

A Digital Tutor for Accelerating Technical Expertise

J. D. Fletcher
Institute for Defense Analyses
Alexandria, VA
fletcher@ida.org

William D. Casebeer
Lockheed Martin Advanced Technology Laboratories
Cherry Hill, NJ
William.d.casebeer@lmco.com

ABSTRACT

Information Technology is as vital and ubiquitous in the national economy as it is in military operations. A 16-week Digital Tutor for the Navy Information Systems Technicians (IT) rating was produced for DARPA as a way to accelerate the development of expertise in IT and related areas. Graduates from the DARPA Tutor program were found by independent, third-party assessment to exceed significantly ($p < 0.01$) and practically (effect sizes ranging as high as 3.00 standard deviations) the knowledge and practical troubleshooting skills of IT graduates with 35 weeks of classroom training and sailors averaging over 9 years of IT experience in the Fleet. The Tutor was revised and extended to 18 weeks in order to include preparation for higher-level IT certifications and topics relevant to civilian employment. Ninety-seven veterans have completed the newer 18-week version. Eighty-three of these veterans were unemployed after an average 5 years of separation from active duty. There were no academic dropouts from the course. Nearly all who sought employment after finishing the course have found jobs averaging \$65,000 per year. This paper briefly reviews learning and economic findings from the Navy assessments, and provides up-to-date information on learning, quality of life, and economic findings from the veterans' project, including returns to government investment at various scales of implementation. It also provides an overview of strategies underlying intelligent tutoring systems, those used to develop the Digital Tutor, and practical tactics the Tutor applies to accelerate acquisition of technical expertise. This paper is important to the community because it addresses an important topic (Information Technology); the promise of the Tutor's technology; the perennial military need to accelerate development of technical expertise; the magnitude of the assessment findings; and the responsibility to prepare people likely to be separated from the military in the near future.

ABOUT THE AUTHORS

Dr. J. D. Fletcher is a research staff member at the Institute for Defense Analyses where he specializes in personnel and human performance issues. His graduate degrees are in computer science and educational psychology from Stanford University. He has held university positions in psychology, computer science, and systems engineering and government positions in Navy and Army Service Laboratories, the Defense Advanced Research Projects Agency, and the White House Office of Science and Technology Policy. He has served on various science and technology advisory panels for the Department of Defense, National Academy of Sciences, and the National Science Foundation. He has designed computer-based instruction programs used in public schools and training devices used in military training. He is a Fellow of the American Educational Research Association and three divisions of the American Psychological Association. His research interests include intelligent tutoring systems, synthetic environments for education and training, mobile performance aids, analyses of skilled behavior, and cost-effectiveness analyses of education and training.

Dr. William D. Casebeer is Research Area Manager, Human Systems Optimization, Lockheed Martin Advanced Technology Laboratories. Until recently he was a Program Manager in the Defense Sciences Office at the Defense Advanced Research Projects Agency (DARPA). His research interests include the neurobiology of learning and affiliated technology development, social cognitive neuroscience, neuroethics, the evolution of morality, the intersections of cognitive science and national security policy and philosophy of mind. He is author of "Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition" (MIT Press), co-author of "Warlords Rising: Confronting Violent Non-State Actors" (Lexington Books), and has published on topics ranging from the morality of torture interrogation to the rhetoric of evil in international relations in venues such as *Nature Reviews Neuroscience, Biology and Philosophy*, and *International Studies*. Earlier, Dr. Casebeer was the Deputy Head of the Joint Warfare Analysis Center's Technology Advancement Department. A retired Air Force Lieutenant Colonel and intelligence analyst, his most recent intelligence assignment was as the Chief of Eurasian Intelligence Analyst, NATO Military Headquarters.

A Digital Tutor for Accelerating Technical Expertise

J. D. Fletcher

Institute for Defense Analyses

Alexandria, VA

fletcher@ida.org

William D. Casebeer

Defense Advanced Research Projects Agency

Arlington, VA

William.Casebeer@darpa.mil

The value of technical expertise is as evident from empirical research (Ericsson, Charness, Feltovich, & Hoffman, 2006) as it is from casual observation. However, the many years of experience and practice typically needed to develop expertise come at high cost and limit its supply. Empirical demonstrations showing that the time to develop technical expertise can be compressed from years into months are few, but extant. These demonstrations typically use computer-based digital tutoring.

For instance, the Air Force Sherlock project (Lesgold, Lajoie, Bunzo, & Eggan, 1988) was a guided practice environment that prepared technicians to troubleshoot complex problems occurring in a test stand used to troubleshoot components of Air Force avionics systems. Assessments found that 20-25 hours of Sherlock training produced about the same improvement in performing difficult and rarely occurring diagnostic tasks as four years of on-job experience.

Additional evidence was provided by IMAT, the Navy's Interactive Multi-Sensor Analysis Trainer (Wetzel-Smith and Wulfeck, 2010). An at-sea trial found that two days of training with a laptop version of IMAT increased submarine effective search area by a factor of 10.5 (Chatham and Braddock, 2001). In effect, a submarine with IMAT-trained sonar operators could provide the sonar surveillance of ten submarines lacking the IMAT training.

Another example of accelerated technical expertise is the DARPA Digital Tutor, which is the topic of this report. The Tutor provides ab initio (from the beginning) training in information systems technology (IT). The training was intended to develop high-level technical competencies that are characteristic of expertise.

EXPERTISE IN INFORMATION TECHNOLOGY

IT problems can occur with local workstations, connections to a network, the network itself, receiving workstations, security and administration practices, all software involved, user error or intentional misbehavior, and even the actions of IT technicians. The background that IT technicians must master ranges from fundamentals of computer and network hardware, through ever-changing operating system, network, and application software, to a multitude of administrative factors concerning user services, permissions, groupings, and sub-groupings.

IT problem solving typically presents technicians with too much data, too many options, and unknown levels of risk -- characteristics that call for expert levels of competency (Mayer & Wittrock, 1996). Clark & Wittrock (2000) described the ability to solve such problems as the X-ray vision of experts -- an ability to see through surface symptoms and focus on deeper, applicable concepts.

Research by Feltovich, Prietula, & Ericsson (2006), found that experts integrate specific subject matter content with multiple layers of functional, abstracted representations that they deal with to solve complex problems. Sternberg and Hedlund (2002) describe these representations as tacit knowledge -- the latent knowledge acquired by experts from years of experience that they can only partially articulate but that enables them to solve unusually complex, real-world problems. Individuals who strive to achieve expertise consciously abstract principles from the experiences that come their way and deliberately seek out novel situations and challenging problems as they become available. A key to accelerating expertise is to provide opportunities for authentic, deliberate practice that compress years of relevant experience, which ordinarily occur only through happenstance, into weeks or, at most, months.

EXPERTISE AND TRAINING TECHNOLOGY

Over the last 40 years, researchers have attempted to develop computer systems with sufficient ‘intelligence’ to tailor tutorial interactions to individual needs, devise problem solution paths, and abstract specific issues and problem features into the conceptual understandings learners need to attain persistent and transferable levels of competence. Carbonell (1970) identified two basic features that distinguish this instructional approach from other forms of computer-assisted drill and practice: Use of **information structures** rather than pre-programmed, frame-oriented exercises; and Inclusion of mixed-initiative **tutorial interactions** that allow either the computer-tutor or the learner to initiate open-ended questions during an instructional dialogue.

Fundamentally, these capabilities allow computers themselves to assume much of the ‘authoring’ needed for computers to provide instruction. They were a major, early incentive for the application of machine intelligence to education and training (Fletcher, 2009; Fletcher & Rockway, 1986).

THE DIGITAL TUTOR

The Digital Tutor research effort was initiated by the Defense Advanced Research Projects Agency (DARPA) in 2005 and in cooperation with the United States Navy. Development of the Digital Tutor was to serve two broad purposes —meeting operational needs of the Navy and advancing the technology of computers used in instruction. The Tutor was to provide a 16-week course of technical training that would graduate Navy ITs with the knowledge and skills of those with expertise provided by years of Fleet experience. The 16-week duration was chosen to match that of existing qualification training for the Navy’s IT rating, but its instructional objectives substantially exceeded the journeyman level targeted by the Navy training.

The Tutor established a problem-based learning environment for each learner that was tied to information structures functionally similar to those of humans whose expertise in both the subject matter and one-on-one instruction were to be incorporated in a digital tutor.

WHY INFORMATION TECHNOLOGY?

DARPA identified about 40 technical domains as targets for investment. Navy training for the IT rating was among the most prominent of these. Navy materiel systems (e.g., ships, aircraft, submarines) are increasingly dependent upon IT systems. If these IT systems fail, naval operational capability suffers, sometimes catastrophically.

Each year about 2000 sailors, newly graduated from recruit training, attend an “A” school course to qualify for the Navy’s IT rating and various IT certifications. Nonetheless, about 5,000 trouble tickets that cannot be solved by uniformed IT technicians are referred each year to shore-based civilian experts, producing a growing operational demand in the Fleet for IT problem solving expertise.

IDENTIFYING LEARNING OBJECTIVES

The design of the IT Digital Tutor was notable for the extent of its analysis to identify the knowledge, skills, and standards required for expert IT performance. This analysis included numerous observations and interviews of IT personnel in Fleet assignments who were singled out by their managers, peers, and subordinates for IT expertise. Learning objectives and activities were then focused on the key experiences, problems, and insights that contributed to the development of this expertise.

IDENTIFYING HUMAN TUTORS

A second, basic strategy used to design and develop the Tutor’s instructional procedures was to collect, observe, and replicate the best practices of highly skilled human tutors. The objectives and standards suggested by the analysis of IT expertise guided the search for these tutors.

Experts were identified by their peer-acknowledged expertise, significant contributions to IT technology, and authoritative publications covering one or more of the IT sub-domains identified as course objectives. These experts

were then examined (auditioned) for their ability to tutor learners in one-on-one settings, which has been found to differ appreciably from one-on-many classroom settings (e.g., Graesser, D’Mello, & Cade, 2011). Twenty-four individuals were chosen by this process to provide (human) tutorial instruction in their IT sub-domain for capture and replication by computer.

REDUCING COSTS VERSUS INCREASING EFFECTIVENESS

Learners were held to a scheduled time (about a week) devoted to each of its segments. Fast-paced learners who reached targeted levels of learning early were given more difficult problems, problems that dealt with related topics that were not otherwise presented in the time available, problems calling for higher levels of conception and abstraction, and/or especially difficult ‘challenge problems’. The basic approach was to hold time (costs) constant while allowing each learner to achieve all they could in the time available – thereby maximizing learning (effectiveness). It differed from self-paced approaches that attempt to minimize time while holding learning to targeted levels.

TUTOR DESIGN AND DEVELOPMENT

The design of the Tutor emphasizes:

- Active, constant interaction with learners.
- Problem solving in authentic environments.
- Constant, ‘stealthy’, and diagnostic assessment of each learner’s progress.
- Frequent insistence on reflection and explanations of what went well and what did not.
- Focus on higher order concepts underlying problem solving processes and solutions.
- Integration with human mentors.

The Tutor’s information structures enable it to generate evolving models of the learner, generate, adapt, and assign problems to maximize an individual’s learning progress, engage in tutorial exchanges that shadow, assess, and guide the paths learners devise to solve problems, and deepen the learner’s understanding of the issues and concepts embodied within the problems. The Tutor assumed full responsibility for the teaching-learning process, but these structures were capable of sensing occasions when human mentoring and intervention were needed. This mentoring was an infrequent but essential supplement to the Tutor’s dialog.

Development of the Tutor was fundamentally eclectic and pragmatic, based on an iterative, formative evaluation approach – building the Tutor segment by segment, testing each empirically, and then revising until a satisfactory level of effectiveness was achieved. The resulting Tutor provides guided, authentic, diagnostic, individualized, and practical problem solving experience with Navy IT systems, workstations, networks, and administrative policies.

Mirroring its iterative development strategy, the Tutor’s instructional approach is spiral. It presents conceptual material that is immediately followed and applied in solving problems intended to be comprehensive and authentic. Learners interact directly with Fleet IT systems programmed to communicate and share data with each other and the instructional process, which allows the Tutor to observe, track, and model learner progress and solution paths. Once the learner demonstrates sufficient understanding of the conceptual material by applying it successfully, the Tutor advances either vertically, to the next higher level of conceptual abstraction, or horizontally, to new but related conceptual areas, which are in turn followed by authentic problem solving exercises.

The Tutor observes the following prescriptive procedures:

- Per Yerkes/Dodson (1908) through Vygotsky (1978) and thereafter, pose problems that are neither too hard nor too easy, but selected to maximize each learner’s progress in acquiring IT knowledge and skill.
- Frequently question right answers. Promote reflection by asking about antecedents, explanations, consequences, and/or implications.
- Probe vague and/or incomplete responses by the learner.
- Never articulate a misconception.
- Never provide the correct answer.
- Never provide a direct hint. In the case of a learner impasse, review knowledge and skills the learner has acquired and ask for explanations and reasons why something did or did not happen.

- Require logical, causal, and/or goal-oriented reasoning in reviewing or querying steps taken by the learner to solve problems.
- If the learner's responses suggest absent or misunderstood concepts, refocus the dialogue on these concepts before the learner proceeds.
- If the learner fails to apply or misapplies a concept for which he or she has earlier demonstrated mastery, allow problem solving to continue until the error becomes obvious to the learner.
- Always verify learner understanding of any didactic material presented by the tutor.

In brief, design and development of the Tutor's instructional approach was basically a matter of using known, but high-quality ingredients and applying them in the right proportions as determined by systematic trial and error.

NAVY ASSESSMENT

As described in a data report (Fletcher & Morrison, 2012), the Tutor was assessed five times during its development – initially by the Navy four more times by the Institute for Defense Analyses (IDA) as an independent third party. The fifth assessment for the Navy provided summative evaluation of the completed Tutor. It was conducted in late March and early April 2012.

Thirty-six participants were examined in two, separate 5-day sessions. Eighteen participants (six from each of three groups) were assessed in each session. The groups were:

- 12 graduates of the 16-week Digital Tutor (DT) course. Six individuals were drawn at random from each of two 20-student classes that completed the DT instruction on 2 March 2012 and 19 March 2012.
- 12 graduates of the 35-week ITTC (Information Technology Training Continuum) classroom course, which finished about the same time. These were the first 12 in a 30-student class who passed a certification exam.
- 12 Fleet ITs with more than 4 years (average of 9.1 years) experience as Fleet ITs. They were chosen to be representative of highly capable Fleet ITs at this level of experience.

ASSESSMENT OBJECTIVES

Among other matters, the Navy assessment examined (a) how well the Tutor provided its graduates with relevant Fleet-required IT competencies; (b) how the competencies acquired by Tutor graduates compared with those of experienced Fleet ITs; (c) how the competencies acquired by Tutor graduates compared with those of graduates who had received classroom training; and (d) whether and to what degree the Tutor accelerated acquisition of expertise. Items (a) and (b) primarily concerned preparing individuals for Navy IT duties; items (c) and (d) concerned advances in training technology.

Both ITTC and DT training content covered the general topics of IT. These topics include hardware, servers, routers, Internet Protocol, Windows Operating Systems, the Open Systems Interconnection model, Domain Name System, Dynamic Host Configuration Protocol, Exchange Server, Group and Sub-Group Policy, and so forth. A Navy review (Mitchell, 2012) of enabling objectives targeted by the 35-week ITTC and the 16-week DT training found that the Tutor included fewer Navy-identified enabling objectives than ITTC, but that it dealt with them in greater cognitive depth, requiring Bloom's Level 5 (using old concepts to create new ideas) and Level 6 (comparing, contrasting, and evaluating concepts, ideas, and theories) in contrast to Level 1 (Recall of information, facts, and nomenclature), which received more attention in ITTC training.

In assessing different instructional approaches, content similarly may be less important in training than in education. Differences between education and training are not absolute. They appear at opposite ends of a common continuum that might be called instruction. Pure education seems as rare as pure training. Assessment in education focuses on content knowledge obtained because it prepares individuals for a lifetime of varied and unanticipated application. In training, content similarity matters less because assessment must focus on relatively better-known and anticipated job-relevant performance, which is central and can be directly measured for IT and other technical areas. Knowledge still matters to help ensure transfer and retention of performance in training, and the Navy assessment examined it as well as performance. But content aside, the core product for training using the Digital Tutor was job-related performance and skill.

SUPPORT TEAMS

A White Team made up of six senior Navy ITs and three members of the Navy's Fleet Support Engineering Team (FSET) performed essential services in conducting the Navy assessment. White Team members interviewed participants in Review Boards, provided oversight of participant teams for the practical exercises, ensured that exercise parameters and procedures were observed, scored all performance in the practical exercises, and coordinated assessment activities with the Technical Support team. The Technical Support team was responsible for injecting practical problems into the assessment and for proper initialization, management, and operation of the necessary hardware and software. There were no technical disruptions during the Navy assessment.

FACILITIES

Participants were tested in three separate classrooms provided by the San Diego Naval Base. Two 3-member teams from each group worked in each classroom. Each classroom contained three IT systems—one physical system, with a full complement of servers and software, and two identical virtual systems running on hardware hosted on a single server supporting the same software as the physical systems. One system could be set up for the next exercise while the two teams worked with the remaining two systems. The systems were designed to mirror those typically found on Navy vessels and installations. Classrooms were instrumented with video cameras and microphones. Participant activity was available live and time stamped for later review and analyses using tools prepared for the program.

SCHEDULE

The assessment was performed in two successive weeks with 18 cohorts assigned at random from the three participating groups (DT, ITTC, and Fleet) for each week. The assessment for each week consisted of five activities: a Written Knowledge Test, which was administered before the week began; interviews with a Review Board conducted throughout the first day; two and a half days of Troubleshooting, a half day Security Exercise, and System Design and Development on the last day. Results from the Troubleshooting and Knowledge tests are reported here. Descriptions of all tests and their results are provided by Fletcher and Morrison (2012).

MEASUREMENT

Troubleshooting

Troubleshooting problems were drawn from a database of 20,000 trouble tickets that had been referred from the Fleet to the onshore FSET team for assistance. They provided the core assessment for the Tutor. Differences in team performance were based on group averages calculated from four data points (four teams for each group). Six teams were tested in each week's session. Troubleshooting problems were presented as they are at Navy duty stations -- as Trouble Tickets.

On each day, the same problems were presented to each group in the same order. Fifteen minutes after a problem was presented, the team working on it was free to move to the next problem whenever they believed they had solved the problem or had decided they could not. Team members used their own notes along with IT reference materials on compact discs just as they would in Navy duty station assignments. Scoring was determined by consensus among the three White Team members who awarded 0-5 points for each problem. These points were assigned as shown in Table 1.

The assessment addressed three issues: quality of problem solution, unnecessary steps taken while solving the problem, and harmful changes made during troubleshooting and were left in the system after exiting the problem. Solution quality was determined by consensus among the three White Team examiners.

Table 1. Troubleshooting problem scoring

Score	Description
5	Solved as described in the instructions or deemed equal in quality.
4	Solved, but omitted items such as documentation or full implementation.
3	Weak solution with explanation (e.g., work-around that requires later upgrading).
2	Solution that relieves the symptom but not the underlying problem.
1	Solution does not solve the problem.
0	No attempt.

A problem was considered correctly solved if its solution was rated four or five in accord with the descriptions listed in Table 1. Problem difficulty ranged from 1 (least severe) to 5 (most severe) as shown in Table 2.

Unnecessary steps provide a direct measure of efficiency in problem solving. Their cost is monetary but they can also be operationally critical during naval maneuvers and engagements at sea. Unnecessary step scores were tallies of these steps, as judged by the White Team, taken during a solution attempt. The maximum score for unnecessary steps for a single problem was held to 5, even if more than five unnecessary steps were taken – problem approaches described by the Navy as “Easter eggging” are not uncommon.

Harmful changes are especially pernicious. About 20 percent of Fleet trouble tickets may be attributed to efforts by IT technicians to solve other problems. The White Team tallied harmful changes left during troubleshooting and scored their severity depending on the skill level needed to find and correct them. The harmfulness of these changes was scored as suggested by the difficulty of removing them (Table 2, again).

Knowledge Test

The IT Knowledge Test consisted of 3 parts totaling 272 one- and two-point items worth 349 points in all. Most test items were short-answer questions, but some multiple choice and paragraph length items were also included. IDA professional and technical staff and the Navy Network Warfare Command vetted all items for their central relevance to Navy IT duty assignments.

Table 2. Problem difficulty ratings

Level of Difficulty	Description
1-Very Easy	Solved by the average “Power User.”
2-Easy	Solved by the average IT technician.
3-Average	Solved by an average network administrator.
4-Hard	Solved only by experienced network administrators.
5-Very Hard	Solved only by seasoned IT professionals.

All participants completed the Knowledge Test before beginning other elements of the Navy Assessment. All three parts were administered under “closed-book, closed-notes” conditions. The test was intended to be sufficiently difficult to avoid ceiling or floor effects. Participants were given 75 minutes to finish Part 1, 75 minutes to finish Part 2, and 90 minutes to finish Part 3. Nearly all participants finished each part in less than an hour. All finished the test in the time available.

RESULTS

Troubleshooting

Because the trouble tickets were drawn from problems that were reported from the Fleet to the shore for assistance, preliminary determination of problem difficulty could be estimated from e-mail traffic. Data reported here are based on efforts to solve 140 troubleshooting problems, the maximum attempted in the exercise.

Successful Solutions to Troubleshooting Problems

A success was defined as receiving a score of 4 or 5 on the problem. Figure 1 shows troubleshooting problems attempted and solved. The figure arranges problems by difficulty. DT teams attempted a total of 140 problems and successfully solved 104 of them (74 percent) with an average score of 3.78 (1.91). Fleet teams attempted 101 problems and successfully solved 52 (51 percent) of them with an average score of 2.00 (2.26). ITTC teams attempted 87 problems and successfully solved 33 (38 percent) of them with an average score of 1.41 (2.09). One-way analysis of variance found significance for the group means -- $F(2,417) = 41.23$, $p < .0001$. All pairwise differences are statistically significant. The DT teams outscored both the Fleet and ITTC teams with effect sizes¹ of 0.74 (large) and 1.03 (very large) respectively. Overall, Figure 1 shows that DT teams attempted and solved more troubleshooting problems with a higher probability of success than either Fleet or ITTC teams, and they attempted and correctly solved 60 percent of the very difficult problems, none of which were solved by the other two teams.

¹ All effect sizes reported here were calculated as Hedges “g” (Hedges, 1981).

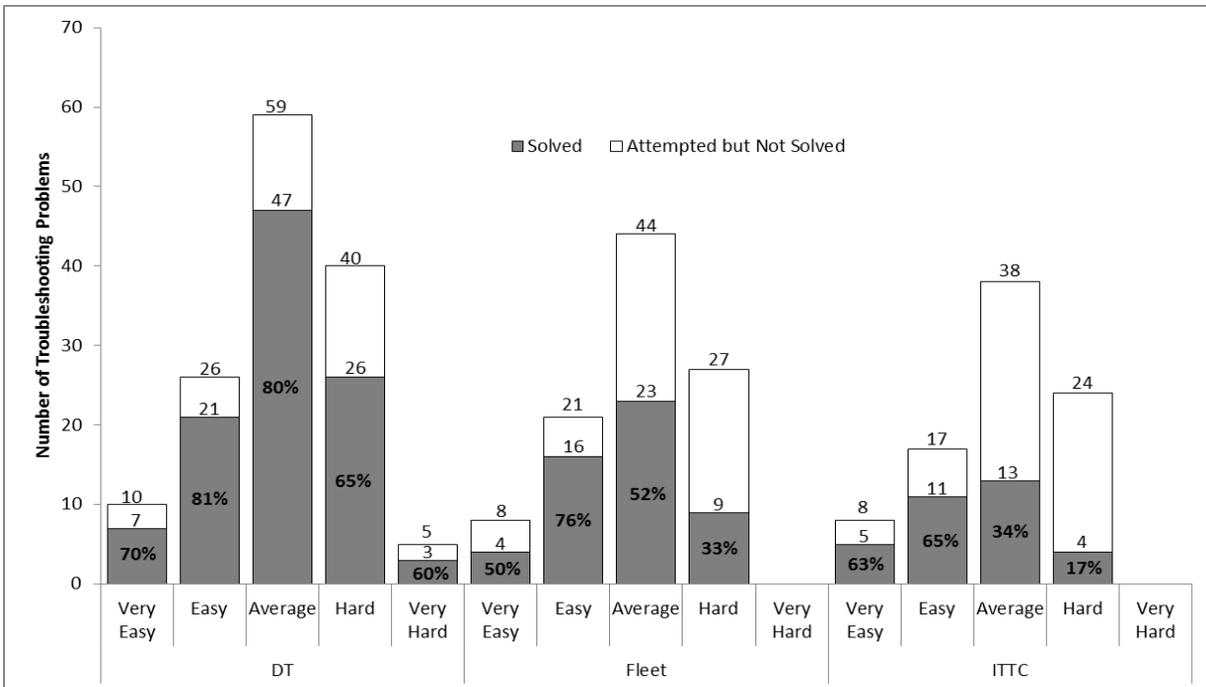


Figure 1. Troubleshooting problems solved by DT, Fleet, and ITTC teams

Unnecessary Solution Steps

The proportion of solution attempts with at least one unnecessary step divided by total attempts was 0.26 for the DT teams, which was about half that for Fleet and ITTC teams (0.51 and 0.55, respectively). Analysis of variance found the difference between groups to be statistically significant – $F(2,325) = 13.54$, $p < 0.001$. The DT proportion of unnecessary steps was significantly lower ($p < .0001$) than those for either Fleet or ITTC teams with effect sizes of -0.46 and -0.54, respectively. Frequency of unnecessary steps taken by the Fleet and ITTC teams did not differ significantly.

Harmful Errors

The proportion of harmful errors left behind during problem-solving by a DT team was less than half that of either the Fleet or ITTC teams: 0.14 versus 0.41 and 0.33 respectively. Analysis of variance found the difference between groups to be statistically significant – $F(2,325) = 11.95$, $p < 0.0001$. Harmful error rates in DT teams were statistically significantly lower than those for either Fleet or ITTC teams with effect sizes of -0.55 and -0.40 respectively. Frequency of harmful errors taken by the Fleet and ITTC teams did not differ significantly.

If DT graduates had simply matched Fleet IT performance in troubleshooting, the goals of the program to accelerate acquisition of expertise would have been met. It is notable that the DT students generally outscored the Fleet participants by substantial margins. From a training and monetary standpoint it is notable that they outscored ITTC graduates who spent more than twice the time in training as DT students.

KNOWLEDGE TEST

The averages (and standard deviations) for total scores on the Knowledge Test were 237.92 (27.95) for DT graduates, 104.79 (41.09) for Fleet ITs, and 133.25 (20.08) for ITTC graduates. ANOVA found that the overall difference between groups was significant: $F(2,33) = 61.59$, $p < 0.001$. All pairwise differences were statistically significant. Effect sizes were 3.66 for DT scores compared to the Fleet scores and 4.15 for DT scores compared to ITTC scores. ITTC graduates outscored Fleet participants with an effect size of 0.85. Even though Knowledge is not the core outcome in training, it is not inconsequential. Scores on the Knowledge Test accounted for about 40 percent ($r = 0.64$) of troubleshooting variance among DT students.

The Fleet ITs were likely to be at a disadvantage on the Knowledge test. The test assesses up-to-date knowledge at considerable breadth and depth. Although the Fleet ITs attend follow-on “C” schools and receive sustainment training and technical updates, this training is likely to have varied in content, quality, and recency, and was less current than training received by both the DT and ITTC students. Also Fleet duties may limit the range, if not the currency, of their experience. These explanations do not apply to the ITTC graduates who had just completed 35 weeks of IT qualification training combined with training for advanced IT certificates.

ADDITIONAL ANALYSES

Several additional analyses were performed in this research. Among them was an effort to determine the effectiveness of this digital tutoring compared to human tutoring, its dependence on reading ability, and its return on investment.

Digital tutoring versus human tutoring

Early work reported by Meehl (1954), Dawes (1971), and others compared the effectiveness of ‘clinical’ with ‘statistical’ decision-making. In these comparisons, regression techniques captured as accurately as possible the processes used by humans to determine such matters as graduate school admissions and medical treatments. Surprisingly, decisions based on statistical processes were found to be superior to those made by the humans on whose processes the statistical processes were based. The algorithmic character of regression analysis suggests the possibility that digital tutoring could be superior to human tutoring, even when the digital tutoring simply clones the effective tutoring practices of humans.

Such a comparison was to some degree possible in this study because the design of the Digital Tutor was an attempt to precisely capture and clone the processes first used by expert human tutors to present the course. Further, Parts 1 and 2 of the Knowledge Test were precisely the same and administered under the same conditions in both the human and tutored version of the course. They provided an approximate opportunity for comparing the knowledge obtained from human and digital tutoring. Similar analysis was not possible for troubleshooting because of differences in test procedures and conditions.

Knowledge Test comparison found that the overall mean score was higher for digital tutoring -- 156.7 (25.2) -- than for human tutoring -- 143.0 (16.9) -- but not statistically significant, $t(22) = 1.56, p < 0.10$. Further analysis found that the digital tutor condition was statistically superior for Part 1 of the test -- 96.3(13.0) compared to 80.5 (11.0), $t(22) = 3.23, p < 0.01$ -- but not Part 2 -- 60.3 (13.7) compared to 62.5 (8.3), $t(22) = 0.47, p < 0.60$. It may be, as the earlier clinical versus statistical studies found, that quality control of learning was more consistent, comprehensive, and/or accurate with digital tutoring than human tutoring. These findings are by no means conclusive, but they suggest the desirability of further research on this issue.

Dependence on reading ability

Because most instruction in the Digital Tutor is conducted through reading, the extent to which reading ability affects students’ progress was of interest. The Gates-MacGinitie reading test (GMRT) was used to assess this possibility and administered prior to assessments of instructional treatments.

This assessment found that no GMRT scores accounted for more than 11 percent of DT Knowledge Test scores, but GMRT Total Reading and Reading Vocabulary accounted for 25 and 35 percent, respectively, of variance on Knowledge Test scores for the ITTC students. These results suggest that superior IT knowledge rather than reading ability was more responsible for the higher Knowledge Test scores ($g = 2.93$) of DT participants.

Return on Investment

The DARPA Digital Tutor project was a \$35.5M research and development effort. This cost that may well diminish with the development of tools and routinized production techniques, but it is still large relative to other instructional techniques. A study by Cohn and Fletcher (2010) compared the costs to develop, implement, and employ the Digital Tutor with the costs of sending less prepared ITs to the Fleet for the years of experience and on-the-job (OJT) training aimed to produce in 16 weeks expertise at a 7-year level of Fleet IT experience. Cohn and Fletcher considered a 12-year cycle, which assumed 5 years of Tutor development costs, instant overnight replacement of the current “A” school training at the beginning of the sixth year, and 7 more years to bring all ITs in the current

training pipeline up to 7-year expertise levels. Monetary savings for the Fleet exceeded \$1.3 B over this 12-year period.

Summary

The results reported here suggest the following conclusions: (a) The Digital Tutor provided its graduates with substantial relevant IT knowledge and skill. Its graduates were found to be technically prepared to perform a wide variety of advanced IT duties in Navy assignments; (b) The knowledge and skills acquired by Tutor graduates substantially exceeded those of ITTC graduates who had spent more than twice the time in training as DT graduates and those of ITs with years of experience in the Fleet; (c) Digital tutoring can substantially accelerate the acquisition of expertise.

VETERANS PROJECT

Preparing veterans for civilian employment is widely noted as national challenge. The percentage of veterans who are unemployed after 5-10 years of service is currently about 9 percent for Gulf War Era Veterans (Collins, et al., 2014). The Department of Veterans Affairs (DVA) has done much to provide access to education and training to improve this situation. However, preparing veterans for high-demand, rewarding occupations is generally time-consuming and often limited to preparation for entry-level positions. Finding ways to accelerate development of genuine expertise, well beyond entry-level or even journeyman competencies, is a serious and growing challenge, not just for veterans, but generally, for all sectors of the economy (Ericsson, Charness, Feltovich, & Hoffman, 2006). The challenge is especially severe in the ab initio training needed for veterans. The DVA sought possibilities for meeting this challenge in adapting the DARPA's Digital Tutor for the civilian IT workplace and initiating a pilot program in the Fall of 2013 intended to train 100 Gulf War Era veterans in 2014.

Accordingly, the Tutor was modified for civilian IT occupations as an 18-week IT training course and given to 5 cohorts of 20 veterans each. Some corroboration of the Navy Assessment findings is being obtained through this program. Because there was no comparison group for this effort, corroboration of its value must be obtained from its graduates' success in finding remunerative employment.

Job fairs, Facebook, recruit email, and Twitter were used with minor results to recruit Veterans for this study. Most Veterans applied in response to advice from Department of Veterans Affairs (VA) counselors, the VA web site, and (surprisingly) Craig's List. Initial registration was received from 2408 veterans and of those, 837 completed the application process, which was itself a test of interest, grit, and determination and required submission of Armed Services Vocational Battery (ASVAB) scores. Of the final 837 applicants, about 200 met acceptance criteria to a sufficient degree: recent (post-9/11) service; honorable discharge; a score of 222 or higher on the Army and Navy's qualification test for IT training (calculated from ASVAB scores); limited or no prior civilian or military IT training; and either under-employment or no civilian employment. Table 3 provides data on the characteristics of the veterans chosen.

Of the 100 veterans who began the course, 97 successfully completed it. There were no academic dropouts. Two veterans left for medical reasons and one left because of a death in the immediate family. Eighty-three of the veterans were unemployed before beginning the course, 89 lacked full-time employment. Forty-seven possessed only high school or General Education Development (GED) degrees, 11 had associate degrees. Of the 53 veterans with associate or bachelor degrees, 4 had formal computer or IT training. Of the 47 with only high school or GED degrees, 5 received military training in electronics or information technology.

Two cohorts of 20 veterans began training in February 2013 and finished in June 2013. The remaining 3 cohorts began training in July and August 2013 and completed the course in October and November. The veterans' accomplishments and the Tutor's capabilities are indicated the enthusiasm of major commercial enterprises in hiring course graduates – generally at levels reserved for information systems technicians with 5-10 years experience and

Table 3. Veterans chosen for training.

Total Selected	100
Total Graduated	97
Average Years of Age	30.4
Civilian Computer Ed	4
Military IT Training	5
Average years of Service	5.2
Average years separation	6.5
Married	30
HS/GED Degree Only	47
Associate Degree	9
Bachelor Degree	44
Employed	17
Employed full time	11
Honorably Discharged	100

with offers averaging \$65,000 per annum. These results support the Navy's findings, as well as the conclusion that substantial acceleration in the acquisition of technical expertise is possible and within the grasp of our training technology.

Return to the government from investment in this training, including consideration of costs for research and development, was assessed for the Veterans project in terms of taxes on earnings paid over a 20-year period by DT graduates. Basic assumptions for this analysis were (1) Annual salary increases for no IT training: 4% for years 1-10, 2% thereafter; (2) Annual salary increases for others: 8% years 1-10, 4% thereafter; (3) Gross income set aside for deductions: 15%; (4) Starting salaries: DT \$65K; 2-year \$30K; 4-year \$40K; 4-year tech \$45K; no IT training \$30K; (5) State tax at 5%, Federal at 2013 rates, Payroll taxes included. Results of the analysis indicated a net present value return in tax revenue of \$832,000 for DT, compared to \$298,000 from a government supported 2 year IT Associate Degree, \$284,000 from a government supported 4 year college IT degree, \$304,000 from 4 years of government supported IT technical training, and \$269,000 for no IT training. The analysis also determined that for every Veteran hired for an IT occupation, at least one additional job would be created elsewhere in the economy. These results may be trivial compared to the quality of life improvements for veterans taking the course.

FINAL WORD

Findings from the Digital Tutor assessments suggest that substantially accelerated development of technical expertise is possible and within the state of the art. Aside from considerable time and effort, there was no "secret sauce" used in the design and development of this tutor -- the secret was careful selection of known quality ingredients, prepared properly, and applied in the right proportions. At a minimum the Digital Tutor assessments show what can be done with sufficient resources applied to best effect and that this investment realizes returns of substantial value. The acceleration of learning in this work is monetarily, practically, and operationally significant, if not a breakthrough, but it requires additional corroboration, assessment, and analysis – as well as application in other subject areas. Such investment appears justified by the results reported here.

DISCLAIMER

This work was supported by The Defense Advanced Research Projects Agency (government contract D07PC75445). The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense.

REFERENCES

- Bloom, B., Engelhart, M., Furst, E., Hill, W., & Krathwohl, D. (1956) *Taxonomy of Educational Objectives, the Classification of Educational Goals, Handbook I: Cognitive domain*. New York: David McKay Company.
- Carbonell, J. R. (1970) AI in CAI: An Artificial Intelligence Approach to Computer-Assisted Instruction. *IEEE Transactions on Man-Machine Systems, 11*, 190-202.
- Chatham, R.E. & Braddock, J.V. (2001). *Training Superiority and Training Surprise*. Washington, DC: Defense Science Board, Department of Defense.
- Clark, R., & Wittrock, M. (2000). Psychological principles in training. In S. Tobias & J. D. Fletcher (Eds.) *Training and Retraining: A Handbook for Business, Industry, Government, and the Military* (pp 51-84). New York: Macmillan Reference.
- Cohn, J., & Fletcher, J.D. (2010). What is a Pound of Training Worth? Frameworks and Practical Examples for Assessing Return on Investment in Training. *Proceedings of the InterService/Industry Training, Simulation and Education Annual Conference*. Arlington, VA: National Training and Simulation Association.
- Collins, B., Dilger, R., Dortch, C., Kapp, L, Lowry, S., Perl, L. (2014). *Employment for Veterans: Trends and Programs* (R42790). Washington, DC: Congressional Research Service.
- Dawes, R.M. (1971). A case study of graduate admissions: Application of three principles of decision-making. *American Psychologist, 26*, 180-188.
- Ericsson, K.A., Charness, N., Feltovich, P.J., & Hoffman, R.R. (Eds.) (2006). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.

- Feltovich, P.J., Prietula, M.J., & Ericsson, K.A. (2006). Studies of Expertise from Psychological Perspectives, in K.A. Ericsson, N. Charness, P.J. Feltovich, & R.R. Hoffman (Eds.) *The Cambridge Handbook of Expertise and Expert Performance* (pp. 41-67). New York: Cambridge University Press.
- Fletcher, J.D. (2009). Education and Training Technology in the Military. *Science*, 323, 72-75.
- Fletcher, J.D., & Morrison, J.E. (2012). *DARPA Digital Tutor: Assessment Data* (IDA Document D-4686). Alexandria, VA: Institute for Defense Analyses.
- Fletcher, J.D., & Rockway, M.R. (1986). Computer-based training in the military. In J.A. Ellis (Ed.) *Military Contributions to Instructional Technology* (pp. 171-222). New York, NY: Praeger Publishers.
- Graesser, A., D'Mello, & Cade, W. (2011) Instruction based on tutoring. In R.E. Mayer & P.A. Alexander (Eds.) *Handbook of Research on Learning and Instruction* (pp. 408-426). New York: Routledge.
- Graesser, A.C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology*, 9, 495-522.
- Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128.
- Lesgold, A., Lajoie, S., Bunzo, M., & Egan, G. (1988). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J. Larkin, R. Chabay, & C. Scheftic (Eds.), *Computer Assisted Instruction and Intelligent Tutoring Systems: Establishing Communication and Collaboration* (pp.201-238) Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mayer, R.E., and Wittrock, M.C. (1996). Problem solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 47-62). New York: Macmillan.
- Meehl, P.E. (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of the Literature*. Minneapolis, MN: University of Minnesota Press.
- Mitchell, C. (2012). *16-Week Education Dominance Curriculum Review and Evaluation 19 July – 13 August 2010*. San Diego, CA: Center for Information Dominance Learning Site San Diego.
- Sternberg, R., & Hedlund, J. (2002). Practical intelligence, g, and work psychology. *Human Performance*, 15, 143-160.
- Vygotsky, L.S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wetzel-Smith, S.K. & Wulfeck, W.H. (2010). Training incredibly complex tasks. In J.V. Cohn & P.E. O'Connor (Eds.) *Performance Enhancement in High Risk Environments* (pp. 74 - 89). Westport, CN: Praeger.
- Yerkes, R.M., & Dodson, J. D. (1908) The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.