruction. *Journal of Instruction Delivery Systems*, Summer, 28–31.

Sikorski, J. F., Rich, K., Saville, B. K., Buskist, W., Drogan, O., & Davis, S. F. (2002). Student use of introductory texts: comparative survey findings from two universities. *Teaching of Psychology*, 29(4), 312–313.

Skinner, B. F. (1956). A case history in scientific method. *American Psychologist*, 11, 221–233.

Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1963). Reflections on a decade of teaching machines. *Teachers College Record*, 65, 168–177.

Skinner, B. F. (1968). *The Technology of Teaching*. New York: Macmillan.

Thorndike, E. L. (1898). Animal intelligence: an experimental study of the associative processes in animals. *Psychological Review Monograph Supplements*, 2(4).

van der Linden, W. J. & Glas, C. A. W. (2000). *Computerized Adaptive Testing: Theory and Practice*. St. Paul, MN: Assessment Systems Corporation.

Vargas, E. A. & Vargas, J. S. (1992). Programmed instruction and teaching machines, in West, R. P. & Hamerlynch, L. (Eds.), *Designs for Excellence in Education: The Legacy of B. F. Skinner*. Longmont, CO: Sopris, pp. 33–69.

Verplanck, W. S. (1992a). Verbal concept "mediators" as simple operants. *The Analysis of Verbal Behavior*, 10, 45–68

Verplanck, W. S. (1992b). A brief introduction to the Word Associate Test. *The Analysis of Verbal Behavior*, 10, 97–123.

Wainter, H., Dorans, N. J., Eignor, D., Flaugher, R., Green, B. F., Mislevy, R. J., Seinberg, L., & Thissen, D. (2000). *Computerized Adaptive Testing: A Primer*, 2nd ed. St. Paul, MN: Assessment Systems Corporation.

au: added note OK?