WHAT ARE COMPUTING EXPERIENCES GOOD FOR:
A CASE STUDY IN ON-LINE RESEARCH

John H. Mueller, D. Michele Jacobsen,
University of Calgary,
and Ralf Schwarzer, Freie University of Berlin

ABSTRACT
The argument has been made that discovery-based experiences in learning to program
in LOGO on computers reduce mathematics anxiety. Broadening this perspective, we
speculated that the result should be a decrease in test anxiety and an increase in self-
confidence, with enhanced general academic performance, not just mathematics
performance. Further, we hypothesized that this benefit would not be limited to the
LOGO language, rather it should be a general by-product of self-instruction (discovery)
experiences with computers in general. We developed an on-line survey to explore this
possibility. This chapter will summarize some of the results of this survey and our
experiences with on-line data collection.

Address all correspondence to:
John Mueller
EDPS Department
University of Calgary
Calgary, Alberta
T2N 1N4
Canada
mueller@ucalgary.ca
http://mueller.educ.ucalgary.ca/
403-220-5664 (work)
403-288-7342 (home)
403-282-9244 (fax)
What good comes from all this time and money spent on computers? Among others, Internet widows and widowers alike seem to ask this question daily. This chapter chronicles our efforts to achieve an answer to this question from data collected via an on-line survey. We view this survey as a case study, an exploration of the question and a test of on-line technology as a data collection format, rather than a theory-testing venture. Thus any answers will necessarily be tentative. We report here some preliminary results from about two thousand respondents. In addition to summarizing our findings, we will also address how our own experiences may provide some “lessons learned” for others contemplating such data collection.

We note in passing that in the “real world”, out there beyond our ivory tower, the "payoff" for using computers is often thought of simply in terms of "productivity increases". In other words, the “good” is reflected in annual profit reports, stock values and such. In spite of substantial investment, it has been difficult to document regular, direct gains from computer expenditures in the workplace (e.g., Gibbs, 1997; Landauer, 1995; Norman, 1988, 1993). The disappointing results on the “return on investment” issue are beyond the scope of the present discussion, though one may reasonably suspect that there is some connection with the issues that we will discuss here, compounded by missteps in the manner of technological implementation (cf. Adams, 1996).

**Personal Growth as Value**

In the educational domain, computing experiences have long been justified in other terms, in terms of benefits to the individual, warm and fuzzy things compared to the bottom line. In academe, computing experience is viewed as a possible contributor to an individual’s skill set, either in terms of some specific task mastery (e.g., graphics
software) as in learning a trade, or some broader personal improvement related to
general intellectual stimulation.

This personal growth perspective was emphasized in *Mindstorms*, a popular
book by Seymour Papert (1980). Papert was concerned with declining mathematics
competence, as well as increased mathematics avoidance in course selection and related
problems. He argued that declining mathematics competence might be traceable to the
conventional classroom assessment environment, whereby performance is usually
dichotomized as "right" or "wrong." In the traditional classroom context, failing to solve
mathematics problems is an unpleasant experience. Being "wrong" connotes personal
failure, and eventually fear of being wrong generates anxiety and avoidant strategies
designed to preserve self-esteem. It was argued that declining mathematics scores
might be traced to this characterization of “error” as “bad me”.

As an antidote for this fear-of-failure approach, Papert seized upon the LOGO
programming language as a way to reduce mathematics anxiety in young children.
Children learn to make the "turtle" cursor move about the computer screen using
LOGO commands, in an unstructured, play-like environment. In the process, they learn
that programs do not always work as planned, even for the teacher and other adults.
Papert argued that this experience of an error as a mere step in problem solving
changes the experience of making a "mistake" to something less personal. Errors
become less threatening to self esteem, and thus merely a challenge to further
constructive (i.e., on-task) action. In this discovery framework, a mistake has at least
some benefits, rather than becoming a stimulus for avoidant or escape behavior with
concomitant negative emotional reactions.

Thus Papert expected to find improved mathematics performance following
LOGO experiences, at least if LOGO was experienced in a "discovery" manner rather
than being "taught", and a substantial literature developed to examine LOGO (e.g., Littlefield, Delclos, Lever, Clayton, Bransford, & Franks, 1988). At the risk of overgeneralizing, students do routinely report more positive attitudes about school following LOGO training, but the evidence for improvement in mathematics or problem-solving is less consistently obtained. Still, the anecdotal reports of teachers that students seem to have fun during LOGO sessions is consistent with the notion that a changed perception of the significance of mistakes has taken place. This recharacterization of the meaning of an error should be beneficial for the child's self image and other academic endeavors if that more relaxed perception of errors can be generalized beyond the LOGO setting. (In passing, it is worth noting that the association of the computer with play extends beyond LOGO and computer games. I have long found that if my office door is open and I am at my computer keyboard, passersby do not hesitate to pop in and see what I am doing -- I am seen as “playing” with my computer, so interruption is socially acceptable. However, if I was sitting there at a typewriter, a potential visitor would quietly pass by, I would be perceived as “working” -- and there is less often interest in sharing work.)

Although the LOGO research does provide some context for our project, we do not believe that it is broad enough to provide a complete framework, and so we note LOGO only in passing. For present purposes, “structured” instruction will refer to the conventional format, that is, syllabus-driven, going from A-Z on each topic in sequence, with the prospect of formal evaluation. “Discovery” learning here will refer to alternative approaches such as the undirected play-like format described by Papert, but we will also include other learner-driven formats, such as self-directed problem solving. Obviously there is no inviolate definition of these, and unfortunately some uses of the term “discovery” have a legacy of connotations that may mislead. And surely there are
any number of formats that would be problematic of classification. However, the extremes of such a continuum seem identifiable, and our concern is that the extremes provide environments potentially different in terms of the perceived meaning of making a mistake,

Observing children at the computer, one may see that structured approaches to learning about computers may not be optimal. Specifically, the child will confront a specific, limited problem or issue of immediate interest, such as, “how do I change the font for this material.” This specific problem is pursued to satisfaction, and then the next specific issue is dealt with. In other words, that is not the time to intrude with a comprehensive discourse about the differences between bitmap, truetype, and postscript fonts, screen versus printer appearance differences, kerning, ad nauseam. Unfortunately the comprehensive format is all too characteristic of formal course presentation, covering all aspects of a topic whether needed at the moment or not, in contrast to the "just in time" nature of learning that seems to occur naturalistically. More broadly, this just-in-time perspective is championed by other observers of the way in which computers can promote children’s learning (e.g., Tapscott, 1998).

On closer reflection, we decided Papert’s analysis stopped short of a complete realization. In particular, it was too conservative in viewing the benefits as limited to mathematics performance. We think that there is appeal to the idea that computer learning experiences may alter one’s perception of errors during learning, but we extend Papert’s analysis in two ways. First, there seems no reason to think that this benefit would be restricted to just LOGO experiences to the exclusion of other programming languages, provided that the computer experience occurs in a "discovery" mode. That is, the benefit should accrue whenever the learning experience occurs for immediate problem-solving benefit rather than under the structured
instruction format. For this reason we will not pursue an extended review of the literature on LOGO per se. Although LOGO is an intellectual ancestor for our survey, the original concerns are just too specific. Likewise, to appreciate this potential benefit from an altered view of errors, one need not be a constructivist theorist, because even the classic Thorndikian tradition will suffice. That is, any viewpoint whereby errors tell us something and lead us part-way toward a solution will do -- error elimination as information gain, not errors as “bad me”. Therefore, the main question for us became “do self-instruction and/or discovery experiences involving computers in general yield a changed attitude about how errors fit into problem solving, compared to more traditional structured experiences?”

Second, we see no reason to think that the benefits of such a change in attribution about the meaning of an error would be restricted to just mathematics performance, nor, for that matter, to just computer use. That is, reduced test anxiety and enhanced self-confidence should show broad, general performance gains, as reflected in measures such as cumulative GPA rather than scores in a single course or even a single subject area. If so, then assessing the benefits of such a changed attitude about errors requires looking beyond changes in grades in mathematics courses. Further, because these more general effects involve changes in self-confidence and related dimensions, the effects may not be fully manifest in improved general performance until a few years later, another reason to look beyond effects in a single, immediate course.

Therefore, what are computer experiences good for? We decided to look at the way computer experiences might be associated with learner features such as test anxiety and self-efficacy, and whether there was a benefit to academic performance. Do they just affect mathematics performance or are the results broader? Do the results
derive equally for structured courses and self-instruction? More specifically, does an open-ended computer experience reduce evaluative anxiety? Does it increase self-efficacy? Do these self enhancements improve overall academic performance, but perhaps only in a long-term or delayed timeframe, years after the elementary school experience with the LOGO turtle?

**Approaching the Question**

How to address these questions? One traditional strategy would be to follow a group of kindergarten children through their elementary and high-school years. However, longitudinal research at best provides exasperatingly delayed reinforcement to the investigator, and further a longitudinal sample like this would be quite restricted in terms of size, locale, curriculum, and so forth. Another traditional source of respondents would be to survey first-year students at University. However, the campus sample misses out on those who do not continue on to college, it restricts the experiences to a specific geographic locale and educational regimen, and has other generalization problems as well.

We decided it was appropriate to throw this question open to the world by setting it up so that the survey could be accessed by anyone via the Internet, using a web browser such as Netscape. This procedure raises some potential issues of sample representativeness, but it does assure a wider variety of background experiences in computing, it will tap ages beyond high school and University, and it allows the assessment of long-term benefits. Further, it potentially taps a wider range of test anxiety and efficacy than would be the case by restricting the sample to University students. That is, students who really have test anxiety may be less likely to continue on to University, so a University subject pool would tend to be lower in test anxiety than the population at large. (Of course, it might be asked whether really high-anxious
people would seek out a WWW survey, but presumably an on-line survey is not
construed as a personal evaluation and thus would not elicit such evaluative concerns.)
Likewise, a college-only sample would presumably reflect a group with self-efficacy
scores biased to the high end, warranted or not.

The most obvious "bias" to a sample gathered on-line is that the respondents will
all be computer users, but that after all is what this particular study is about -- what
benefits derive from computer experiences. No matter how we might have
alternatively approached the question, it would have involved computer users. Having
said that though, there may be general issues of representativeness for other survey
topics, but some of our results may be heartening in that regard even, and there may
be some general strategies which will help allay some concerns about generalizability.
On the plus side for any topic, there are trade-offs to be gained in terms of breaking
down regional educational practices and other localized sampling constraints. That is, if
we avoid the homogeneity of the Introductory Psychology class, we will get
heterogeneity and thus perhaps results closer to the population at large. However, it is
not necessary to always throw your survey open to the world, one can restrict it to a
recruited local subject pool, even use password access, basically just using the web
browser software in lieu of paper and pencil. Likewise, in some cases it may be useful to
simultaneously conduct a traditional paper-and-pencil survey for comparison (e.g.,
Pasveer & Ellard, 1998).

These and related sampling issues have been discussed by Smith and Leigh
(1997) and others (including this volume), and though we might have wished for such
counsel when we started this project (December, 1995), the best resolution of these
cconcerns remains to collect the on-line data to see whether the potential concerns
materialize or not. Interestingly, as will be discussed below, some of the concerns we
have often heard expressed (e.g., gender bias) simply have not shown up in our sample, even with minimal control on access.

Ultimately we must be realistic: a survey is a very limited way to test a hypothesis, it’s a far better way to set up a hypothesis. Thus the more reasonable expectation in most cases should be to gather some evidence in an exploratory fashion, then pursue a controlled laboratory experiment if indicated.

**Getting Started**

With this idea in late 1995, we were stimulated by the appearance of commercially-oriented surveys implemented as web pages, something which by now has become ubiquitous. These marketing surveys made it clear that this on-line methodology was mechanically possible. It would have helped to have had models and turn-key software to set up our survey web page (e.g., Baron, 1998, and Schmidt, 1997), but that part actually turned out to be fairly straight-forward, just using local help files and on-line guides to programming in hypertext markup language (HTML), a little discovery learning of our own. Your institution likely provides some common shared (CGI, or common gateway interface) scripts for processing forms, and likely allows you to host your survey on the institution’s server, but if not almost any commercial Internet service provider will do so for a very nominal fee. Looking back, the HTML part of the project was not much of a problem at all, and by now undergraduates can do it over lunch. But we have enhanced our own self-efficacy as a result.

**Ethical Considerations**

We noted in our initial deliberations that there did appear to be some notable omissions in the commercial or non-academic surveys. For example, in contrast to our often tedious campus procedures for informed consent, non-academic surveys generally launched right into the questions without any component such as our
obligatory consent form. Specifically, these non-academic surveys involved no or minimal assurance about the anonymity and/or confidentiality of the answers that a respondent provides. Even when such "assurance" was offered, the credibility of the offer seemed suspect as soon as follow-up e-mail contacts began to arrive. Although this is not the place to pursue it, there does seem to be an issue here having to do with differential standards in the academic and private sector with regard to data collection on-line.

Nonetheless, as academics we thought we needed to model the paper-and-pencil informed consent format. We included a preliminary consent form as a web page with the usual basic information about the purposes of the survey. It also assured the potential respondent that they could quit at any time with no penalty. The consent form also included the usual assurance about anonymity and confidentiality. Most routines save the data with the numerical IP address as part of the subject’s data, and this is usually sufficiently anonymous (e.g., 136.159.200.9, or 24.64.69.5). In general, to further reinforce the assurance of anonymity, it is a good idea to not ask for an e-mail address. Alternatively, one can ask for an e-mail address (e.g., to provide follow-up results) after storing the data, on a separate web page, so that the e-mail address goes into a separate file from the survey answers.

Another variation from a paper survey is the impossibility of getting the web-page consent form “signed” to acknowledge consent. With the web-page, the participant can only click on the “Agree” button. One can imagine the campus lawyers worrying profusely at this point. However, from a broader, real-world perspective, this practice has been common-place with computers for sometime now. For example, when installing software one invariably has to only click on a series of “Accept” buttons for the License Agreement. Thus we decided that if clicking an “Agree” button is good
enough to be binding for Microsoft’s lawyers, it should suffice for on-line research. Guidelines incorporating these and other points for the consent page can be found at Mueller (1997), and also in Smith and Leigh (1997).

In reality, it is hard to imagine a more “voluntary participation” format than web surfing. For this reason if no other, it seems wise to keep the consent form as brief as possible, within the constraint of informed consent. If the participant elects to stop, there aren’t any face-to-face social cues to stop them going elsewhere, no chance to answer questions for clarification, etc. And, for similar reasons, it is hard to imagine that there could be any sense in which a penalty could be imposed for a decision to withdraw. In pragmatic terms, one can hardly imagine a more superfluous clause to be issued to someone half-way around the world than the assurance they could stop at any point without penalty. There may be circumstances that dictate a “high entrance barrier” to minimize drop-out (cf. Reips, this volume), but we think that one should avoid implementing that strategy except by deliberate design.

It will be detached in time from the consent form, but it also seems desirable to provide a debriefing document after the survey answers are submitted. This is done as a matter of courtesy, and as a matter of completing the informed consent agenda, but it also should avoid a flood of e-mail asking repetitious questions.

Thus the usual academic constraints for informed consent can be honored. Further, the web-based survey is cheap compared to a hard-copy mailed survey, the survey is there 24 hours a day without the need to pay a research assistant, there is no need to barter with a department head for lab space, and the data so collected routinely accumulate in a data file without even the need for data entry for statistical analysis. Clearly there are merits to the on-line approach, whether the sample is controlled or open-ended, and whether the survey topic concerns computer-related issues or not. But
in any event a question about the effect of computer experiences certainly seemed appropriately examined on-line.

**Method**

Thus, in January of 1996, we launched a virtual survey, exploring the issue of how computer experiences might be related to self-efficacy and similar traits, and whether there would be any concomitant benefits to academic performance or other measures of achievement. First we located some measures of test anxiety and self-efficacy to convert to the on-line format. This on-line use requires honoring the usual conventions around copyrights, so we obtained such permissions, and noted this fact on the web page as a caution for others.

We adopted the *Test Anxiety Inventory* (TAI; Spielberger, 1980) to assess test anxiety. The TAI provides two subscale test-anxiety scores: *worry* (task-irrelevant cognitions), which reliably has a negative association with performance, and *emotionality* or arousal per se, which may or may not hinder performance. We elected to stay with these traditional two subscales for analysis, though there is recent evidence to suggest that other factors may be identified (e.g., Hodapp & Benson, 1997).

To assess self-efficacy, we adopted the *Generalized Self-Efficacy Inventory*, developed by Schwarzer and Jerusalem (1995). In other words, we did not just look at "computer anxiety" (the negative perspective traditionally taken in educational technology research), nor even its positive counterpart, "computer efficacy". We could have looked at specific efficacy as well as general, but as noted above re the consent form, there is a need to keep the on-line experience digestibly short. In fact, finding an effect for generalized self-efficacy seems more problematic, and thus our procedure constitutes a more conservative or demanding test than a more focused measure such as "computer efficacy."
We then developed several demographic items having to do with the usual suspects, age, gender, education, grade point average, but also income, geographic region, handedness, and extraversion. Handedness was assessed by a 9-point self-report rating, having to do with writing, sighting, swinging a hammer or racquet (1=left, 9=right). Extraversion was also a 9-point self-report rating, with 1 being “extraverted and liking to work in groups” and 9 being “introverted and liking to work alone”. Although there are other reasons for looking at interactions between anxiety measures and extraversion, in this case the purpose was more modest: presumably extraversion could have some connection with preferences for self-instruction (solitary) versus structured (group) classes, for example. Single-item assessments such as this may not usually have high reliability, but a longer assessment would have extended the on-line experience. Therefore, given the exploratory nature of the project and because handedness and extraversion were secondary concerns, single-item coverage seemed sufficient at this stage.

Finally we developed some items having to do with specific computer experiences, especially programming languages, including LOGO, BASIC, HTML. We also asked how that expertise was obtained, that is, self-instruction vs. courses, with no detailed explanation of these formats. To some extent, self-instruction may be somewhat age-linked, in that some years ago there were few courses in computer programming, and thus most older respondents may be self-taught without choice have been a factor.

The complete survey is at http://www.acs.ucalgary.ca/~mueller/tai-consent.html. To start soliciting participants, we seeded this site’s address (URL) into several web search engine databases (e.g., WebCrawler™, YAHOO™, Alta Vista™, and so forth). Thus web users searching for “anxiety”, “efficacy”, “computer
programming” and such at these search engines would find our web site. Entering the URL into several search engines seemed desirable as a way to overcome any selection bias that might be present for visitors to a specific database.

Respondents finding it by those search engine databases would be somewhat delayed because updates to the database may take several days or even weeks. Therefore, to get more immediate response, we also directly announced the survey URL to several educational technology discussion groups (both Listservs and Usenet newsgroups), psychology discussion groups, computer magazine web pages, educational technology sites (e.g., International Society for Technology in Education), Psychology sites (e.g., American Psychological Society), and otherwise promoted its existence more directly than the web search engine route.

At this point in time, over two years later, any differences in respondent origin have presumably changed and averaged out. That is, the early respondents likely came primarily from the discussion groups, whereas later respondents came largely if not entirely from the search engines. Because of the anonymous nature of the identification in the date file, we are not able to classify respondents in terms of how they found the survey. In hindsight, one could duplicate the survey with multiple URLs, and then seed one URL to Psychology discussion groups and the other URL to Technology discussion groups, for example. Then one could record the probable origins of the respondents by simply creating separate data files from each URL, but we did not incorporate this scheme here. By using a variety of search engine seedings and a variety of discussion group postings, we presumably obtained broader representation than would be the case had the topic dictated a more focused target sample. The general issue of why some of the recipients choose to respond and others did not is still present, as it is in any survey format.
In these postings, we tried to carefully follow another set of ethical considerations, namely those which by convention constitute good taste for on-line communication. For example, we did not post the announcement to groups that would likely not be interested, we identified it clearly as a survey solicitation in the subject line, we kept the invitation to participate short, did not repeat solicitations, and otherwise tried to behave in a manner consistent with good netiquette (Net Etiquette, e.g., Rinaldi, 1998).

Sample Characteristics

Out of 2,123 useable respondents (March, 1996, through December, 1998, scanned for duplicate records and other problems), roughly half were 25 years of age or younger, but about 20% were 40 years or older. Importantly, only 54.5% were male, lower than might have been projected given popular stereotypes re gender differences in computer use. (For purposes of discussion, here and in the following, results will be described as “significant” when they are associated with p < .05 or less.) As Figure 1 shows, this gender distribution was true across ages, with a nonsignificant chi-square for Age X Gender, $X^2 (12) = 9.49$. The most common education level was "some college," but education level ranged from about 28% with "completed high school" to about 20% with a Masters degree or beyond, and the Education X Gender chi-square was nonsignificant, $X^2 (7) = 4.39$.

The average reported income level was the $30,000 to $45,000 category, but income ranged up to the low six figures and included a number or respondents below $15,000, and did not show a significant chi-square for Income X Gender, $X^2 (8) = 11.77$. The respondents were largely from North America, some 76% being from the USA or Canada, with about 10% from Europe. The Region X Gender chi-square was significant, $X^2 (5) = 25.38$, p < .001, as men outnumbered women in the three most populous
regions and not elsewhere, there being a floor effect in the latter regions.

These demographics compare well to the on-going Georgia Tech GVU (1998) surveys of on-line users, and other surveys about WWW usage (e.g., EuroMarketing, 1998; InfoQuest, 1998; Klopfenstein, 1998; NUA, 1998). That is, the sample we obtained is not just “freshman male college students in engineering” as popular media stereotypes of computer users might have us fear. Nor is it the profile of the typical Introductory Psychology class, or even the University freshman class. Perhaps most notably, it is far more gender balanced than most stereotypes allow, in fact more gender balanced than Introductory Psychology classes. Of course, we must also remember that in striving to achieve a sample that is more representative of the real world than a freshman Introductory Psychology class, the result is heterogeneity -- variability associated with relaxed control in sampling (e.g., greater age range, greater income range, geographic and presumably cultural variation, etc.). Perhaps one can see this in two ways, either as the price one pays in lost control for going off campus, or as the reward, but we prefer the latter perspective. There also are problems associated with having a large sample, such as small differences becoming statistically significant though lacking clear practical importance. However, being unaccustomed to an abundance of data, we have to carry on and make the best of this problem.

**Replicated Results**

Another way to check the validity of the on-line sample is to show that some well-known results from conventional laboratory research were also obtained here; if so, then that would increase confidence in accepting the more novel results here. In the case of test anxiety research there are a number of marker patterns that can be used, and these are thoroughly documented in a meta-analysis by Hembree (1988) and a new
treatise on test anxiety by Zeidner (1998). For example, it is commonly found that women report significantly higher levels of test anxiety than men, and that was the case here for both the worry subscale, $t(1981) = 5.60$ ($M_s = 13.3$ vs. 12.1), and the emotionality subscale, $t(1981) = 8.64$ ($M_s = 16.8$ vs. 14.6).

It is also generally found that test anxiety is negatively correlated with global performance measures such as GPA, and that was found here for total test anxiety, $r(1823) = -.26$. This negative correlation is also usually greater for the worry subscale than for the emotionality subscale, and that was the case here, $r_s(1823) = -.34$ and -.18, respectively. Granted that these are not large effects in some respects, but in fact the same correlations in more controlled experiments are typically modest as well. In other words, the correlations are consistently in the direction of poorer performance with higher test anxiety, but the magnitude of the relationship is never large enough to indicate that test anxiety is the only thing affecting performance. Nonetheless, most researchers are satisfied with the reality of the relationship, and most students would prefer to minimize the handicap.

Therefore, although there can always be some question about a sample’s validity, several common test anxiety marker effects were present here. These replicated marker effects thus seem to lend some credibility to the interpretation of relationships that are examined for the first time here. In addition to the replicated test anxiety results, an item analysis and factor analysis of the self-efficacy scale results here proved to be highly similar to those for the many pre-existing datasets using the self-efficacy scale (e.g., Schwarzer & Jerusalem, 1995), and these data have been reported in more detail elsewhere (Schwarzer, Mueller, & Greenglass, 1999). Thus these replications add to the accumulating evidence (e.g., Pasveer & Ellard, 1998; Smith & Leigh, 1997; and chapters in this volume) that on-line samples yield results that mirror those found in
more controlled settings. Granted, continuing to monitor the issue of how results derived from conventional formats compare to on-line results seems prudent, but it begins to appear that such concerns may have been over-stated in terms of many areas of inquiry.

In sum then, insofar as these on-line data overlap with well-known effects for both test anxiety and self-efficacy, we can proceed with some confidence to examining the newer patterns of interest here. Designing some such marker patterns into one’s survey seems a prudent and useful strategy for any on-line survey. Marker patterns may lack something in statistical elegance or finality in the interpretation of cause-effect relationships, but confirming old truths in a novel context is still comforting when trying to assess the legitimacy of newer findings.

**Newer Results**

**Efficacy**

Table 1 presents some basic patterns, a correlation matrix for the test anxiety subscales, self-efficacy, and the demographic variables. Although we have a large number of degrees of freedom here, most of these relationships do seem plausible and thus perhaps they are not merely statistical artifacts. Efficacy, as might have been expected, was negatively correlated with both test anxiety subscales, $r_s (1974) = -.35$, and positively correlated with grades, $r (1813) = .19$. Efficacy was negatively correlated with self-reported extraversion, that is, extraverts reported higher efficacy, $r (1993) = -.14$ (but since extraversion did not correlate with income or GPA perhaps that is simply a positive attitude as opposed to achievement feedback). The relationship is quite small, but efficacy was positively correlated with level of education, that is, higher education went with higher self-efficacy, $r (1980) = .08$, but efficacy was not correlated with age. Efficacy was significantly lower for women than for men, $t (1969) = 5.72$ ($M_s = 28.5$ vs. 28.9).
Test anxiety

In addition to the gender differences for test anxiety, and the common GPA deficit for high test anxiety, anxiety was significantly negatively correlated with education level for the worry subscale, $r (1982) = -.14$, and even for emotionality, $r (1982) = -.05$. The inverse correlation of anxiety with education level may mean that with further education one becomes less threatened by evaluation. For example, over time in college, students may learn better coping skills to cope with their test anxiety, and perhaps experienced students also develop better skills for dealing with exams, a test-wiseness effect. Alternatively, or additionally, the anxiety-education correlation may also mean that high test-anxious students progressively drop out at each higher education level. If that is true, then what we see at the post-secondary levels is a more restricted range of test anxiety, missing much of the extreme upper portion of the dimension. This anxiety-education correlation is not a large effect, but it fits readily with theoretical conceptions of test anxiety based on laboratory research.

In the final analysis, higher levels of worry were associated with lower income, $r (1665) = -.08$, and this anxiety-income relationship also was found for emotionality, $r (1665) = -.08$. These are both small effects, but nonetheless plausible.

Computer use

Our initial motivation was to determine whether there were any benefits to self-efficacy or test anxiety from learning to use computers. Although this is an easy question to formulate in the abstract, there are many specific manifestations with some bearing on the answer. One historically important question is how LOGO instruction may affect measures of efficacy, anxiety, and performance. As noted above, Papert
(1980) argued that LOGO instruction would defuse mathematics anxiety. We do not have data here on mathematics performance per se, so instead we deal with a more tenuous relationship, namely the sense in which a benefit from LOGO might be more general than mathematics confidence, and more delayed or enduring in impact than a grade in an immediate mathematics course.

There are a couple of ways to examine this question. One involves a comparison of efficacy scores for those with and without any LOGO experience (ignoring how the expertise was acquired), as shown in Figure 2. The other requires a closer examination of the type of LOGO instruction within the subset of respondents who have had LOGO expertise, as shown in Figure 3.

**LOGO per se.** Respondents with LOGO experiences of any kind (n = 337) reported significantly higher generalized efficacy scores than respondents lacking any LOGO experience, t(1991) = 4.41 (Ms = 30.4 and 28.9, SDs = 5.16 and 5.22). Remember, this outcome reflects an assessment far removed in time from the actual course, not just a short-term end-of-course evaluation. Furthermore, it is not just how a computer experience diminishes computer or mathematics anxiety (or increases computer efficacy), instead this is apparently a broader increment in efficacy that persists over a long period of time.

There was a difference for the worry subscale of test anxiety, in that those with past LOGO experiences reported significantly less worry, t(1991) = 2.92 (Ms = 11.9 vs. 12.8, SDs = 4.39 and 5.07), and likewise they reported lower emotionality, t(1991) = 3.82 (Ms = 14.4 vs. 15.8, SDs = 5.10 and 6.03). However, there was no difference for GPA or income due to the presence or absence of LOGO experience, ts < 1.

**LOGO method.** For those who reported LOGO expertise, the instructional
method was classified as either discovery \((n = 111)\), structured \((n = 179)\), or self-taught \((n = 120)\), and these three categories were used as the basis for a one-way analysis of variance. There was a significant difference for generalized efficacy, \(F(2, 407) = 6.76\), as efficacy was lower for the structured group \((M = 29.3, SD = 5.61)\) than for either discovery or self-taught, the latter two not being significantly different \((Ms = 30.4 \text{ and } 31.5, \text{respectively, } SDs = 5.01 \text{ and } 4.87)\). Efficacy being higher with self-instruction seems consistent with the idea that self-instruction avoids the “failure” connotation of error, as does the discovery procedure.

LOGO instruction method did not lead to significant differences for either the worry component of test anxiety, \(F < 1\), or the emotionality component, \(F = 1.21\). Method of instruction likewise failed to show an effect with GPA, \(F = 1.11\). There was a significant effect for instruction method in terms of income, \(F(2,340) = 3.37\), as income was lower for the respondents who had structured LOGO training than for either discovery or self-taught LOGO, the latter two not differing.

Self-taught vs. Courses. Some of the same expectations can be extended to any other computer experiences obtained in a self-taught as opposed to a structured course format. One of our survey items asked subjects how they had acquired most of their expertise: self taught or through courses (cf. Figure 4). Reported efficacy was not different for self-taught \((n = 983)\) vs. course-based \((n = 126)\) backgrounds, \(t(1115) = 1.58\) \((Ms = 29.4 \text{ vs. } 28.6, \text{respectively})\). However, self-taught respondents reported significantly less worry, \(t(1115) = 2.24\) \((Ms = 12.6 \text{ vs. } 13.6)\), and less emotionality, \(t(1115) = 2.26\) \((Ms = 15.4 \text{ vs. } 16.6)\). There was no significant difference for income, \(ts < 1\), but self-taught respondents did report higher GPAs, \(t(1016) = 2.16\). Self-taught respondents did rate themselves as more introverted (i.e., work alone) than
respondents who said they had more often taken courses, $t(1115) = 4.52$ ($M_s = 6.2$ vs. 5.3).

BASIC course. Respondents could similarly be classified as either having or not having expertise in BASIC programming. BASIC is a common programming language, but not one with a specific rationale associated with it such as that which Papert (1980) provided for LOGO. Nonetheless it offers a way to assess the generality of the LOGO findings in regard to associations with self-efficacy: Does LOGO experience have some unique benefit, or is the enhancement of efficacy also associated with learning to program in BASIC? If BASIC also has a comparable effect, then the decision to distance this project from the legacy of the LOGO literature would seem justified.

As Figure 5 shows, those with expertise in BASIC ($n = 781$) showed significantly higher generalized efficacy scores than those respondents who did not ($n = 1233$), $t(1991) = 3.70$ ($M_s = 29.8$ and 28.9, $SD_s = 5.10$ and 5.30). There also was a difference for the worry subscale of test anxiety, in that those with BASIC experiences reported less worry, $t(1991) = 2.44$ ($M_s = 12.3$ vs. 12.9, $SD_s = 4.66$ and 5.16), and likewise they reported lower emotionality, $t(1991) = 2.99$ ($M_s = 15.1$ vs. 15.9, $SD_s = 5.54$ and 6.11). However, there was no difference for GPA or income due to BASIC experiences, $t_s < 1.$

CONCLUSIONS

This has been working paper for some time, but at this point some conclusions seem in order with regard to the substantive questions, and also with regard to the value of the web-based format for addressing these questions.
It seems clear that there is some evidence consistent with the general hypothesis. Specifically, it appears (Figures 2 and 3) that discovery and self-taught LOGO experiences were associated with subsequent higher generalized efficacy -- not just efficacy in the domain of computing or mathematics, and simultaneously lower test anxiety. However, the efficacy increment was not associated with higher achievement, at least as assessed here in terms of GPA (nor as perhaps assessed to a lesser extent by income).

It was of interest to determine whether this pattern of benefits was found for other programming languages. The results of the omnibus “self-taught” analysis (Figure 4) above are supportive of the idea that the benefits are not confined to LOGO experiences. Furthermore, the BASIC course results (Figure 5) are likewise consistent with the idea that a variety of computer experiences may be associated with higher efficacy and lower test anxiety. In sum, there seems to be nothing specific to LOGO, rather the result is a more general phenomenon as we argued at the outset.

So we can tentatively conclude that there are apparently some positive benefits associated with computer experiences, especially if those experiences do not involve formal courses. However, there are the usual cause-effect concerns associated with survey data. That is, although there may be correlations between computer experiences and self-reported efficacy (and anxiety), it cannot be rigorously claimed that the computer experiences directly caused the changes. It is a possibility though, so we cannot simply dismiss the possibility because of the correlational nature of these data anymore than we can accept it.

Converging data from other investigations will help resolve the concern with causation. Such data could and should come from any number of inquiries, but a couple of illustrations should clarify how, in general, one could proceed to tease apart
correlation and causation here. For example, one could now consider a situation where children are assigned to either a computer project or other projects, such as reading literature, writing essays, or second language learning, in each case half with the self-instruction or problem-driven format and half with a comprehensive syllabus-driven format. By taking measures of efficacy (anxiety) before and after these sessions, one could determine the extent to which computer experiences do or do not uniquely produce (cause) self-efficacy, among other things. Alternatively, one could arrange a psychology class where projects involve using computers and on-line communication, such as turning in assignments as web pages, with a corresponding class where the work is not at all related to computers. After administering a preliminary self-efficacy assessment, students could then be given a choice as to which section to enroll in, and then performance through the semester would be monitored. If high efficacy students opt into the computer-based section more readily than the conventional non-computer format, that would seem to suggest more correlation than causation perhaps, though historical experiences with computing would have to be examined as well.

There is also the question of the limited magnitude of the patterns observed here, that is, perhaps the apparent patterns here are merely being driven by the large sample size. Practically speaking, by comparison, even in laboratory experiments the magnitude of the anxiety-performance relationship is modest, though quite repeatable (cf. Hembree, 1988, Zeidner, 1998). Presumably the smaller effects are obtained because the constructs in question are multiply determined. That is, even though self-directed computer experiences may be one factor that enhances efficacy, no doubt there are many other factors which contribute to increased self-efficacy, and likewise for the factors which reduce test anxiety. Further, no doubt there are factors other than efficacy and anxiety which contribute to better performance. Thus it is reasonable that any one
component might show a limited effect no matter how valid or repeatable. Again, converging data will help resolve this concern.

Nonetheless, at this time we feel that the on-line procedure has proven to be of value in this case. At this point, we are reasonably satisfied with the results, and with the methodology of a web survey. One question is whether the particular question we examined here could have been assessed as well, or better, using more conventional data collection procedures. It does not seem likely that a more definitive answer would have emerged with either a direct intervention in a classroom or a longitudinal design, as those approaches substitute other limitations of their own, as noted earlier. For example, an intervention to teach computers to some students in a specific school would produce a result limited geographically, pedagogically, and in other respects. Methodologies involve trade-offs, and our interest here, as noted at the outset, was exploratory rather than theory testing, assessing the utility of the on-line data collection format at least as much as the computers and efficacy question itself. With a more heterogeneous sample than otherwise readily achievable, we have data consistent with the idea that self-directed computer experiences may yield enhanced self-efficacy and reduced test anxiety, perhaps because the psychological significance of an error had been changed.

Is our satisfaction with on-line research simply a unique circumstance produced by the mating of a computer-based data collection format with a computer-related topic? To the contrary, we don’t think the suitability of the format is limited to just computer-related content. For one thing, the other efforts reported in this volume show the applicability of the on-line format to other, non-computer topics. Furthermore, we have become sufficiently attracted to this form of data collection that we have begun other on-line research, where the topic does not involve computer-
related content. One of these project considers the extent to which the state of knowledge in the neurosciences can be applied to classroom educational practice. The answer to this is somewhat different when seen from the perspective of neuroscientists as opposed to the way educators understand the state of neuropsychology’s findings on cognition (McCrea & Mueller, 1997). In this case, we are using the web to minimize costs and enhance the reach of the sampling demographics, and there is no computer-related content. Another project uses the on-line format to extend the reach in participant recruiting in a study of attribution in brain-injury cases compared to other traumatic injuries (Logan, 1998). In this case, the number of such participants in any given geographical region is somewhat limited, but the on-line format allows participation from across the country. Another on-line project has used the web format to examine the pattern of technology adoption among university faculty members (Jacobsen, 1997). In this case, the pattern of responses was examined for those who responded on-line versus those who would only complete a paper-and-pencil version, among other comparisons. All in all, we think the results of these various projects vindicate the time we spend at the keyboard!
REFERENCES


TABLE 1
Correlations between anxiety, efficacy, and demographic variables (* p < .05).

<table>
<thead>
<tr>
<th></th>
<th>Emotion</th>
<th>Efficacy</th>
<th>Hand.</th>
<th>Extrav.</th>
<th>Age</th>
<th>Educ.</th>
<th>GPA</th>
<th>Income</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worry</td>
<td>.76*</td>
<td>-.35*</td>
<td>.01</td>
<td>.01</td>
<td>-.05</td>
<td>-.14*</td>
<td>-.34*</td>
<td>-.08*</td>
<td>.13*</td>
</tr>
<tr>
<td>Emotion</td>
<td>-.36*</td>
<td>.01</td>
<td>.03</td>
<td>-.04</td>
<td>-.05*</td>
<td>-.18*</td>
<td>-.08*</td>
<td>-.19*</td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td></td>
<td>-.14*</td>
<td>.03</td>
<td>.08*</td>
<td>.19*</td>
<td>.18*</td>
<td>-.13*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handedness</td>
<td>.04</td>
<td>.03</td>
<td>.03</td>
<td>.02</td>
<td>-.02</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extravert</td>
<td>-.06*</td>
<td>-.10*</td>
<td>.03</td>
<td>-.01</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.49*</td>
<td>.07*</td>
<td></td>
<td>.22*</td>
<td>.02</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td>.18*</td>
<td>.06</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.11*</td>
<td>.10*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure Captions

1. Distribution of survey respondents by gender for age, education, income, and regional categories (y-axis is number of respondents).

2. Test anxiety and efficacy for respondents who either had or did not have a course in LOGO programming ($n_s = 337$ vs. 1786).

3. Test anxiety and efficacy by LOGO instruction method (only for the 337 respondents who had a LOGO course).

4. Test anxiety and efficacy by preferred instruction method.

5. Test anxiety and efficacy for respondents who either had or did not have a course in BASIC programming.
Figure 1
Figure 2

![Bar chart showing scores for EFFICACY, WORRY, and EMOTION with two categories: LOGO and No LOGO. The chart displays higher scores for EFFICACY compared to WORRY and EMOTION for both categories.](image-url)
Figure 5

A bar chart showing the comparison between two conditions: No BASIC and BASIC, across three categories: EFFICACY, WORRY, and EMOTION.

- EFFICACY: No BASIC has a score of 30, and BASIC has a score of 35.
- WORRY: No BASIC has a score of 10, and BASIC has a score of 15.
- EMOTION: No BASIC has a score of 15, and BASIC has a score of 15.

The y-axis represents the score, ranging from 0 to 35.