This product is part of the RAND Corporation monograph series. RAND monographs present major research findings that address the challenges facing the public and private sectors. All RAND monographs undergo rigorous peer review to ensure high standards for research quality and objectivity.
Performance Evaluation and Army Recruiting

James N. Dertouzos, Steven Garber

Prepared for the United States Army
Approved for public release; distribution unlimited
The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2008 RAND Corporation
All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2008 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org
Performance metrics are the standard by which individuals and organizations are judged. Such measures are important to organizations because they motivate individuals and influence their choices. In the context of Army recruiting, choices made by recruiters can have a major impact on the ability of the Army to meet its goals. Designing and implementing performance metrics that support Army goals requires analysis of how different metrics would affect recruiter behavior and, in turn, recruiters’ contributions toward achieving the Army’s goals. In addition, performance measures should not be heavily influenced by random factors affecting enlistment outcomes that might be reasonably attributable to luck or fortune. The present study focuses on performance measurement for Army recruiting to provide incentives that induce behaviors that support achievement of Army goals and are acceptably insensitive to random events.

We compare and evaluate, theoretically and empirically, various performance metrics for regular (i.e., active component) Army recruiting. Some of them have been used by the United States Army Recruiting Command (USAREC); others are original to this study. Previously used performance measures—which we refer to as traditional measures—can be computed from readily available data for command or organizational units of various sizes (e.g., stations, companies, battalions) and for intervals of varying lengths. Traditional Army metrics for recruiter performance include the following:

• How many contracts were signed per on-production regular Army (OPRA) recruiter?
• How many high-quality contracts—namely, enlistments of high school seniors and high school graduates scoring in the top half (i.e., categories I to IIIA) of the Armed Forces Qualification Test (AFQT)—were signed per OPRA recruiter?
• By what percentage did the command unit exceed or fall short of recruiting missions (the Army’s version of sales targets or quotas) for high-quality or total enlistments?
• How often did the command unit achieve its recruiting targets—the regular Army mission box—which, during the period we analyze, included separate targets for (a) high-quality high school graduates, (b) high-quality high school seniors, and (c) other youth?1

Our analysis demonstrates that all these measures—and all others that can be computed with readily available data—are flawed because they fail to provide strong incentives (1) for current recruiters to put forth maximum effort or (2) for soldiers who have good skills or aptitudes for recruiting to volunteer for recruiting duty. Moreover, such metrics can be viewed as inequitable; hence, using them can undermine morale and, as a result, reduce the effort levels of recruiters.

Consider an example involving hypothetical recruiting Stations A and B. Suppose that the market territory of Station A is uncommonly fertile for recruiting youth to enlist in the Army. Recruiters in Station A are, in fact, often able to conform to the adage, “make mission, go fishin’.” In contrast, the recruiting territory for Station B is uncommonly barren for Army recruiting. Suppose further that USAREC recognizes that Station A has a much better market than does Station B and, in response, assigns to Station A recruiters missions that are double those of Station B. Suppose, finally, that Station A exactly achieves its missions, but that Station B fails to meet its mission. According to all four of the traditional measures described in the bulleted list above, Station A has outperformed Station B. This conclusion, however, is suspect. In particular, to decide which station has really performed

1 The term other refers to the total contracts minus senior high-quality contracts minus graduate high-quality contracts.
better, one must inquire: *How much* better is Station A’s market than is Station B’s?

Much of the research reported here focused on developing and empirically implementing methods aimed at measuring recruiting performance while taking adequate account of variations in the difficulty of enlisting youth (a) falling into different contract categories (such as high-aptitude seniors versus high school graduates) and (b) located in the market territories of different recruiting stations. Our previous research (Dertouzos and Garber, 2006) demonstrated that the local markets of Army recruiting stations vary significantly in the effort and skill required to enlist high-quality prospects (seniors and graduates combined) and that variations in stations’ high-quality missions do not adequately reflect these differences. Thus, performance assessment based on the traditional metrics does not accurately reflect effort exerted and skill applied. In principle, incentives for exerting effort and for persuading the right soldiers to volunteer for recruiting could be stronger.

In Chapter Two, we present a framework for estimating determinants of the numbers of enlistments in various categories that enables estimation of the difficulty of recruiting youth in different categories and in the market areas of different stations. This empirical analysis relies on a microeconomic model (detailed in Appendix A) of recruiter decisions to direct effort toward recruiting youth of different types. Extending previous analyses (Dertouzos and Garber, 2006, Chapter Four), this model emphasizes and distinguishes two general factors that determine recruiting outcomes: recruiter productivity (effort plus skill) and the quality of the organizational unit’s market area. We then present a preferred performance metric (PPM) for recruiting stations that explicitly distinguishes among multiple enlistment categories while accounting for variations in local markets that affect the difficulty of enlisting youth who fall into these separate categories. The advantages of this metric come at a cost, however. In particular, implementing them requires econometric analysis to estimate the difficulty of enlisting youth of different types in different local recruiting areas.

In Chapter Three, we present estimates of econometric models of recruiting outcomes using monthly, station-level data for fiscal
years 2001 to 2004. The main purposes of our empirical analysis are to quantify the factors that affect the difficulty of recruiting and to use this information to develop estimates of the difficulty of recruiting youth of various types in various locations. We first estimate a model distinguishing the three categories of youth that are missioned separately: high-aptitude, high school graduates; high-aptitude, high school seniors; and “other” enlistments. These estimates provide the empirical foundation for comparing alternative PPMs.

Key findings regarding determinants of enlistments for the three missioned categories of youth include the following:

- At intermediate levels of difficulty of achieving a station’s recruiting goal (mission plus Delayed Entry Program (DEP) losses charged that month)—which depend on the level of the goal, the quality of the local market, and other factors—recruiter effort increases as goal difficulty increases.
- The positive marginal effect of increasing goals on effort—and, in turn, on contracts produced—declines as goals increase.
- The marginal effect of goal increases on contract production is substantially higher for stations that have been more successful in recruiting in the recent past. Relying on research literature in psychology and management, we interpret this effect as indicating that success increases recruiters’ confidence, their morale, or both.
- Market quality is also an important determinant of recruiter effort levels.
- Market quality in a station’s territory depends on many factors, such as qualified military available (QMA) youth per OPRA recruiter, economic and demographic factors, and the strength of competition the Army faces from other services.
- Some determinants of market quality differ substantially across the three missioned categories.
- Actual missions do not adequately reflect differences in market quality. This suggests that enlistment outcomes might be improved through use of performance measures that better reflect such dif-
ferences or by setting missions to more accurately reflect market quality.

We conclude Chapter Three with an exploratory empirical analysis of four enlistment categories that are not missioned separately: (a) high-quality males, (b) other males, (c) high-quality females, and (d) other females. The capability to estimate and account for variations in market quality among these four market segments could be invaluable both for measuring performance and for improving the efficacy of missioning. The analysis uses information on categories of military occupational specialties (MOSs), including combat arms, combat support, blue-collar, and white-collar jobs.\(^2\) Key findings include the following:

- Market quality in the four dimensions varies considerably across station areas.
- Due to variations in local demographics, economic conditions, or both, the difficulty of recruiting youth in one of the four categories provides little, if any, guidance about the difficulty of recruiting youth in other categories. For example, a recruiting station’s territories may be difficult for recruiting in one category (e.g., high-quality males) while having an ample supply of prospects in another (e.g., high-quality women).
- Missioning and the distribution of MOSs available to be filled can have important effects on the volume and distribution of enlistments.
- Combat support jobs and white-collar jobs have special appeal to both high-quality men and high-quality women. Combat MOSs are most attractive to lower-quality men; blue-collar jobs draw more “other” women. This finding implies that the distribution

\(^2\) We define MOS categories as follows: (1) combat arms MOSs = all occupations that are not available to women; (2) combat support MOSs = all other jobs that have no obvious private sector counterpart, such as weapon system maintenance or missile operators; (3) blue collar MOSs = jobs with private sector counterparts that are considered blue collar, such as construction, truck maintenance, and transportation jobs; and (4) white collar MOSs = jobs with private-sector counterparts in office, service, or professional occupations, such as nurses, clerical, or accounting.
of available occupations can differentially affect the difficulty of recruiting in different local markets.

- Incremental returns associated with adding recruiters diminish. However, additional recruiters can potentially expand the market for lower-quality males and all females. This suggests that the high-quality male market is closer to saturation.

In Chapter Four, we compute three versions of the PPM at the station level for fiscal year (FY) 2004 for 1,417 stations with complete data, and we compare them to five diverse, traditional Army recruiting performance measures. Our key findings:

- The traditional performance measures are not highly correlated with any of the three alternative PPMs. More specifically, the range of rank correlations (across stations) for a full year of station-level performance is 0.42 to 0.68.
- These fairly low correlations indicate that classifying stations on the basis of their ranks on any of the five traditional measures is an unreliable guide for assessing station performance. For example, among the stations that would be ranked among the top 25 percent of performers based on frequency of making regular Army mission box, only 46 percent would be ranked in the top quarter using a PPM. Moreover, about 20 percent of those ranking in the top quarter based on mission-box success would be ranked in the bottom half using a PPM.
- In sum, performance evaluation using traditional measures can be very misleading.

In Chapter Five, we consider choice of organizational and temporal units for evaluation. First, we consider alternative performance windows or time periods over which performance is assessed (e.g., a month, quarter, or year), using estimates from our model of the three missioned enlistment categories. Key findings include the following:

---

3 Alternative versions of the PPM, based on different weighting schemes for combining contract counts in different categories, yielded similar results.
• Randomness or luck is a leading determinant of contract production during a month or a handful of months.
• Aggregating performance over months tends to average out this randomness, and thus makes performance measurement over longer time intervals considerably less subject to randomness.
• For example, during a single month, the proportions of the variation in production levels that are predictable using our estimates are only 0.32, 0.10, and 0.27 for high-aptitude graduates, high-aptitude seniors, and other contracts, respectively.
• Much of this randomness averages out using a performance window of six months, with the proportions of variance explained exceeding 0.65 for both graduates and other enlistments, but only 0.31 for seniors.
• Even over 24 months, considerable randomness remains. Senior contracts are the least predictable, with less than 0.60 of the variance explained by the model.
• Unmeasured factors that are station-specific and persist over time account for 75 percent or more of the unexplained variation that remains—even for two-year performance windows.

Next, we use data for FYs 1999–2001 to evaluate the efficacy of individual-recruiter versus station-level missioning and performance evaluation. The analysis takes advantage of a “natural experiment” made possible due to the sequential conversion of brigades from individual to station missioning. More specifically, two brigades were converted at the beginning of FY 2000, and the other three were converted at the beginning of FY 2001. Key findings include the following:

• Station missioning increased production of high-quality contracts by about 8 percent overall during this time.
• For individual stations, the effect of moving to station missioning depends on the level of mission difficulty. For example, as predicted by a theoretical analysis (detailed in Appendix B), for stations for which missions were unusually easy or unusually difficult, converting to station missioning tended to reduce productivity.
• Converting to station missioning appears to have reduced productivity for about 10 percent of the station-month pairs in our data.

Chapter Six concludes with a discussion of the policy implications of the research findings.

Implications for Policy

Based on our findings, we believe that the Army should adopt modest, perhaps gradual, changes in the ways that recruiters are evaluated. Although the traditional performance metrics are clearly flawed, we are reluctant to recommend an immediate and wholesale adoption of our preferred performance metric for four major reasons:

1. The current missioning process and associated awards (such as badges, stars, and rings) are a deeply ingrained part of the Army recruiting culture. Sudden, dramatic changes are likely to meet resistance and could undermine recruiter morale and productivity at a time when declines in enlistments would be extremely costly.

2. Relative to traditional measures, the PPM is conceptually complex and requires fairly sophisticated econometric analysis of local markets. Implementation of the PPM would place an additional burden on USAREC and, perhaps more importantly, will not be transparent and intuitive to personnel in the field. Although we believe that this new performance metric would be more equitable, perceptions are sometimes more important than reality.

3. The form of the PPM depends on assumptions about enlistment supply, recruiter behavior, and the relative value of enlistments in different categories. Additional analyses should be conducted before settling on a particular version of the PPM.

4. Our research focused primarily on three markets: high-aptitude seniors, high-aptitude graduates, and all other contracts.
Our exploratory work distinguishing males and females indicates that substantial distinctions exist between these segments that should also be considered in the design of a performance metric. Indeed, there are likely to be other segments—based on education or MOS preferences, for example—that are worth considering.

Despite these caveats, we recommend that USAREC consider some short-term adjustments to its procedures. In particular:

- Improve mission allocation algorithms to reflect variations in market quality and differences in market segments.

  Current mission allocations do not do a very good job of adjusting for station-area differences in crucial economic and demographic factors that affect the productivity of recruiter effort and skill. Many of the discrepancies between the PPM and the traditional measures stem from the failure to account adequately for market differences in allocating missions. The pattern of such differences varies by market segment and can also be influenced by command-level policies such as the distribution of MOSs.

- Lengthen the performance window to at least six months or smooth monthly rewards by reducing emphasis on station-level mission accomplishment.

  Until at least six months of production have been observed, recruiting outcomes at the station level are dominated by randomness. The importance of this randomness is magnified by a system that provides a discontinuous and significant reward (i.e., bonus points) for making mission in a single month. A substantial number of stations

---

4 Recently, USAREC began implementing a battalion mission-allocation model based on past enlistments in the Army and other services. We believe that this approach has merit and is likely to improve missioning, at least at the battalion level. Such a model, however, is not used to allocate the battalion mission to the station level, nor is it flexible enough to adjust adequately to relative changes in the local recruiting environment.
that often achieved regular Army mission are actually less, or no more, productive than many of their counterparts that make mission less frequently.

- Consider a more refined system of rewards for additional enlistment categories such as males, youth with more education or higher AFQT scores, youth with skills enabling them to fill critical MOSs, or those willing to accept longer terms of service.

Our exploratory research distinguishing males and females suggests that other market segments that we have not analyzed may also differ significantly in terms of recruiting difficulty. These differences are likely to vary systematically across the market areas of different stations. They could also vary when there are changes in the distribution of needed enlistments by MOS. It would not be advantageous to allocate missions for detailed subcategories, however, especially if mission accomplishment in a single month continues to lead to substantial bonus points. This is because doing so would increase the importance of randomness in performance evaluation and rewards. However, explicit recognition of market-quality differences in establishing recruiting goals or even a supplemental reward system providing points based on the distribution of enlistments among categories of differing importance to USAREC would better reward productivity as well as provide additional incentives for recruiters to help meet overall Army objectives.

- To minimize resistance, include education and outreach when implementing reforms.

Organizational change is always difficult, especially when there are perceived winners and losers. Modest efforts to explain and justify the changes could substantially increase their acceptance. If the performance measures are perceived as fair, with every station given a reasonable chance of success if their recruiters work hard, resistance will be reduced.
Although it is impossible to quantify the productivity gains that could emerge from such reforms, they are likely to dwarf any costs of implementation. As demonstrated in previous research (Dertouzos and Garber, 2006), better mission allocations during 2001 to 2003 could have improved average recruiting productivity for high-quality enlistees by nearly 3 percent. Much of this gain would have been due to an increased willingness on the part of stations that had a recent history of success (by conventional measures) to work harder and be more responsive to mission increases. There is substantial reason to believe that using a performance metric that better reflects Army values and more accurately assesses recruiter effort and skill would also have significant benefits.