HISTORICAL MUSINGS ON COST/BENEFIT ANALYSIS OF MILITARY TRAINING

Stanley A. Horowitz
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, Virginia 22311-1882
Ph: 703-575-4685
E-mail: shorowit@ida.org

ABSTRACT

This paper reviews a body of literature on the cost-effective benefits of training and new technology with regard to measurable improvements in performance.

It observes that resource allocation is often not guided by these insights, but that funding decisions can be made more output-oriented—and it discusses one effort to make them so.

Finally, it encourages the training analysis community to remember that training is one of many ways to improve performance, and that military resource managers would be well served by analyses that explicitly compared them (H830, I210, D730).

I. INTRODUCTION

I have worked periodically on topics relating to the cost-effectiveness of training for almost 40 years. This paper draws on some of the observations, speculations, hopes, and disappointments that have accumulated over that time.

There is a non-negligible body of work demonstrating that training improves the performance of military personnel and units. This has been demonstrated by studying the determinants of both unit readiness and operational effectiveness.

One strain of the literature demonstrates the effectiveness of advanced training technology: simulators, simulations, networked training, distance learning, and the like. Another goes beyond addressing effectiveness to demonstrate improved efficiency. This paper reviews research in these areas.

The work of the training analysis community has not been without impact. Important training innovations have been adopted as a result of analytic insights, but it seems to me that we could progress faster. One key limiting factor on the rate of training improvements is the way in which resources are allocated to training. There is little explicit attention to assessment when budget decisions are made. This paper reviews an ongoing effort to more tightly link training resource allocation to high-level training goals.

As important as it is to understand the value of training, it is important to remember that training is but one way to improve military performance and that it is not always the most cost-effective one. The training research community should broaden its horizons by analyzing the payoffs to different ways of improving performance—training of various kinds, job performance aids, better personnel selection, and more automated equipment. This paper discusses some efforts to make such cross-domain comparisons.

II. EARLY ANALYSES OF TRAINING AND PERFORMANCE

Many studies have demonstrated the value of operational training. These studies cover such diverse areas as ship readiness, aircraft readiness, aviation performance, and Army performance.

Ship Readiness Analyses

In the mid-1970s, I examined the productivity of Navy enlisted personnel in six maintenance occupations (Horowitz and Sherman 1977). The success of a work center was measured by the fraction of time that work center was free of serious mission-degrading equipment failures. This was not a sophisticated analysis of any particular kind of training. Our key measure of training received was related to time in the Navy. A similar analysis at the ship rather than the work center level was performed by colleagues of mine a few years later (Quester et al. 1989). In both cases, the research showed that ships with a higher fraction of senior personnel, who were more experienced and presumably more knowledgeable, tended to be significantly more ready than other ships.

Aircraft Readiness Analyses

Similar work was done for carrier-based aviation squadrons (Marcus 1982). Readiness was measured by the number of sorties generated in a quarter. The analysis examined the relationships between sortie generation and the characteristics of squadron enlisted personnel. The data covered 292 quarters of squadron operation between 1977 and 1980.

The principal finding of the study was that squadrons with more senior personnel were significantly more ready. Indeed, the addition of junior personnel tended to reduce readiness, presumably because more junior people required their superiors to spend more time providing on-the-job training and less time making sure tasks were accomplished.

Aviation Performance

Moving beyond readiness-related output measures, I have worked on a series of studies containing data on aircrew performance, reflecting either expert assessments or objectively measured results (Hammon and Horowitz 1990; Horowitz et al. 1987; Hammon and Horowitz 1996; Hammon and Horowitz 1992; and Cedel and Fuchs 1986). These included kill probabilities in instrumented air-combat maneuvering exercises, bombing accuracy, airdrop accuracy, accident rates, torpedo exercise scores, carrier landing grades, and the overall results of operational readiness evaluations for carrier-based squadrons.

Different analyses focused on the experiences of different people (or groups of people)—sometimes pilots, navigators, and co-pilots for airdrops, and the entire tactical team (including sensor operators) for torpedo exercises. The studies addressed squadron, crew-level, and individual performance in a wide range of circumstances covering Navy, Marine Corps, and Air Force operations.

We consistently found that both recent and career flying hours were generally significant predictors of performance. Not surprisingly, career hours were more important. It is difficult for additional short-term training to make up for long-term experience.

Army Performance

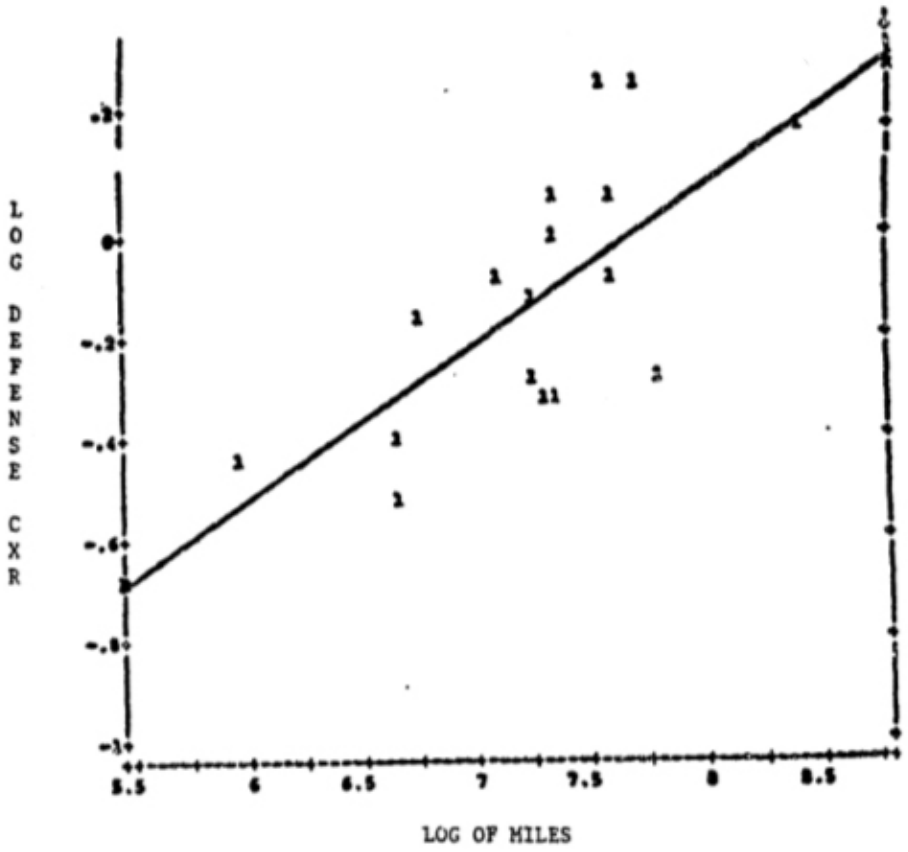
One of my favorite studies was performed by researchers at the Army Research Institute in the late 1980s, although it had too small a sample size to be truly compelling: only seven observations. I liked it because it was able to look at something approximating a real ground combat situation. It is based on the performance of brigades at the National Training Center (NTC) at Fort Irwin, California, as assessed by professional observer-controllers. Units were

ranked according to the percent of missions rated as “successfully accomplished.” The NTC has over 2500 square kilometers of maneuver space and ranges and is highly instrumented.

Statistically significant correlations were found between miles driven in train-up and performance on both force-on-force offensive missions and on live-fire defensive missions. Figure 1 shows the relationship between miles driven and the exchange ratio in defensive missions (Keesling et al. 1992; and Hiller et al. 1990).

FIGURE 1

Exchange Ratio in Live-Fire Defense vs. Miles Driven in Training



Source: Keesling et al., “The Determinants of Effective Performance of Combat Units at the National Training Center,” Army Research Institute, June 1992.

III. THE EFFECTIVENESS OF ADVANCED TRAINING TECHNOLOGY

The research findings discussed so far mostly find that more training yields better performance. Training has either explicitly been live training—for example, flying hours and vehicle miles—or has been an amalgam of all training garnered through experience. This section will address the value of various kinds of simulator-based training.

Several of the aviation studies cited earlier addressed time spent in simulators as well as flying time. They found that: time spent in C-130 simulators improves the accuracy of airdrops, both career and recent simulator time improves bombing accuracy for F/A-18s, and increased simulator time for enlisted acoustic operators improves P-3 torpedo exercise scores.

A program to provide additional simulation-based training devices, courseware, and strategies to units in the Idaho National Guard resulted in significant improvements in performance at NTC and other training venues (Metzko and Morrison 1999).

The Interactive Multisensor Analysis Trainer (IMAT) is a PC-based tool that allows a sonar operator and a submarine's tactician to visualize a very complicated acoustic situation and determine how to best use their sensors. New sonar technicians trained with IMAT were found to perform better than personnel with years of fleet experience (Fletcher 2009; Braddock and Chatham 2003; and Wulfeck and Wetzel-Smith 2007).

IV. INSIGHTS INTO COST-EFFECTIVENESS

While the effectiveness of training on performance was a major focus of the research previously discussed, studies of its cost-effectiveness have also been performed.

The 1989 ship readiness study mentioned above found that a six percentage point increase in personnel experience was associated with a nine percentage point increase in ship availability.

While there are definitely costs associated with retaining more experienced personnel, it is much less expensive to increase ship availability by buying more experience than by buying more ships.

When comparing various types of training, it appeared in some cases that an extra simulator hour may have improved bombing and airdrop accuracy more than an extra flying hour. Simulator time is obviously much less costly.

However, it is not always the case that more training or more sophisticated training is the most cost-effective way to improve performance. An analysis comparing the payoff for tank training, both live and simulated, relative to buying more sophisticated equipment yielded unclear results (Deitchman 1990). Conversely, performance gains from the IMAT training are much less expensive than similar gains from improved hardware.

This paper largely focuses on operationally oriented training rather than on the provision of basic occupational skills, which usually takes place in classrooms. Nonetheless, there is an extensive literature, summarized by my colleague Dexter Fletcher (2010), that argues persuasively that computer-aided instruction is a cost-effective alternative to traditional podium-based instruction. Some of this literature dates back to the 1960s.

The thrust of this literature is characterized by Fletcher as demonstrating the rule of thirds. Computer-aided instruction can either reduce instructional time by one-third on average or increase the skills and knowledge acquired by one-third. At the same time, it can reduce the cost of instruction by one-third. Especially today, computer-aided instructional tools can tailor themselves to the needs of individual students, approximating the effectiveness of individual tutoring.

A recent example is the Digital Tutor program developed by the Defense Advanced Research Projects Agency (DARPA) for training information technicians in the Navy. Not only do Digital Tutor students perform better than traditionally schooled students on both written and practical tests, but in some cases they outperform the people who teach the traditional courses (Fletcher 2011).

V. ADOPTION OF INNOVATIVE TRAINING

Training Has Improved

There are significant areas where military training has been improved by the adoption of new technology, and sometimes this has been facilitated by analysis.

The increased emphasis on instrumented realistic combat training characterized by NTC and air warfare training, like the Navy's Top Gun program and the Air Force's Red Flag, are widely recognized as having markedly improved the readiness of deploying units.

Perhaps analysis has combined with budget pressures to convince the Services to increase their use of flight and other combat simulators. IMAT plays an important role in the operational training of sonar operators.

Since 2004, the U.S. Defense Department has spent over $3 billion on the Training Transformation, or T2, program. T2's purpose is to improve training—particularly joint training and training with partner organizations, including foreign militaries. It has funded a program to certify that joint training provided by the Services and Combatant Commands is realistic and has appropriate joint context. It has developed an integrated federation of simulations to facilitate staff training, to facilitate synchronous training from disparate locations, and to enhance the ability to include rare, expensive assets in training activities. It has also built a widespread, easily

configurable, and even portable, network to further enable training from far-flung sites. T2 has also championed the use of computer-aided instruction for individual-level training.

Sometimes Progress Seems Slow

All that notwithstanding, from an analyst's perspective, the speed at which training improves often seems slow, and the role of analysis in enabling improvement seems less prominent than it might be.

For example, there are persistent problems that might benefit from focused investments. The U.S. training community has long lamented the difficulty of training with key players from other organizations, domestic and foreign. Training for cyber-related disruptions has proven difficult.

While some tasks that are undertrained may be difficult to simulate with needed fidelity, some deficiencies can be overcome through better simulation and better connectivity. And, as was noted earlier, despite extensive research, legacy modes of training—like traditional podium instruction—persist to a greater degree than analysis indicates is warranted.

While training is assessed by the leaders of units and their superiors, little is done to build databases to support a better understanding of what works and what doesn't, and where there are training problems and where there aren't.

Finally, I observe that the allocation of training resources is generally not based on a systematic examination of the expected impact of those resources. Why might this be so?

Impediments to Progress

Perhaps the way in which training requirements are developed could be more flexible. Current practice calls for the identification of mission-essential tasks, the conditions under which

they must be performed, and the standards they must meet. Objectives are set and training programs developed to meet them. This defines how much training is enough. By and large the objectives are achieved, but often the possibility of achieving them more cost-effectively is not routinely considered.

Elaborate processes exist for documenting training-related problems, derived from both training and operational activity. These feed lessons-learned systems. The lessons are available for interested parties to read about, but there often are no mechanisms to ensure that initiatives are developed to address them. One might think of the systems as mostly documenting lessons *observed*, because without systematic follow-up they are unlikely to really be *learned*. What little data exist on training performance at the task or mission levels that could support analysis are rarely used either to systematically identify problems or to guide investments.

Perhaps there is, to some extent, a good news story here. Training managers, who do have considerable knowledge of what is desired, are largely satisfied—but analysis might show that training could be better or more efficient.

The attention of top leadership to the specifics of training problems and opportunities is not consistent. The T2 program was set up when the Deputy Secretary of Defense, the number two person in the Department, became personally involved, but that is unusual.

There are undoubtedly other impediments to progress, but the last one I will mention is that the allocation of training resources is largely based on minor perturbations of who had what last year, not on our priorities. There is also little done to ensure that the training expenditures made last year achieved explicit goals.

VI. AN INNOVATIVE APPROACH TO RESOURCE ALLOCATION

The T2 program, which allocates roughly $600 million a year to improve training, makes a significant effort to allocate training resources more analytically. The process for allocating the funds has been as follows: Program stakeholders—the Services and Combatant Commands—submit proposals, which are reviewed collectively by the stakeholders and prioritized. The results of this review are then presented to the training leadership, which makes the final allocation decisions.

However, the process has not relied on systematic analysis. Starting this year, leadership is seeking to improve the analytic basis for its choices. A new process being adopted to support that goal will ultimately provide four kinds of information to decision makers:

- First, it will categorize proposed investments according to their functional emphasis—do they: seek to improve training by developing new training tools and techniques; support previously existing training activities with personnel, operational support, or capital investments; or manage training?
- Second, it will evaluate proposals with respect to their focus on strategically chosen focus areas. These areas reflect emphases in national security guidance documents and training priorities identified by the leadership.
- Third, it will evaluate measures of effectiveness that the sponsors of proposals must submit.
- Finally, it will track measures of effectiveness for continuing proposals to see whether the investments are accomplishing what was promised.

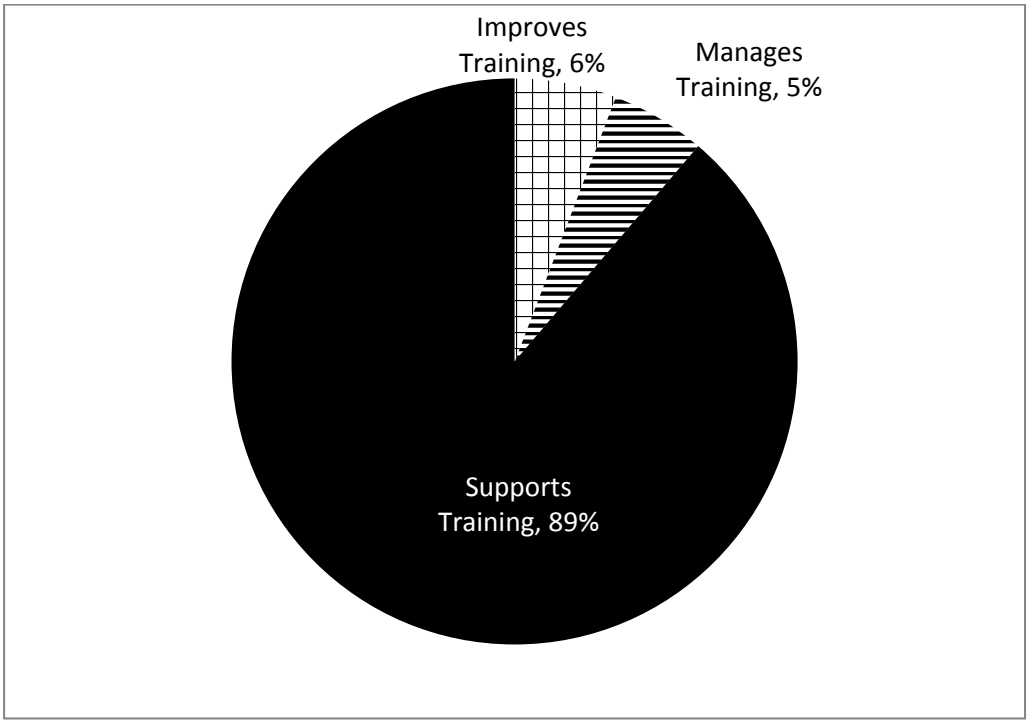
The goal is to use this information in making resource allocation decisions.

Characterization of Proposed Investments

Figure 2 presents an evaluation of last year's funding requests. It indicates that almost 90 percent of them focused on supporting training, which included some capital expenditures. Only six percent were designed to improve training tools and techniques.

FIGURE 2

Initial Categorization of T2 Funding Proposals: How is Requested $1.1 Billion for Joint Training Allocated?



Focus Areas

The focus areas shown in Table 1, which will be used to categorize proposals, have been explicitly identified by the training leadership as priorities for increased emphasis, based on high-level guidance documents. There is more detail to the structure, but it is not presented here.

This year the procedure will provide a standard format that the organizations submitting proposals must use to characterize their emphases. Personnel in the Training Resource and

Strategy Office in the Office of the Under Secretary of Defense for Personnel and Readiness will review their characterizations.

The purpose of this review is two-fold: first, to provide more insight into the goals of individual proposals, and second, at least as important, to allow the leadership to see what priorities are best and least addressed by the T2 program as a whole.

TABLE 1

Focus Areas for Joint Training Readiness

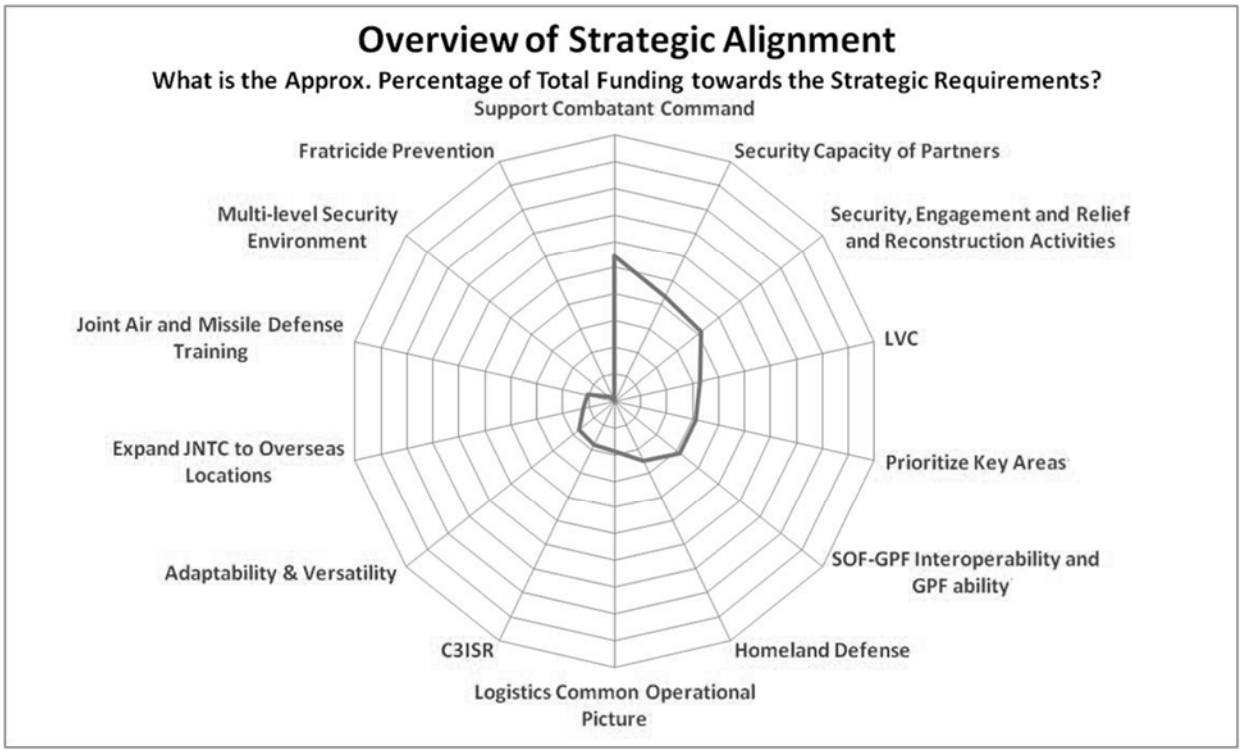
1.	Train for Irregular Warfare Threats (counterterrorism, unconventional warfare, foreign internal defense, counterinsurgency, and stability operations)
2.	Train for Operations to Deter and Defeat Aggression
3.	Enhance Integration with Partners
4.	Strengthen Security and Resilience at Home
5.	Improve Capabilities for Cyber, Space, and Information Operations
6.	Enhance Unit and Individual Adaptability
7.	Improve our Ability to Train Realistically and Efficiently

The last two focus areas are a bit different from the first five. They are not about developing and maintaining specific operational capabilities; rather, they concentrate on the people being trained and the methodology being used to train them.

Figure 3 illustrates one of the major uses of the focus areas. Last year the Training Readiness and Strategy Office did a prototype analysis of the funding proposals that had been submitted. A different set of focus areas was used in the prototype, but the chart shows that the portfolio of proposals paid far more attention to some priority areas than others. Relatively little emphasis was given to joint air and missile defense, fratricide prevention, and operating in a multi-level security environment. This kind of analysis tells the training leadership that if it is serious about the under-emphasized priority areas, it must take action to change the emphasis.

FIGURE 3

Emphasis Given to Priorities in Funding Proposals



Measures of Effectiveness

The third thrust of the effort to make the allocation of training resources more analytic is to give more attention to measures of effectiveness (MOEs) for individual projects.

In the past, MOEs have been required in the standard project submission form, but they have not received much attention. Many have not met common-sense criteria. Some could not be measured with available data, some were not clearly relevant to program goals, some did not have goals, and some were hard for non-specialists to understand. Table 2 shows the characteristics the training leadership is trying to ensure the submitted MOEs possess, including increased emphasis on relevance, practicality, usability, and even cost-effectiveness.

Perhaps the biggest flaw concerning the MOEs is that in the past nobody was required, or even encouraged, to track performance with respect to them; they were just promises with no follow-up and no consequences for failure to deliver. Starting this year, quarterly reporting will be required.

TABLE 2

Guidelines for Good Measures of Effectiveness

Content	Structure	Language
Measurable	Quantifiable	Terms clearly defined
Credible (clear cause and effect)	Has threshold	Understandable (to non-specialist)
Relevant (aligned with program goals)	Simple	
Significant (represents performance)	Not anecdotal	
Useful (provides actionable feedback)		
Timely		
Reliable (accurate)		
Attainable (data available)		
Cost-effective (not too expensive)		

VII. BROADENING OUR HORIZONS

I have touched just a drop on whether investments in training are a more cost-effective way of improving military capability than other kinds of expenditures. That is a major issue that we, the training analysis community, do not often address—and neither does anyone else.

Military performance can be improved in many ways: training, larger forces, smarter people, more sophisticated equipment, job performance aids, and more spare parts among them. Sometimes training is the most cost-effective way to make marginal expenditures and sometimes it isn't. The analytic community should compare the cost-effectiveness of investments along these many dimensions. Usually we don't.

But we can. It has been done, sometimes with surprising results.

In the 1980s, economists at West Point asked whether attention to attracting and retaining high-quality personnel was more or less important as equipment became more sophisticated. Looking at scores in tank gunnery exercises, they found that tank commanders and gunners who scored higher on mental tests made a big difference in older M60 tanks, but hardly any difference at all in newer M1s (Smith et al. 1986).

Fletcher and Johnston (2003), in a study of a computerized job performance aid to help F-16 technicians troubleshoot avionics problems, found that relatively untrained junior personnel with the job performance aid did better than highly trained avionics specialists who did not have the aid, and about as well as highly trained personnel who did have the aid.

As I said earlier, analysis of the determinants of ship readiness indicated that more training was a more cost-effective way to increase available combat power than buying more forces, although improvements to the supply system might be even more cost-effective.

Our resource allocation processes are very often stove-piped; routine comparisons across resource areas are not made, so efficiencies go unidentified. The analytic community could help remedy this weakness.

VIII. CLOSING OBSERVATIONS

There is a long-standing body of research demonstrating that training provides measurable improvements in performance, that new technology is often more effective than old technology, and that training in general and new technology in particular are often cost-effective ways to improve performance.

Unfortunately, resource allocation is often not guided by these insights, but funding decisions can be made more output-oriented—and I discussed one effort to make them so.

Finally, the training analysis community should remember that training is one of many potential ways to improve performance and that military resource managers would be well served by analyses that explicitly compared them.

REFERENCES

Braddock, Joe and Ralph E. Chatham, co-chairmen. “Defense Science Board Task Force on Training for Future Conflicts.” June 2003.

Cedel, Lt Col Thomas E., USAF and Lt Col Ronald P. Fuchs, USAF. “An Analysis of Factors Affecting Pilot Proficiency.” Washington, DC: Air Force Center of Studies and Analyses. December 1986.

Deitchman, Seymour J. “Further Explorations in Estimating the Military Value of Training.” IDA Paper P-2317. Alexandria, VA: Institute for Defense Analyses. January 1990.

Fletcher, J. D. “Training via Simulations and Games.” Presentation to the Committee for Learning Science: Computer Games, Simulations, and Education. Washington, DC. October 2009.

Fletcher, J. D. “Research Foundations for the Advanced distributed Learning Initiative.” IDA Document D-4118. Alexandria, VA: Institute for Defense Analyses. August 2010.

Fletcher J. D. “DARPA Education dominance Program: April 2010 and November 2010 Digital Tutor Assessments.” IDA Document NS D-4260. Alexandria, VA: Institute for Defense Analyses. February 2011.

Fletcher, J. D. and R. Johnston. “Effectiveness and Cost Benefits of Computer-Based Decision Aids for Equipment Maintenance.” IDA Document D-2824. Alexandria, VA: Institute for Defense Analyses. February 2003.

Hammon, Colin P. and Stanley A. Horowitz. “Flying Hours and aircrew Performance.” IDA Paper P-2379. Alexandria, VA: Institute for Defense Analyses. March 1990.

Hammon, Colin P. and Stanley A. Horowitz. “Relating Flying Hours to Aircrew Performance: Evidence for Attack and Transport Missions.” IDA Paper P-2609. Alexandria, VA: Institute for Defense Analyses. June 1992.

Hammon, Colin P. and Stanley A. Horowitz. “The Relationship between Training and Unit Performance for Naval Patrol Aircraft – Revised.” IDA Paper P-3139 (Revised). Alexandria, VA: Institute for Defense Analyses. December 1996.

Hiller, Jack H., Howard McFann, and Lawrence G. Lehowixz. “Does OPTEMPO Increase Readiness? An Objective Answer.” Army Science Conference Proceedings, Volume II. 1990.

Horowitz, Stanley A., Colin P. Hammon, and Paul R. Palmer. “Relating Flying-Hour Activity to the Performance of Aircrews.” IDA Paper P-2085. Alexandria, VA: Institute for Defense Analyses. December 1987.

Horowitz, Stanley A. and CDR Allan Sherman. “Crew Characteristics and Ship Condition.” CNA Study 1090. Alexandria, VA: Center for Naval Analyses. March 1977.

Keesling, J. Ward, Patrick Ford, Francis O’Mara, Howard McFann, and Robert Holz. “The Determinants of Effective Performance of Combat Units at the National Training Center.” Army Research Institute. June 1992.

Marcus, A. J. “Personnel Substitution and Navy Aviation Readiness.” Professional Paper 363. Alexandria, VA: Center for Naval Analyses. October 1982.

Metzko, John and John Morrison. “Assessment of SIMITAR.” IDA Paper P-3522. Alexandria, VA: Institute for Defense Analyses. October 1999.

Quester, Aline, Russell Beland, and William Mulligan. “Ship Material Readiness,” Professional Paper 467. Alexandria, VA: Center for Naval Analyses. March 1989.

Smith, D. Alton, Robert H. Baldwin, and Robert L. Phillips. “Are Smart Tankers Better? AFQT and Military Productivity.” *Armed Forces & Society*. Winter 1986.

Wulfeck, W. H. and S. K. Wetzel-Smith. “Use of Visualization Techniques to Improve High-Stakes Problem Solving.” In *Assessment of Problem Solving Using Simulations*, edited by E. Baker, J. Dickieson, W. H. Wulfeck, and H. F. O’Neil, 223–238. New York: Lawrence Erlbaum Associates, 2007.