Use of Self Assessments in Estimating Levels of Skill Retention

Joel D. Schendel and John C. Morey
US Army Research Institute Field Unit
Fort Benning, Georgia

M. Janell Granier and Sid Hall
Auburn University

Abstract

The focus of this research was on one approach toward predicting task retention and refresher training requirements--self assessment. Prior to zeroing their weapons for annual M16A1 rifle qualification (record fire), 153 permanent party soldiers completed a questionnaire designed to collect information on their previous marksmanship experiences as well as their beliefs about how they were going to shoot at record fire. No special instructions were provided about the meanings of questions or possible responses. Questionnaire items then were correlated with rifle qualification scores based on 40 target exposures. Predicted scores accounted for about 10% of the variance associated with record fire scores. However, for those soldiers whose confidence in the accuracy of their predictions exceeded 90%, predicted scores accounted for 25% of the record fire score variance. The best predictors of record fire performance were remembered most recent record fire performance and predicted performance. Subjects, generally, overestimated their actual performance and were biased heavily toward predicting success. The usefulness of the self-assessment approach to skill retention estimation is discussed together with practical suggestions toward refining the precision of questionnaire techniques.

Introduction

Most Army jobs require that soldiers achieve and maintain proficiency on scores of tasks. Because of resource constraints it is impossible to train every soldier on every task to the degree necessary even to minimize forgetting. As a result, some amount of periodic refresher training must be provided. This much is clear. What is not clear is how frequently individuals require refresher training to sustain proficiency on particular tasks. As Schendel and Hagman (1982) indicate, if time intervals between successive sessions are too long, then performance may fall below acceptable levels and entail considerable risk. Emergencies may arise requiring corrective action before an individual has had a chance to retrain. If time intervals between sessions are too short, then administrative costs necessarily are inflated.

What can be done to remedy this problem? One possible approach toward managing refresher training involves the use of self assessments . Under this approach, priorities would be assigned to tasks to be trained (managerial decision). Unit commanders then would determine where they are weakest, next weakest, etc by sampling soldiers' estimated skill levels on these tasks. Priorities would be weighed against weaknesses as well as resources (time, travel requirements, equipment constraints, etc) and unit refresher training schedules developed accordingly. Of course, self assessments need not be used alone. They may be used to identify soldiers requiring training on particular tasks and, consequently, may be used to refine predictions derived from other, more generalized approaches (e.g., Rose & Ford, 1982).

The advantages of the self assessment approach are that it is simple and economical--self assessments can be collected using only a questionnaire. Self assessments also can be very accurate (e.g., Levine, Flory, & Ash, 1977). The main disadvantage of self assessments is that they are subject to distortion, both intentional and unintentional. For example, the stereotype of a "good soldier" may influence self assessments. As a result, soldiers may be reluctant to admit needing training on tasks which they should know. Alternatively, soldiers may unintentionally overestimate their abilities to perform a task as a result of their overconfidence. Several methods for handling these problems are discussed later in this report. Many of these methods also have been covered by Burnside (1982). This research examines the feasibility of using self assessments in predicting soldiers' task retention and refresher training requirements.

Method

Subjects

The subjects were 153 male (n = 147) and female (n = 6) permanent party soldiers assigned to the 1st and 2d Infantry Training Brigades and Infantry Training Group at Fort Benning, Ga., who fired record fire between 26 and 27 March and 4 and 5 June 1982, and who volunteered to participate in this research. Subjects completed the questionnaire in groups ranging in size from approximately 15 to 60. Reported times since last record fire ranged from 1 to 60 months (n = 147, M = 12.80, SD = 10.93).

Materials and Procedure

Questionnaire data. Prior to firing record fire, soldiers reported to a 25-meter range to zero their weapons. Zeroing involves firing series of three-round shot groups at a specially designed 25-meter target and adjusting the rifle's sights until point of bullet impact coincides with point of aim. Prior to zeroing, soldiers assembled in bleachers to receive a safety briefing. After soldiers had assembled in the bleachers but before they had received this briefing, they were informed about the purpose of this research. Subjects were told merely to answer the questions as accurately as possible. No special instructions were provided about the meanings of questions nor were possible responses discussed.

Performance data. After zeroing, all soldiers were transported M16A1 rifle qualification range. On arriving, each subject was assigned randomly to one of eight firing lanes. All scoring was done by independent support personnel. These personnel were fully informed about scoring procedures and the purpose of this research prior to the onset of testing. In addition, an experimenter and numerous range personnel were available to assist in scorekeeping and to answer questions.

The course-of-fire consisted of 14 "F-type" silhouettes and 26 "E-type" silhouettes. F-type silhouettes are designed to appear like the head and shoulders of a man and were seen at 50 ($n = 5$) and 100 ($n = 9$) meters. E-type silhouettes, designed to appear like the head and torso of a man, appeared at 150 ($n = 10$), 200 ($n = 8$), 250 ($n = 5$), and 300 ($n = 3$) meters. The first half of this course was shot from the foxhole supported position; the second half from the prone unsupported position. Targets fell when hit. Record fire scores and related shooting classifications are as follows: 0 to 22--Unqualified; 23 to 29--Marksman; 30 to 35--Sharpshooter; 36 to 40--Expert.

Results

Pearson correlation coefficients (rs) first were computed on all subjects' questionnaire and record fire data. Subjects' predicted and actual record fire scores correlated positively, $r = .26$ ($n = 150$, $p < .01$). A similar result was obtained when computing the r between subjects' predicted shooting classifications and their actual classifications, $r = .25$ ($n = 152$, $p < .01$). While these rs appear low, they were much stronger than others which we expected may be quite strong. For example, reported experience outside the Army firing a rifle or a shotgun failed to correlate with performance at record fire, $r = -.10$ ($n = 152$, $p > .05$).

One of the best predictors of record fire performance was soldiers' remembered shooting classifications from their most recent record fire, $r = .38$ ($n = 133$), $p < .01$). Remembered most recent record fire scores also correlated with record fire performance, $r = .29$ ($n = 121$, $p < .01$). Also, remembered classifications and scores related highly to those predicted. Remembered classifications correlated $r = .65$ ($n = 132$, $p < .01$) with predicted classifications and $r = .53$ ($n = 130$, $p < .01$) with predicted scores; remembered scores correlated $r = .67$ ($n = 120$, $p < .01$) with predicted classifications and $r = .63$ ($n = 118$, $p < .01$) with predicted scores.

More extensive analyses then were carried out on questionnaire data corrected for internal consistency. This correction involved determining whether each subject's response to the question "How many targets out of 40 do you think you will hit at record fire?" was consistent with his (or her) responses to the following questions: "What is the highest number of targets out of 40 you feel you are likely to hit today?" "What is the lowest number of targets out of 40 you feel you are likely to hit today? For example, if a subject indicated he (she) believed he (she) was going to hit 27 targets at record fire, but then indicated that the highest (lowest) number of targets he (she) felt likely to hit was less (more) than 27, the subject's data were not included in the analyses that follow.

Most of the rs computed following this correction were similar in magnitude and direction to the rs reported above. Particularly noteworthy, however, was the r between soldiers' predicted and actual record fire scores which showed some improvement, $r = .32$ ($n = 124$, $p < .01$). In addition, the r between reported months since last record fire and record fire performance achieved significance, $r = -.19$ ($n = 123$, $p < .05$). Reported months since last record fire should correlate negatively with record fire performance, at least to the degree that this variable reflects the length of the interval between successive refresher training periods.

As indicated earlier, subjects were asked to predict their shooting classifications. They also were asked to estimate their chances of being correct in these predictions. In general, subjects who estimated their chances of being correct at 90% or 100% were more accurate in their predictions than subjects who gave lower confidence estimates. In particular, predicted shooting classifications and record fire scores correlated $r = -.01$ ($n = 36$, $p > .05$) for subjects who estimated their chances of accurate prediction at less than 60%. This r was .24 ($n = 47$, $.05 < p < .10$) for subjects who estimated their chances of accurate prediction as being between 60% and 80%. And, r equalled .42 ($n = 41$, $p < .01$) for subjects who estimated the likelihood of accurate prediction at 90% or 100%. Similarly, the rs between predicted and actual record fire scores were nonsignificant for subjects who estimated their chances of accurate prediction at less than 60%, $r = .23$ ($n = 36$, $p > .05$) or between 60% and 80%, $r = .04$ ($n = 47$, $p > .05$). However, this r was highly significant for subjects who estimated their chances of accurate prediction at 90% or 100%, $r = .50$ ($n = 41$, $p < .01$).

Estimation errors (i.e., algebraic errors) were inversely related to record fire scores, $r = -.73$ ($n = 124$, $p < .01$). An examination of these errors indicated that the predictions of subjects who fired well at record fire, generally, were more accurate than those who fired poorly. This suggests that subjects who performed well also were more skilled at self assessment. On the other hand, this result may merely reflect a general tendency to overestimate shooting ability: 75% of the subjects predicted they would hit more targets than they actually hit. If most subjects thought they would do well, good shooters naturally would show less error in their predictions than poor shooters.

Only five (4%) subjects predicted that they were going to fail to qualify (i.e., hit less than 23 targets). Actually, 34 (27%) subjects failed. More interestingly, however, three of the five subjects who predicted that they were going to fail did in fact fail. And, the remaining two who passed, passed only by two points. Conclusions are limited by the small sample size. However, this result suggests that, while subjects were biased heavily toward predicting success, those who predicted failure were quite accurate.

Discussion

Correlations between predicted and actual scores were not high. At best, predictions accounted for only about 10% of the variance associated with record fire scores. Nevertheless, these data must be regarded as encouraging for the following reasons:

Record fire scores are notoriously unreliable. Equipment failure, scorer bias, variations in light conditions, vegetation, and terrain all contribute to this unreliability (e.g., Marcus & Hughes, 1979).

No special instructions were provided subjects as they completed the questionnaire. The idea was to obtain a baseline estimate of soldiers abilities to assess their own skills under field conditions.

The data appear logically consistent. Predicted and actual performances correlated positively as did predicted and remembered performances. The correlation between predicted and actual scores showed some improvement when data were corrected for internal consistency. And, subjects who expressed the most confidence in their predictions also were most accurate.

Future research should focus on the problem of improving the accuracy of subjects' self assessments. Several manipulations already are known to improve this accuracy (e.g., Burnside, 1982). Our experience conducting this research suggests at least three further manipulations:

1. Provide subjects as much information relevant to the formation of accurate self assessments as possible. This information is conveyed largely by descriptions of the task, conditions, and standards. However, it also may be beneficial to remind subjects how they did previously or to refresh their memories for a task using a demonstration.

2. Have subjects provide confidence ratings along with their self assessments. Confidence ratings appear a good index of the accuracy of these assessments. In this research, the predicted scores of subjects who estimated their chances of accurate prediction at 90% or 100% accounted for 25% of the variance associated with record fire scores. Similar results have been obtained by Fischhoff and MacGregor (1981).

3. Instruct subjects to check their responses for internal consistency as they complete their self-assessment questionnaires.

References

Burnside, B. L. Subjective appraisal as a feedback tool (ARI Technical Report). Alexandria, Va: U.S. Army Research Institute for the Behavioral and Social Sciences, May 1982.

Fischhoff, B., & Mac Gregor, D. Subjective confidence in forecasts (PTR Technical Report PTR-1092-81-12). Woodland Hills, Ca: Perceptronics, December 1981.

Levine, E. L., Flory, A., & Ash, R. A. Self-assessment in personnel selection. Journal of Applied Psychology, 1977, 62, 428-435.

Marcus, A., & Hughes, C. R. An evaluation of a technique for using combat training theater (CTT) for periodic rifle marksmanship proficiency training and qualification (ARI Research Problem Review 79-7). Alexandria, Va: U.S. Army Research Institute for the Behavioral and Social Sciences, May 1979.

Rose, A. M., & Ford, P. Acquisition and retention of soldiering skills: Report of year 1 progress (AIR Technical Report AIR/IR-88601-4/82). Washington, DC: American Institutes for Research, April 1982.

Schendel, J. D., & Hagman, J. D. On sustaining procedural skills over a prolonged retention interval. Journal of Applied Psychology, 1982, 67, 605-610.