

THE STRENGTHS AND LIMITATIONS OF LECTURE-BASED TRAINING IN THE ACQUISITION OF ERGONOMICS KNOWLEDGE AND SKILL

Jeffrey K. Liker, Susan M. Evans, Sheryl S. Ulin

Center for Ergonomics, Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109 (U.S.A.)

and Bradley S. Joseph

Ford Motor Company, 104 Central Laboratory, 15000 Century Drive, Dearborn, MI 48120 (U.S.A.)

(Received June 17, 1988; accepted in revised form June 10, 1989)

ABSTRACT

A common approach to training designers of workplaces to incorporate ergonomic considerations in their designs is a two- to five-day course based primarily on lectures by experts. A quiz, designed to test the acquisition of ergonomics factual knowledge and skill in judging the degree of physical stress in various job configurations, was given to 147 participants before and after four days of a five-day short course based principally on lectures by university faculty and staff. The major findings were as follows. First, there was a considerable lack of factual knowledge and a high level of error in judging the level of stress prior to the training. Second, the training increased participants' factual knowledge but had little impact on their ability to accurately judge levels of stress in slides depicting real work situations. Third, participants' knowledge and skills before the training and their improvement as a result of the training were unrelated to prior education or training in ergonomics or experience with repetitive, manual work. These results are interpreted in light of prior research on design of effective health and safety training.

RELEVANCE TO INDUSTRY

The most common way of preparing staff in industry to apply ergonomics to the workplace is the several-day short course based primarily on lectures. This paper identifies the strengths and weaknesses of this form of training and suggests the types of additional training needed to skillfully apply ergonomics.

KEYWORDS

Ergonomics, health and safety, skill acquisition, training, cognitive complexity.

INTRODUCTION

Ergonomic research and development on the relationship between workplace design and physical stresses on operators has yielded a large and important body of knowledge and practical tools for designing ergonomically sound workplaces. Perhaps the primary mechanism for transferring this scientific knowledge to actual designers of workplaces is training. Indeed, it has been argued that lack of training is one of the principle barriers to effective use of ergonomic tools and knowledge (Rhomert and Laurig, 1977; Liker et al., 1984; Smith and Smith, 1984; Montreuil and Lavile, 1985; Sherwood, 1986). Sherwood (1986, p. F-20) notes that in Britain many people with little prior training in occupational health and safety "in mid-career are being given new health and safety responsibilities", a situation which is paralleled in the U.S. The main approach to providing these new health and safety practitioners with skills and knowledge needed for their jobs is continuing education courses of relatively brief duration (several days or weeks), typically using lectures as the primary teaching approach.

Two important assumptions are embedded in this training model. First, it is assumed that the intuition people naturally develop on the job is not sufficient, in and of itself, to appropriately diagnose poorly designed jobs and take corrective action. Second, it is assumed that relatively short training courses based on lectures can impact this ability.

This paper addresses both of these assumptions by looking at participants' *factual knowledge* and *skills* before and after four days of lecture-based ergonomics training. Participants were trainees in a five-day short course on occupational ergonomics taught by faculty and staff at the University of Michigan (U of M) in the summer of 1986. A quiz designed to test both knowledge and skill was administered *prior* to the training and again *after* the first four days of training (after all relevant material had been covered). This study design treats subjects as their own controls—no separate control group was included. By looking at the specific areas in which trainees improved, as well as the types of errors trainees persisted in making, we gain insights into the effectiveness and limita-

tions of this approach to training. We suspected that lecture-based training, while strong in its ability to efficiently transfer large quantities of information, has serious limitations in helping students acquire real practical skills. An earlier pilot study (Liker and Joseph, 1986) provides some evidence for this.

In this paper the term *factual knowledge* will refer to understanding of facts, basic terminology (e.g., what is wrist flexion?), and some simple bivariate relationships (e.g., relationships between lifting posture and stress on the lower back). *Skill* refers to the ability to apply that knowledge in *novel* situations, in this case, the ability to look at jobs that were *not* used as examples in the training and determine their physical stressfulness. The distinction between "factual knowledge" and "skill" is analogous to what cognitive psychologists (cf. Anderson, 1980) call "declarative knowledge" (facts, simple relationships) and "procedural knowledge" (how to actually apply the knowledge to do things), respectively. We emphasize application to novel situations since every work situation is different and thus mere memorization of how instructors evaluated a job is of limited usefulness in practice.

We acknowledge up front a major weakness of the study—the lack of a comparison group that went through an alternative form of training such as a more experientially based training course. Thus, we can only make empirically derived statements about the types of knowledge and skills that showed the most and least improvements as a result of the course, not what kinds of training are superior to other kinds for knowledge and skill acquisition. Nonetheless, the last part of the paper speculates on the characteristics of training that most effectively lead to skill acquisition drawing on published literature in cognitive psychology.

HYPOTHESES

This study builds on an earlier pilot study conducted in a manufacturing plant with a smaller number of trainees who went through a much shorter training course (Liker and Joseph, 1986). In that study 43 employees were given a quiz before and after a four- or eight-hour training course. The quiz, which formed the basis of the

quiz in the present study, included a set of true/false and multiple choice questions, the factual part of the quiz, as well as a rating task in which trainees were asked to evaluate physical stress in slides of real work situations in their plant, the practical portion of the quiz. The hypotheses below draw partially on the results of this earlier study and partially on cognitive psychology literature.

H1: *The training will result in improvement in scores on both the factual and practical portions of the quiz.*

This was found in the earlier study. Related to this general hypothesis are two more specific hypotheses about which aspect of the quiz will benefit most from lecture-based training.

H1a: *The biggest improvements will be found in the factual portions of the quiz as compared to the more practical portion of the quiz.*

H1b: *Particularly large improvements will be found in the practical portions of the quiz in which trainees rate stress in computer-generated slides as compared to slides of real people in real work situations.*

These two more specific hypotheses focus on the strengths of lecture-based training in transferring simple concepts and factual information quickly, as well as the limitations in transferring complex cognitive skills. The factual portions of the quiz largely reflected rote memory. The questions reflected the same information as was presented in the course in a form similar to that presented. By contrast, the practical portions required applying and integrating the knowledge in a novel situation. Thus, H1a predicts the most improvement in the portion of the quiz that is most compatible with the mode of teaching.

H1b focuses on a related phenomena which has been described in the instructional learning literature (Gagne, 1970; Howe, 1984). Some of the slides were generated by computer and depicted a wireframe model of a human lifting boxes, while other slides showed real people performing jobs. The computer-generated slides include only relevant information and therefore are in a form more compatible with the simple decision rules presented in the training. By contrast, the real industrial slides include a great deal of irrelevant distracting stimuli which makes direct application of simple decision rules more difficult.

H2: *In rating back compression force, which in reality depended on three manipulated factors, trainees will use more simplistic algorithms that focus on one or two salient pieces of information. This type of error will be present before and after the training.*

We suspected that one of the problems with relying purely on lectures as a means of training is that trainees never really *integrate* the knowledge they gain. In practice there seems to be a general human tendency toward cognitive simplification of complex problems as has been shown in the literature on multi-causal reasoning (Nisbet and Ross, 1980; Kahneman et al., 1983).

H3a: *Trainees with appropriate formal education will score higher on the factual part of the quiz, while persons with relevant experience with manual labor will perform better on the practical portions of the quiz.*

This hypothesis is based on the earlier study (Liker and Joseph, 1986). Years of formal education was found to correlate only with scores on the factual portion of the quiz, while years of direct experience in manufacturing correlated only with the ability to rate the slides.

H3b: *Personal background will not be correlated with improvements in scores resulting from the training.*

Prior to the pilot study we expected that persons with more years of formal education would learn more from the lectures than their less educated counterparts. However, this was not the case. As stated in H3b, neither formal education nor manufacturing experience was related to improved scores on any part of the quiz.

STUDY METHODOLOGY

The training

The test instrument was administered to each student participating in an Occupational Ergonomics short course offered through The University of Michigan Engineering Summer Conference in the summer of 1986. The open enrollment course was staffed by Ph.D.s and M.D.s in ergonomics and related disciplines, most of whom were university professors. Six lectures were presented each day, with breaks between lectures. The course

introduced concepts and applications of psychology, physiology, anthropometry, biomechanics and engineering to workplace and work methods design. Emphasis was on control of injury, illness, lost time, and productivity problems due to upper extremity disorders, lower back stress, or excessive energy expenditure. It was emphasized that force, frequency, and posture work together to cause physical stress. Lectures incorporated extensive graphical examples drawing on the experience of lecturers (e.g., slides of actual jobs) and were augmented with demonstrations of software (e.g., 2D static-strength model) and other analytic methodologies (e.g., the NIOSH Work Practices Guide). There were also limited class discussions and some participation in problem-solving exercises, though total hands-on experience using the concepts presented in lectures was limited. * No homework was assigned.

The majority of the first two days of the course, approximately 11.5 h, was devoted to issues related to upper extremity cumulative trauma disorders: their measurement, control, and prevention. Of that time, approximately two hours focused specifically on job analyses for the upper extremity (i.e., shoulder and wrist). The following two days (13 h total) focused on issues related to whole body biomechanics, and low-back pain. Of these 13 h, approximately five were divided among topics on low-back pain in industry, manual job analyses and workplace design, and applications of the NIOSH Work Practices Guide (NIOSH, 1981). Most segments included presentation of basic physiology, simple principles to determine what were good and bad work situations, analytic procedures for evaluating jobs, and actual examples of good and bad jobs taken primarily from industry.

The distribution of quiz items, which focused largely on upper extremity cumulative trauma and lower back stress, is not representative of the time

allocated to various topics in the course. Although the allocation of time to specific quiz-related topics may seem low, the materials covered throughout the four days preceding the post-training quiz could all be considered relevant to and reinforcing of the primary issues of low-back and upper extremity stress covered in the quiz.

The quiz

The test instrument was an elaboration of a quiz developed by Joseph (1986). The quiz was divided into two major sections reflecting the distinction between *knowledge* and *skill*. The total time for administering the quiz was about 45 min.

Factual knowledge measures

Twenty-four questions, 19 true/false and five multiple-choice questions, were designed to test the knowledge and understanding of basic ergonomic facts presented in the training. These twenty-four questions covered general factual knowledge, anthropometry, and the stressfulness of specific work situations. The following are examples of the *true/false* questions:

"In most cases tools that are used by men can be used equally well by women."

"Wrist flexion causes a nerve to be pinched inside the wrist."

"An object that weighs ten pounds is so light that it can always be lifted safely."

"Hoists always reduce low-back stresses on the body when lifting heavy objects."

Ergonomics skill measures

The skill tested was the participants' ability to recognize the tendency of various working conditions to cause physical stress on particular parts of the body. Participants were shown 33 slides depicting worker postures and task conditions: 19 were actual photographs taken from industrial settings and 14 were slides of computer-generated figures lifting objects (see Fig. 1). Varying degrees of back stress were presented in 19 slides (five photographs of jobs and 14 computer-generated figures), while shoulder and wrist stresses were depicted by seven slides each. The weight of objects lifted or pushed/pulled, as well as task frequency, were provided to participants verbally and in written form on the quiz.

* The one exception was the NIOSH Work Practices Guide. The slide rule was handed out and participants used it to solve a problem. Several optional workshops on specific topics were offered as three-day follow-ons to the five-day short courses and these include extensive practice using the software and techniques presented in the five-day course. These follow-on workshops were *not* evaluated using the quiz.

Participants were instructed to look at the person performing the job in the slide, were read a statement describing the job, and were asked to rate the degree of stress for a specific body region using a five-point scale. An example of the job description and rating scale used to rate the degree of stress on the worker's **back** follows:

| Acceptable (Low Stress) | | Marginal Stress | | Unacceptable (High Stress) |
|----------------------------|---|--------------------|---|-------------------------------|
| 1 | 2 | 3 | 4 | 5 |

Slide 1. Place 25 lb. box on fourth tier of pallet. Box location is 24 inches in front of the body, at waist height. Task frequency is 48 boxes per hour.

The five-point rating scale above was "behaviorally-anchored" (Landy and Farr, 1980). That is, the levels of stress were related to the *need for action* to improve the job (NIOSH, 1981). The actual definitions of scale points were included in the quiz (*italics added*) as follows:

Acceptable (1–2) Stresses are sufficiently low that the job can be safely performed by virtually all workers. There is no need to change the operation (*below Action Limit in NIOSH, 1981*).

Marginal (2–4) Stresses could result in problems (e.g., fatigue, discomfort, injury) to some workers. Administrative or even engineering controls *may* be necessary (*between Action Limit and Maximum Permissible Limit in NIOSH, 1981*).

Unacceptable (4–5) Stresses are sufficiently high to cause problems for most workers (including fatigue, discomfort, or injury). Engineering controls are *recommended* to correct the problems (*above Maximum Permissible Limit in NIOSH, 1981*).

Note: "2" and "4" are on the borderline between the three categories.

"2" = high acceptable and low marginal stress.

"4" = high marginal and low unacceptable stress.

The "correct" score was determined by Ph.D.-level ergonomics experts. For lower-back slides, the *Work Practices Guide* (NIOSH) in combination with a computerized biomechanical static strength model (Garg and Chaffin, 1975) was used to determine the correct rating. For upper extremity slides, only *judgments* of the experts were used since the technical tools for upper extremity analysis were not as clear cut. There was considerable agreement among five experts on the "correct" stressfulness rating for upper extremity slides (Joseph, 1986, Appendix A). Each of the five experts rated eight shoulder slides and nine wrist slides. Correlations were computed between the stress ratings for all possible pairs of experts. The average correlation for the shoulder slides was 0.88 and for the wrist slides a more modest 0.61.

Computer-generated manual lifting figures

The computer-generated figures were carefully selected and warrant a more detailed description. The use of standardized, computer-generated postures made it possible to focus the trainee's attention on the primary risk factors associated with low-back pain (NIOSH, 1981; Chaffin and Andersson, 1984)—i.e., force, frequency, and posture—while systematically varying these factors. By regressing participants' stress ratings on the characteristics of the slides being rated, we could statistically model the judgment process of trainees. That is, we could study which factors were taken into account by trainees before and

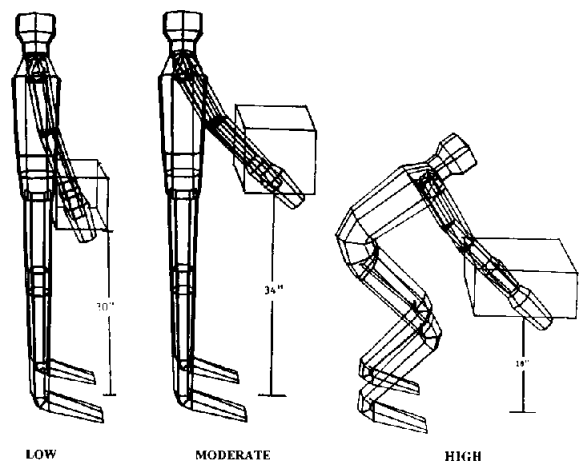


Fig. 1. Low, moderate, and high stress postures used for low back evaluation.

TABLE 1

Levels of force, frequency and posture used in computer-generated manual lifting slides

| Stress level | Force | Frequency | Posture |
|--------------|--------|--------------------|--|
| Low | 10 lb. | 48 units/ hour | hands at knuckle height, 6" in front of ankles |
| Moderate | 30 lb. | (none) | hands at waist height, 16" in front of ankles |
| High | 50 lb. | 196 units/ hour | hands at mid-thigh height, 18" in front of ankles |

after the training. The stimuli were designed so that trainees had to use all three variables to accurately rate all slides.

The computer-generated figures represented three postures, three forces, and two task frequencies. The postures were generated using the posture prediction capability of Garg and Chaffin's (1975) three-dimensional biomechanical strength prediction model. Hand locations and general posture identifiers, such as stand, stoop, or squat, drove the posture prediction. An integrated ergonomic design system (described in Evans and Chaffin, 1986) which accessed several computer-based human performance models, including the strength prediction model, produced the hard-copy graphical renditions of the postures shown in Fig. 1.

The levels of each condition, shown in Table 1, were selected such that any condition, taken at its lowest level, would produce an acceptable (i.e., "low") level of back stress. Increasing the level of any factor would increase the stress to either marginal or unacceptable levels. Predictions of back compression force and NIOSH lifting limits (see NIOSH, 1981), provided through the integrated design system, were used to assess the expert rating for the stressfulness of each condition.

Fourteen slides were selected from the possible set of 18 combinations; the actual stressfulness of the selected slides was uniformly distributed over the rating range. The order of slide presentation was randomized, but held constant across the pre- and post-training quiz.

The sample

All students, 147 practitioners from throughout the United States, participated in the pre-training

quiz, but attrition reduced the numbers taking the post-test to 131. To test for biases due to attrition, the pre-training quiz scores of those people who took the post-test were compared to those who did not take the post-test. Differences were small and not statistically significant.

The training attracted a very heterogeneous group. In terms of formal education, 16 percent had no college degree at all, while 49 percent had Associate or Bachelor's degrees, 25 percent had Master's degrees, and 10 percent had a Ph.D. or M.D. Among the college educated, major fields of study also varied a great deal—23 percent were in health-related fields, 35 percent in engineering, 26 percent in general liberal arts, and 17 percent in the basic sciences.

For a substantial portion of the sample this was not their first exposure to occupational health and safety or ergonomics. In all, 67 percent had some college level occupational health and safety courses, though a smaller number, 32 percent, had courses in occupational biomechanics in college. Almost half had some prior training seminars in occupational biomechanics.

Jobs were split almost equally between manufacturing and service sectors. The vast majority of trainees (87 percent) considered health and safety to be a central part of their current job, though relatively few worked exclusively in occupational health and safety. A large majority (66 percent) were responsible for the design of workplaces. Finally, most trainees (72 percent) claimed they had at least some experience actually performing repetitive manual work.

In short, this heterogeneous group included people whose jobs gave them substantial responsibility for health and safety and the majority had direct responsibility for workplace design. Almost all had some education or personal experience relevant to the material presented in the ergonomics short course.

RESULTS

H1: Scores before and after training

Table 2 presents the percent of correct responses to each type of question, factual questions and ratings of slides, before and after the training.

TABLE 2

Percent correct on factual questions and ratings of slides before and after training ($n = 131$ subjects)

| | Expected by chance ^a | Pre-training | Post-training | Change |
|------------------------------------|------------------------------------|--------------|---------------|-------------------|
| <i>Factual questions</i> | | | | |
| All questions (24) | 45% | 72% | 80% | + 8 ^c |
| Job analysis questions(5) | 27% | 39% | 59% | + 20 ^c |
| <i>Slides ratings</i> ^b | | | | |
| All slides | 20% | 34% | 36% | + 2 |
| Back (computer generated) | 20% | 38% | 41% | + 3 |
| Back (lifting) | 20% | 36% | 30% | - 6 |
| Back (push/pull) | 20% | 11% | 12% | + 1 |
| Shoulder | 20% | 38% | 40% | + 2 |
| Wrist | 20% | 31% | 34% | + 3 |

^a This is the score that would be expected on average if participants merely randomly guessed.^b Correct in the case of the slides meant the subject's rating on the five-point scale matched exactly the expert's rating for that slide.^c $p < 0.05$.

Also shown are the percent correct that would be expected by chance alone, assuming participants simply guessed. The pre-training scores show that participants had some knowledge and skill at the outset, substantially more than one would predict by chance alone. In most cases participants did over one and one-half times better than would be expected by chance. However, there was clearly room for improvement (scores ranging from 11 percent to 72 percent correct).

The findings partially support H1. The differences in scores before and after the training

show substantial improvement in ergonomics knowledge, but only marginal or no improvement in skills (all statistical tests of change are based on paired *t*-tests). Knowledge of ergonomics, as measured by the factual questions, gained significantly. Before the course, the average percent of these questions answered correctly for the 141 trainees was 71.6 percent. The average score on the post-training test was 80.3 percent, an eight-point improvement ($p < 0.05$). Five of the multiple choice questions dealt explicitly with anthropometry and workplace design. The average pre-

TABLE 3

Trainee's average ratings before and after training compared to expert ratings

| Slide type | Number slides | Mean expert ratings ^a | Trainees ratings of slides ^b | | | Inaccuracy = Subject rating - Correct ^c | | | |
|------------------|---------------|----------------------------------|---|--------------------------------|-------------------------|--|--------------------------------|-------------------------|-----------------------|
| | | | Pre-training ($n = 147$) | Post-training ($n = 131$) | Change ($n = 131$) | Pre-training ($n = 147$) | Post-training ($n = 131$) | Change ($n = 131$) | |
| Back (computer) | Mean | 14 | 3.0 | 3.0 | 3.1 | 0.11 ^{d,**} | 0.89 | 0.79 | -0.10 ^{d,**} |
| | S.D. | | | (1.3) | (1.3) | (1.1) | (0.26) | (0.24) | (0.30) |
| Back (lift) | Mean | 2 | 3.5 | 3.8 | 4.0 | 0.24 ^{d,**} | 0.77 | 0.85 | 0.09 |
| | S.D. | | | (0.84) | (0.80) | (0.92) | (0.55) | (0.50) | (0.71) |
| Back (push/pull) | Mean | 3 | 3.0 | 3.8 | 4.1 | 0.35 ^{d,**} | 1.7 | 1.8 | 0.13 ^{d,**} |
| | S.D. | | | (0.72) | (0.67) | (0.76) | (0.49) | (0.38) | (0.53) |
| Shoulder | Mean | 7 | 2.9 | 3.1 | 3.3 | 0.20 ^{d,**} | 0.89 | 0.81 | -0.08 ^{d,*} |
| | S.D. | | | (1.4) | (1.5) | (1.1) | (0.14) | (0.29) | (0.39) |
| Wrist | Mean | 7 | 3.3 | 3.2 | 3.5 | 0.32 ^{d,**} | 0.97 | 0.90 | -0.06 |
| | S.D. | | | (1.2) | (1.2) | (1.3) | (0.36) | (0.34) | (0.41) |

^a Average expert rating for all slides in category on five-point scale.^b Average trainee rating for all slides in category on five-point scale.^c Inaccuracy, the magnitude of errors in ratings, is the absolute value of differences between trainees' ratings and correct ratings averaged across all slides in each category.^d Paired *t*-test: ** $p < 0.01$, * $p < 0.05$.

training percent correct for these workplace design questions was 39 percent, while the post-training score was 59 percent ($p < 0.05$), a twenty point increase.

By contrast, changes in percent correct for the ratings of slides ranged from a 6 percent reduction in correct ratings (actual photographs of manual lifting) to a 3 percent improvement (computer-generated slides and wrist slides); none of these changes were statistically significant. Thus, as predicted in H1a, *the degree of improvement in factual knowledge was greater than improvement in actual skills in rating the stressfulness of jobs.*

Percent correct is a crude measure in the case of the five-point stress ratings as it assumes trainees are either right or wrong, rather than displaying varying degrees of closeness to the expert's ratings. Table 3 provides a more detailed look at changes in how trainees rated the slides. The first two columns show the number and correct mean ratings for each set of slides as determined by the experts. The final columns consider the amount by which participants deviated from the correct rating—the absolute value of the difference between the ratings by participants and experts (i.e., “inaccuracy”).

The pre-training results show that, on average, trainees were off by about one point (high or low) when compared to the correct ratings. That is, they *may* take more or less action than is actually called for by the job design. The trends resulting from the training were mixed. There was a significant improvement in mean accuracy for the computer-generated back slides and the shoulder slides ($p < 0.05$). For the computer-generated slides this amounted to an improvement of one-tenth of one point on average. However, ratings actually became *less* accurate for photographs of workers lifting and pushing or pulling (large pallet trucks of materials), though only the push/pull changes were significant. These results partially support H1b. The ratings of computer-generated slides improved to a greater extent than all of the photographs of real jobs with the exception of the shoulder slides.

The first three columns of Table 3 show the average ratings of slides by trainees and indicate one reason why participants were making errors. That is, they tended to overrate the degree of stress prior to training, particularly for the

push/pull slides. This type of error became significantly more prevalent ($p < 0.01$) as a result of the training. *Thus, training sensitized trainees to the concept of ergonomic stress to the point where they overrated stress on slides.*

Overall, the training resulted in clear increases in factual knowledge, but only modest gains or even losses in accuracy in rating jobs depicted on the slides. The most notable gain was in the rating of the computer-generated back slides, though this was still modest in size. One source of error noted was the tendency to overrate stress on slides, a tendency that was further reinforced by the training. By looking in greater depth at the judgment process used to rate the computer-generated slides, further insight can be gained into sources of errors.

H2: The judgment process for computer-generated slides

The level of control gained by computer-generated back slides provided an opportunity to statistically model the judgment process used by trainees. That is, we could study the importance trainees attached to each of the three factors manipulated: task frequency, operator posture, and force required to hold or move the object (related to object weight).

Multiple regression was used to model the judgment process. The dependent variable, stress ratings of the slides, was regressed on attributes of the slides expressed as binary (dummy) variables. Four fitted regression equations are presented in Table 4 based on the expert ratings, pre-training trainee ratings, post-training trainee ratings, and the change from pre- to post-training. For these regression equations, one case represents one person's rating of a particular slide—the rating is the unit of analysis. *

* This is actually an overstatement of sample size as each subject was contributing multiple judgments of slides and hence multiple data points to the regression. A given subject's multiple ratings are not statistically independent. An approach which is on the extreme conservative side is to use the number of subjects as the sample size, reducing the n to 131. This was tested and all of the regression coefficients in the before and after equations remained statistically significant.

TABLE 4

Regression of lower back stress ratings on key slide attributes before and after training (computer-generated slides)

| Slide attributes ^a | | Dependent variables | | | |
|-------------------------------|---------|---------------------|----------------------|-----------------------|--------------------|
| | | Expert ratings | Pre-training ratings | Post-training ratings | Change (pre-post) |
| High frequency | b | 0.60 | 1.15 | 1.17 | 0.02 ^{NS} |
| | S.E. | (0.24) | (0.04) | (0.04) | (0.05) |
| High posture | b | 2.60 | 00.97 | 1.24 | 0.28 |
| | S.E. | (0.30) | (0.05) | (0.05) | (0.06) |
| Medium posture | b | 1.12 | 0.46 | 0.75 | 0.28 |
| | SE. | (0.28) | (0.05) | (0.05) | (0.06) |
| High force | b | 2.33 | 1.81 | 1.67 | -0.15 |
| | S.E. | (0.27) | (0.05) | (0.05) | (0.06) |
| Medium force | b | 1.23 | 1.08 | 0.94 | -0.15 |
| | S.E. | (0.31) | (0.05) | (0.05) | (0.07) |
| Intercept | | 0.46 | 1.11 | 1.09 | 0.01 |
| | $R^2 =$ | 0.95 | 0.54 | 0.56 | 0.02 ^b |
| | $n =$ | (14) | (2058) | (1834) | (1834) |

NS = not significant; all other coefficient significant at 0.05 level.

^a Dummy variables with missing category equal to low on frequency, posture and force.^b Overall change equation is significant at 0.05 level based on *F*-test.

The "expert ratings" model shows how individual attributes *should* be weighted in a linear model to best predict the correct rating. This linear regression equation has an extremely good fit to the data ($R^2 = 0.95$) so trainees who considered each factor and weighted the factors properly could come extremely close to matching the ratings by the experts. ** All of the coefficients are highly significant despite the small sample size ($n = 14$ slides) indicating that trainees needed to take into account all levels of all three attributes in their judgments to match the correct answers.

The pre- and post-training ratings show that trainees deviated substantially from the correct model. Frequency was given twice as much weight as it warranted, posture was given half the weight it warranted, and object force was slightly undervalued.

These results also show that the model used by trainees changed significantly ($p < 0.05$) after the training. This can mainly be attributed to a large increase in the consideration given to posture ($p < 0.05$). Trainees became more sensitized to the

importance of posture and were better able to interpret how different postures related to low-back stress. However, this was at the expense of object force. Trainees decreased their emphasis on force further, beyond the pre-training level ($p < 0.05$).

This shift of focus from one factor to another, rather than integration of factors, is consistent with the prediction of H2. After the training trainees increased the degree to which they considered posture in their judgments but at the expense of considering the force of the object. Thus, trainees tended toward simplistic models. The training seems to have led to a substitution of emphasis rather than to more complex models incorporating all three variables.

H3: The role of personal background

An exhaustive analysis of the correlations between the many personal background characteristics measured (i.e., experience with repetitive manual jobs and education) with rating accuracy prior to the training found virtually no significant correlations (results not displayed in tabular form). The only exception was general education. The college-educated slightly outperformed those with no college prior to the training, and among the

** In fact, the equation used to predict back compression force (based on the NIOSH Work Practices Guide) was not simply a weighted average of force, frequency and posture. However, these results indicate a simple linear weighting procedure works quite well (explaining 95% of the variance).

college educated, those who majored in health-related fields or engineering had a slight advantage over those with a liberal-arts background. However, even these correlations were small and inconsistent across different outcome measures. Thus, hypothesis H3a derived from the earlier pilot study was not supported in this study.

Also examined were correlations between background factors and *changes* in accuracy over the four days of the training. In this case hypothesis H3b predicts no correlation and the results supported the hypothesis—there were no significant correlations between personal background and improvement.

DISCUSSION

Even before training, trainees had substantial factual knowledge and some ability to detect stressful postures in slides, at least better than could be predicted by chance alone. Nonetheless, they missed 28 percent of the factual questions and in two-thirds of the cases their judgments were in error compared to the expert's ratings of slides. Thus, their intuition was clearly lacking. Training significantly improved their factual knowledge, but led to only modest gains or even losses in accuracy in rating jobs depicted in slides.

A more detailed look at the findings provides some clues as to why the error rate was so high. First, there was a tendency to overrate stress levels before the training, a bias which became significantly more severe as a result of the training. The training apparently sensitized trainees to the impact of the work situation on stress levels and trainees overcompensated by exaggerating stress levels. In practice, this means trainees are more likely to err on the conservative side which is acceptable for a pure health and safety viewpoint, but may not be economically acceptable. This sensitization may be what Rohmert and Laurig (1977) were observing in their training evaluation. They noted that when plant personnel rank ordered the importance of ergonomics problems in their plants, they ranked the issues covered by the course as more important after the training than they had before the training. The training had sensitized participants to a particular set of prob-

lems, though we do not know from their study whether trainees' judgments after the training were more or less accurate than their judgments before the training.

Second, among the slide portion of the quiz, training improved accuracy principally for stress ratings of computer-generated slides as opposed to the slides of actual people performing jobs. The instructional learning literature demonstrates that training materials should begin with simple visual aids and tests which abstract a small number of bits of directly relevant information (Gagne, 1970; Howe, 1984). The computer-generated slides provided only relevant information about force, frequency, and posture, as did the multiple-choice questions. By contrast, the photographs of real industrial jobs were highly complex stimuli, filled with irrelevant, potentially distracting information, e.g., environmental context, emotional expressions of people, the size and shapes of objects, etc.

Third, the error rates of judgments of push/pull slides were by far highest and for these slides the errors actually increased in the post-test. From a technical viewpoint, the forces that are transmitted to the lower back when humans are pushing or pulling objects are more complex than in manual lifting, and apparently this type of stress is not intuitive to trainees. Perhaps this is because one must consider how a horizontal action is transferred into a vertical back compression force. Also, depending on the posture, other forces may be important as well (e.g., sheer force). The push/pull situation was only briefly covered in course material, but it is unclear why accuracy declined. Conceivably the decision rules provided to understand back compression force in manual lifting somehow misled trainees when applied to pushing and pulling.

Fourth, regression analysis of ratings of the computer-generated slides were used to identify and understand the types of errors made by trainees and examine how the training influenced these errors. The results showed that posture was the most under-emphasized attribute prior to training and substantially increased in importance by the post-test. Thus, one type of error made by trainees was reduced improving their accuracy. However, this was at the expense of force which had been slightly under-emphasized prior to train-

ing and received even less consideration in judgments after training. Frequency was highly overrated before and after training. The relative neglect of posture prior to training might be a result of the relative complexity of this stimuli. Frequency and force were given quantitatively, verbally and in the quiz, but for posture, trainees had to convert a graphic image to a quantitative stress rating. This interpretation of posture was not intuitive to trainees before training, but apparently the training had a major impact on this ability. This learning may have been enhanced by the fact that the training itself relied heavily on diagrams to depict work situations. Overall, the regression results suggest that trainees did not really learn in the training to deal with the complexity of applying a three-factor model to novel work situations.

Fifth, there was virtually no evidence that prior education and experience were associated with rating accuracy before training or changes in accuracy over the four days of training. The lack of correlation with most background factors prior to the training suggests that prior formal education and work experience did little to prepare participants for the ergonomic knowledge and skills tested in the quiz. Wherever the intuitive knowledge evident before the training came from, it was not from formal education or experience as measured by direct experience with manual work or observation of people performing manual work. The lack of correlation with changes in quiz scores suggests that people from very heterogeneous backgrounds all stand to gain from ergonomics training. It should neither be reserved for the educational elite, nor particularly stressed for those with little formal education.

What are the implications of these results for training? The five-day course did improve the participants' base of factual knowledge yet fared poorly in improving their actual skills in judging the stressfulness of jobs. The subtle skill of integrating this information and making complex judgments in novel situations was not learned by listening to lectures. This type of training is clearly not sufficient in and of itself for persons expected to judge physical stress in their professional roles.

What of the intuitive judgments made by the trainees? Did they really need to improve their knowledge and skill? The students in the course

represented a wide range of backgrounds, including hourly employees in plants, professional nurses, degree engineers, and medical doctors, many of whom had prior college coursework or workshops in health and safety and even ergonomics. As a group, prior to the training they performed significantly better than would be predicted by chance alone. Yet, the pre-training quiz revealed significant deficits in their ergonomics knowledge and ability to detect stressful work situations; and their scores on the tests were unrelated to their level of education, prior coursework in ergonomics, professional speciality, and degree of experience with repetitive manual work.

Do these findings suggest that formal education is a waste of time? Do they suggest that intuitive knowledge based on experience does not count for anything in making judgments of the stressfulness of jobs? As for the first part, we suspect that formal education gave these individuals much information (that we did not happen to capture in our factual questions) but it had not yet been translated into the particular practical skill of judging stress in work situations. Perhaps the development of this skill depends on practice, practice which until recently has been lacking as U.S. companies have only recently begun to systematically apply ergonomics to the design of their jobs.

As for the value of intuitive knowledge, the findings here contradicted those of the earlier study conducted onsite in a manufacturing plant (Liker and Joseph, 1986). One difference between the studies was that the slides in the manufacturing example were taken from the plant where the trainees worked and some had personal experience with those jobs. By contrast trainees in the summer course had no direct experience with the jobs depicted in slides. The recent emphasis on participatory ergonomics (Imada and Noro, 1990; Liker et al., 1989) places a high value on the role of participation of those persons who perform the job every day in the evaluation of ergonomics stresses and development of ideas for improvement. This is not to say that the intuition of employees is a substitute for the judgment of trained professionals and the use of formal analysis methods. Rather employees have access to additional data that can be useful in making ergonomics decisions. That is, they know how it feels to do the job. Ergonomics is still an imperfect

science and we cannot afford to ignore any significant sources of data.

WHAT KIND OF TRAINING MIGHT BE MORE EFFECTIVE?

These findings suggest the limitations of short lecture-based courses, but since there were no comparison groups with different training it is unclear how training should be designed to be more effective. For this, we depart from the study reported here and review ideas offered by prior research in cognitive psychology and health and safety education. Based on a literature review, Vojtecky (1985) suggests some health education training principles derived from learning theory. His principles emphasize the importance of positive reinforcement of what is learned in the training and the establishment of new norms of safe behavior for workers. He also emphasizes that instruction should be based on examples specific to the work situation and behaviors being taught. Translating these to ergonomics education, workplace designers should learn about the types of work situations they are likely to face, have an opportunity to practice using what they learn, get feedback on how well they are doing, and the use of ergonomics must become a *reinforced norm* back at their place of work.

These principles are consistent with experimental evidence in cognitive psychology. First, Bruner's (1965) studies suggest lecturers should teach students how to reason through problems by modeling the process (not simply the conclusions) and then asking students to try it. Second, lectures should rely on simple visual aids without distracting irrelevant information (as the computer-aided figures provided) and work up to more complex stimuli (Gagni, 1970). Third, it has been consistently shown that a series of relatively short practice sessions is much more effective than massed practice (Anderson, 1980). * It has also

* This was first shown in studies of Morse code training during World War II. Students were found to learn as rapidly with 4 hours a day practice as with 7 hours of practice. In a more recent example, Gay (1973) showed that spaced practice of algebra rules results in better retention than massed practice.

been shown that practice is even more effective when tasks are varied, as opposed to the identical task repeated over and over (Duncan, 1958). Finally, people learn best if they get feedback on their progress (Bilodeau, 1969; Anderson, 1980). This feedback should be immediate, while their recent performance is still in their active memory, and there should be a minimum delay between feedback and the next attempt which uses the feedback. In this case too much spacing can be counter productive.

In sum, the several-day, lecture-based training course provided a start at developing the factual knowledge base required to apply ergonomics to workplace design. However, this was only a start. Well-designed follow-up training is clearly needed to bolster the skills required to apply these complex principles. Further research is needed on the best ways to transfer ergonomic skills through human and possibly computer-assisted instruction.

ACKNOWLEDGEMENTS

This research was supported by the Engineering Summer Conference, The University of Michigan and, in part, by a grant from the Ford Motor Company. We thank Don Chaffin for permission to conduct this research as part of the summer conference and for helpful comments on an earlier draft of this paper and Jay Elkerton for his helpful comments used in a major revision of the paper. Finally, we thank Deb Liker for editing the final version of the paper.

REFERENCES

- Anderson, J.R., 1980. Cognitive Psychology and its Implications. W.H. Freeman, San Francisco.
- Bilodeau, I., 1969. Information feedback. In: E.A. Bilodeau (Ed.), Principles of Skill Acquisition. Academic Press, New York, pp. 255-285.
- Bruner, J.S., 1965. Learning and thinking. In: R.C. Anderson and D.P. Ausubel (Eds.), Readings in the Psychology of Cognition. Holt, Rinehard & Winston, NY, pp. 76-86.
- Chaffin, D.B. and Andersson, G.B.J., 1984. Occupational Ergonomics. Wiley, New York.
- Duncan, C.P., 1958. Transfer after training on single versus multiple tasks. J. Exp. Psychol., 55: 63-72.

- Evans, S.M. and Chaffin, D.B., 1986. Using interactive visual displays to present ergonomic information in workspace design. In: W. Karwowski (Ed.), *Trends in Ergonomics/Human Factors III*. Elsevier, Amsterdam, pp. 17–24.
- Gagne, R.B., 1970. *The Conditions of Learning*. Holt, Reinhart & Winston, NY, 2nd edn.
- Garg, A. and Chaffin, D.B., 1975. A biomechanical computerized simulation of human strength. *AIIE Trans.*, 7: 1–15.
- Gay, 1973. Temporal position of reviews and its effect on the retention of mathematical rules. *J. Educ. Psychol.*, 64: 171–182.
- Howe, M.J.A., 1984. *A Teacher's Guide to the Psychology of Learning*. Basil Blackwell, NY.
- Imada, A. and Noro, K., 1990. *Participatory Ergonomics*. Taylor & Francis, London.
- Joseph, B.S., 1986. A participative ergonomic control program in a U.S. automotive plant: Evaluation and implications. Ph.D. dissertation, Department of Industrial and Operations Engineering-Center for Ergonomics, the University of Michigan.
- Kahneman, D., Slovic, P. and Tversky, A. (Eds.), 1983. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- Landy, F.J. and Farr, J.L., 1980. Performance rating. *Psychol. Bull.* 87(1): 72–107.
- Liker, J.K. and Joseph, B.S., 1986. Ergonomic knowledge in a U.S. automotive plant: Correlates of practical and theoretical job analysis capabilities. In: H.W. Hendrick and O. Brown, Jr. (Eds.), *Human Factors in Organizational Design and Management*, Vol. 2. North-Holland, Amsterdam, pp. 225–231.
- Liker, J.K., Joseph, B.S. and Armstrong, T.J., 1984. From ergonomic theory to practice: Organizational factors affecting the utilization of ergonomic knowledge. In: H.W. Hendrick and O. Brown, Jr. (Eds.), *Human Factors in Organizational Design and Management*. North-Holland, Amsterdam, pp. 563–568.
- Liker, J.K., Nagamachi, M. and Lifshitz, Y.R., 1989. A comparative analysis of participatory ergonomics programs in U.S. and Japan manufacturing plants. *Int. J. Ind. Ergon.*, 3(3): 185–199.
- Montreuil, S. and Lavile, A., 1985. Ergonomics and workers cooperation in studying postural constraints to change the work situation. Paper presented at the Zadar Symposium in Yugoslavia on The Ergonomics of Working Postures, April 15–17.
- Nisbet, R.E. and Ross, L., 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Prentice Hall, NJ.
- NIOSH, 1981. *Work Practices Guide for Manual Lifting*. U.S. Department of Health and Human Services, March.
- Rhomert, W. and Laurig, W., 1977. Increasing awareness of ergonomics by in-company courses – a case study. *Appl. Ergon.*, 8(1): 19–21.
- Rossi, P.H. and Nock, S.L., 1982. *Measuring Social Judgments: The Factorial Survey Approach*. Sage Publications, Beverly Hills.
- Sherwood, R.J., 1986. Training for industrial hygiene in the UK. *Appl. Ind. Hyg.*, 1(3).
- Smith, L.A. and Smith, J.L., 1984. Observations of in-house ergonomics training for first-line supervisors. *Appl. Ergon.*, 15(1): 11–14.
- Vojtecky, M.A., 1985. Workplace health education: Principles in practice. *J. Occup. Med.*, 27(1): 29–33.