Verbal Estimation of Peak Exertion Intensity

Matthew M. Marshall, Thomas J. Armstrong, and **Marissa L. Ebersole,** University of Michigan, Ann Arbor, Michigan

The objectives of this research were to investigate the accuracy and precision with which trained and untrained participants estimate the magnitude of forceful exertion and to evaluate the mathematical relationship between actual and estimated exertion. Three groups of participants estimated, as a percentage of maximum voluntary contraction (%MVC), the magnitude of submaximal exertion for 12 simulated tasks. In addition to the control group, one group was exposed to one physical benchmark (100% MVC) and another to three benchmarks (25%, 75%, and 100% MVC) prior to force estimation. Error (estimated minus actual) significantly decreased ($p < .0001$) from 14% MVC to 4% MVC with one benchmark and to –3% MVC with three benchmarks, as compared with the control group. Furthermore, the standard deviation decreased significantly ($p < .0001$) from the control group (16.6% MVC) to the one-benchmark group (13.8% MVC) to the three-benchmark group (11.6% MVC), indicating improved precision. Significant interaction effects were observed, but their impact on main effects was negligible. Also, linear, power, and logarithmic regression models described the relationship between perceived and actual exertion equally well (R^2 = .64–.81). Applications of this research include improving the accuracy and precision of field-based psychophysical estimates of forceful exertion for epidemiological research and other field-based analyses.

INTRODUCTION

Researchers and practitioners of occupational ergonomics are faced with a limited number of valid and practical tools for assessing the magnitude of physical exertion. To assess the physical demands of using a screwdriver, for example, one might compare the torque required to fasten the screw with the corresponding strength of the target worker population. Although this approach is conceptually straightforward, a lack of valid tools to efficiently evaluate the physical demands of work poses a challenge. Techniques such as direct force measurement and electromyography (EMG) are often not practical for widespread application in occupational settings, so workers themselves are often relied upon to estimate intensity of physical exertion. This paper describes the results of an experiment that investigated the use of psychophysical magnitude estimation as a tool for evaluating forceful exertion.

The benefits of using psychophysical methods to assess exposure to forceful exertions are numerous. From a practical standpoint, psychophysical methods are less costly, in terms of time and money, as compared with the high cost and setup time of instrument-based techniques (Sinclair, 1995). Also, psychophysical methods provide an integrative measure in that they synthesize the many signals elicited from the peripheral working muscles, joints, and central nervous system (Borg, 1990). A criticism of psychophysical methods, however, is that they are inherently subjective and, consequently, are suspect in terms of validity. Direct force measurement and EMG provide "objective" data and may generally be more credible. This view, however, may be limited. EMG, though objective in that it provides a biophysical measurement of the muscle

Address correspondence to Matthew M. Marshall, 81 Lomb Memorial Dr., Rochester Institute of Technology, Department of Industrial and Systems Engineering, Rochester, NY 14623l; mmmeie@rit.edu. **HUMAN FACTORS,** Vol. 46, No. 4, Winter 2004, pp. 697–710.

load, is subject to many parameters controlled by the researcher, including muscle selection, calibration technique, signal processing, data analysis, and presentation issues. These variables can have a profound impact on the results that are obtained (Solomonow, 2000).

Like any measurement technique, a psychophysical estimate has some degree of error associated with it, and to make meaningful use of the data it is important to quantify the accuracy (agreement between measurement and actual value) and precision (variability of measurement) that should be expected, as well as the factors that affect the error, so that researchers and practitioners can apply this information to collected data. Previous research (e.g., Chin, Bishu, & Hallbeck, 1995; Cooper, Grimby, Jones, & Edwards, 1979; Krombholz, 1985) has investigated the magnitude of estimation error, but those studies focused on a narrow range of exertions that do not represent those performed in occupational settings.

There is some limited evidence that the precision and accuracy with which participants verbally estimate the magnitude of exertion may improve with training. Deeb (1999) investigated the effects of training on participants' ability to estimate the weight of objects. A significant improvement in accuracy was found when they first held weights of known magnitude prior to estimating the unknown weight. Based on these findings, it is hypothesized that the precision and accuracy of estimation will improve when a participant is systematically exposed to "physical benchmarks" to compare with the exertion being estimated. For example, if a participant is exposed to a maximum exertion prior to estimating a submaximal exertion, the estimate will be more accurate than if he or she had no prior exposure to the benchmark. The primary objective of this research was to examine this hypothesis by investigating the effect of training on the accuracy and precision with which participants verbally estimate the magnitude of exertion intensity for simulated tasks that vary with respect to the degrees of freedom needed to perform them.

A considerable amount of attention has been given to modeling the relationship between perceived and actual force production. Early research by Fechner (1860/1966) investigated the relationship between perceived and physical quantities, and from this came the concept of a "just noticeable difference" (JND), which is the incremental change in perception of the stimulus as its intensity is increased. According to the Weber-Fechner law, the JND is a geometric function of stimulus intensity, meaning the perceived sensation is a log function of the stimulus. However, this theory proved inadequate for very low and very high intensities. As a result, S. S. Stevens (1957) proposed a generic power law, in which the perceived quantity grows as a power function of the actual quantity.

Researchers have investigated the relationship between perceived and actual exertion, and their results have varied. Whereas several researchers (e.g., Gamberale, Ljungberg, Annwall, & Kilbom, 1987; Ljungberg, Gamberale, & Kilbom, 1982; J. C. Stevens & Mack, 1959) have supported the model proposed by S. S. Stevens (1957), finding exponents ranging from 1.4 to 2.4, others (e.g., Deeb, 1999; Krombholz, 1985) have found the relationship to be linear. Furthermore, other researchers (Chin et al., 1995; Cooper et al., 1979) have found that the relationship could be described equally well by a power or linear function, noting that the latter is a special case of the former. Still others (Weiss, 1981) have argued that no universal psychophysical law exists and that the relationship between perceived and actual quantities is determined by the specific context of the experiment. In light of the discrepancy in the literature, a secondary objective of this experiment was to evaluate the mathematical relationship between perceived and actual exertions.

METHODS

To achieve the stated objectives, we performed an experiment in which participants estimated the exertion intensity required to just budge a stationary handle on a work simulator. Verbal estimates were compared with the known resistance levels, which ranged between 0% and 100% of the participant's static, maximum voluntary contraction (%MVC). The experiment consisted of testing three participant groups: one control (no-benchmark) group and two other groups, which varied in the number of reference benchmarks to which they were exposed prior to estimating the intensity of submaximal exertions. Each group participated in two data collection phases, although the sequence of these phases varied by group. In the first phase, standardized maximum static strength measurements were collected. In the second phase, participants verbally estimated submaximal exertion intensities. Prior to exertion estimation, one group performed an additional phase in which they were exposed to two additional resistance benchmarks.

Equipment

Equipment consisted of a work simulator and a data acquisition system to record measurements. The simulator (Baltimore Therapeutic Equipment Company, 1992), shown in Figure 1, measures and controls force produced at the exercise head as participants simulate common hand activities, including using a screwdriver, turning a steering wheel, and many others. The work simulator measures static strength exertion and allows the experimenter to control the resistance of submaximal exertions. In knowing a participant's strength and the force setting of the simulator, the exertion intensity, in %MVC, is easily obtained. The advantage of using the simulator is that exertions closely resemble those performed in occupational settings (Powell et al., 1991).

Participants

A total of 48 participants (26 men, 22 women) were recruited for the experiment. Most (85%) were students, and their ages ranged from 19 to 34 years (mean = 22.8, *SD* = 3.4 years). Participants were randomly assigned to one of three experimental groups, each differing in the number of physical benchmarks experienced prior to the exertion estimation phase.

The *no-benchmark* (control) group immediately performed the exertion estimation phase and thus was exposed to no physical benchmark. Force settings were initially approximated, and strength measurement was performed after exertion estimation to determine what the exact %MVC levels were.

The *one-benchmark* group performed strength measurement prior to the exertion estimation phase, which provided a single benchmark of 100% MVC.

The *three-benchmark* group performed strength measurement and two submaximal exertions of 25% and 75% MVC prior to the exertion estimation phase, resulting in three benchmarks (25%, 75%, and 100% MVC).

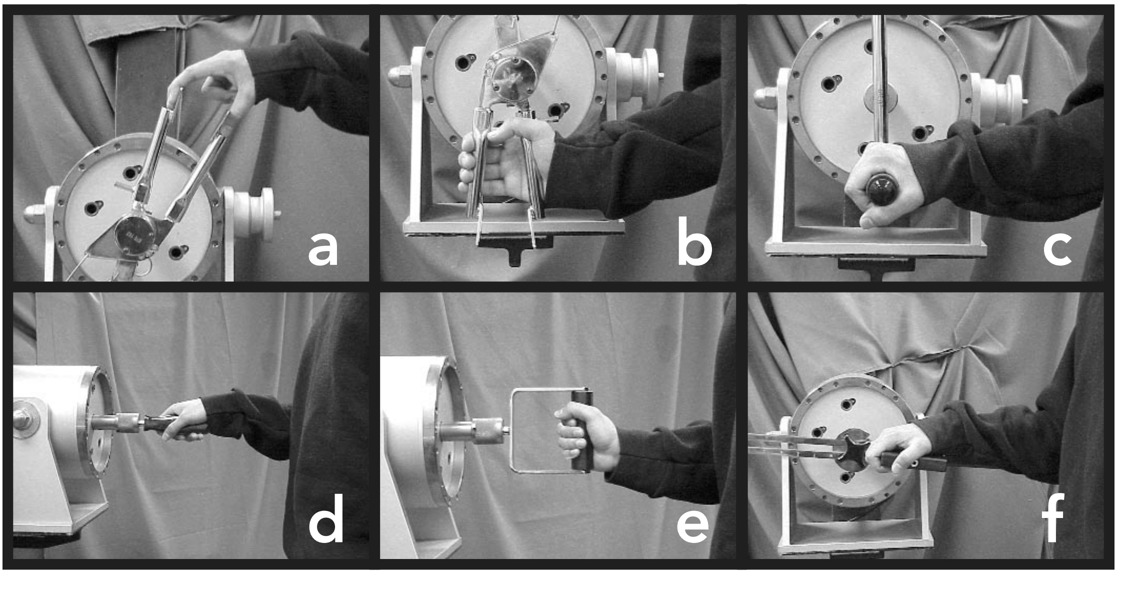

Figure 1. Some of the experimental tasks performed using the Baltimore Therapeutic Equipment Company work simulator: (a) pinch, (b) grip, (c) push-pull, (d) screwdriver, (e) pronation-supination, and (f) press. See text for a full description of all tasks.

Each group consisted of 16 participants and was balanced for age and gender. For each participant, the order of performing each task was random, and this order was maintained throughout each phase of the experiment.

Experimental Procedure

Participants were fitted to the work simulator by adjusting the height of the exercise head to the participant's standing elbow height. Participants performed the experiment using only their dominant hand. The no-benchmark group performed exertion estimation first, followed by strength measurement, whereas the one-benchmark group performed strength measurement, followed by exertion estimation. For the three-benchmark group, strength measurement was performed first, followed by an intermediate training phase in which participants experienced 25% and 75% of his or her maximum. Finally, exertion estimation was performed. Each phase will be described in further detail. For all groups, 30 min were given between phases to allow rest and to set up for the next phase.

Strength measurement. For each task, participants were instructed to "ramp up" from rest to maximum over a 2- to 3-s interval and to maintain the maximum for 2 s before releasing (Chaffin, 1975). Participants were instructed to use only their upper extremity and to not use body weight or lower extremity muscles to assist. For every exertion except the overhead press, the simulator was positioned so that the forearm was parallel to the ground and perpendicular to the upper arm. For the overhead press, the handle was adjacent to the participant's shoulder. Three peak exertion measurements were obtained for each task and averaged. Participants had 1 min of rest between replications and 2 min of rest between tasks. The order in which the tasks were presented was randomized for each participant in order to offset possible effects of fatigue, although fatigue was not measured.

Exertion estimation. For each of the 12 tasks, participants performed an exertion for each of three force levels, for a total of 36 trials. For the one- and three-benchmark groups, exertion levels were randomly selected (in increments of 5% MVC) for each level of low, medium, and high between 10% and 30%, 40% and 60%, and 70% and 90% MVC, respectively. The participant was never made aware of these levels. For the no-benchmark group strength was not initially measured, so it was necessary to approximate these force settings, and it was not known until after strength measurement what the true exertion level was. These approximations were based on strength data collected from pilot participants. Levels of force varied from task to task to avoid a potential learning effect, and the order in which the three levels were performed within each task was randomized. The tasks in this phase were performed in the same order as in the strength measurement phase. Participants performed as many repetitions as needed to assess the magnitude of the exertion (typically two to three repetitions) and were asked, "Please estimate, as a percentage of your maximum, the intensity of the exertion required to just move the attachment." Participants estimated the exertion required to initially budge the attachment (i.e., to change the system from static to dynamic).

Supplemental training. Between strength measurement and exertion estimation, the three-benchmark group was exposed to exertions of 25% and 75% MVC. For each of the simulated tasks, participants performed three repetitions and were told that the force required to move the attachment was either 25% or 75% MVC. Participants were instructed to perform four repetitions of the prescribed force level. The physical benchmarks in this phase were presented in the same order as in the other phases.

Task selection. To reflect the diversity of tasks performed in occupational settings and to evaluate the effect of using multiple and larger muscle groups on the accuracy and precision of exertion estimation, 12 simulated tasks were used, some of which are illustrated in Figure 1.

1. Pinch grip (pinch): The participant pinches paddles together using the thumb and index and middle fingers, using muscles in the hand and forearm.

2. Power grip (grip): The participant squeezes handles together using the thumb and all four fingers, using muscles in the hand and forearm.

3. Forearm supination (supination): The participant grasps a handle in a neutral forearm position and supinates the forearm, using muscles in the hand, forearm, and upper arm.

4. Forearm pronation (pronation): The participant grasps the handle in a neutral forearm position and pronates the forearm, using muscles in the hand, forearm, and upper arm.

5. Turning a key (key): The participant uses a lateral pinch grip to grasp the key and rotate the key clockwise, using muscles in the hand, forearm, and upper arm.

6. Large screwdriver (screwdriver): The participant grasps the handle of a screwdriver (oriented horizontally) and turns it clockwise, using muscles in the hand, forearm, and upper arm.

7. Large knob (knob): The participant grasps a knob using a claw grip and rotates the knob clockwise, using muscles in the hand, forearm, and upper arm.

8. Push (push): The participant grasps a handle with his or her forearm pronated and pushes the handle away from the body at waist level, using muscles in the hand, forearm, and upper arm.

9. Pull (pull): The participant grasps a handle with his or her forearm pronated and pulls the handle toward the body at approximately waist level, using muscles in the hand, forearm, and upper arm.

10. Triceps press down (press): The participant places his or her hand palm down on a paddle and extends the elbow to press the lever down, using muscles in the upper arm and shoulder.

11. Overhead shoulder press (overhead): The participant grasps a handle at shoulder height and presses the lever straight up, using muscles in the hand, forearm, upper arm, back, and shoulder.

12. Steering wheel (wheel): The participant grasps a wheel at the 3 and 9 o'clock positions and rotates the wheel clockwise, using muscles in the hand, forearm, upper arm, back, and shoulder.

These tasks are listed approximately from lowest to highest in terms of the degrees of freedom involved in performing the exertion. In his discussion of motion economy, Barnes (1980) classified hand motions according to how much of the upper extremity is involved. The classification ranges from "finger motions" at the lowest level to "motions involving fingers, wrist, forearm, upper arm, and shoulder" at the highest level. The same basic scheme applies to these tasks. The pinch and grip tasks, for example, are listed first because they utilize primarily the flexor muscles in the forearm. The screwdriver and knob tasks are in the middle of the list because the participant grips the attachment using muscles in the forearm and rotates the attachment, which requires use of muscles in the upper arm. The overhead and wheel tasks are listed last because they utilize not only the forearm and upper arm muscles but also the shoulder and back.

RESULTS

Effect of Training on Estimation Error

The main training effect was significant, $F(2, 45) = 34.32$, $p < .0001$, but several interaction effects were also statistically significant. First, a significant Training × Force interaction effect was observed, $F(4, 90) = 6.28$, $p < .0001$, and is illustrated in Figure 2. As represented by the letter coding, for the no-benchmark group a significantly higher average error was observed for the medium force level (19.4% MVC) than for both the high force (8.1% MVC) and low force (12.3% MVC). This was not the case for the two trained groups, which accounts for the interaction effect. In particular, no statistically significant difference was observed between the low and medium levels of force for either the one- or three-benchmark group.

The second significant interaction was Training × Task, $F(22, 494) = 2.31$, $p = .0007$. As there were 12 tasks, the interaction effects are difficult to interpret because of all the combinations of factors. Some tasks, such as supination and push, had mean values (14.5% and 14.1% MVC, respectively) in which there was no significant difference among training groups, whereas other tasks yielded significant differences among groups. Of all the Training × Task combinations, four produced means significantly larger than those in the other conditions ($p < .05$). Those conditions all involved the no-benchmark group and occurred for the grip, knob, overhead, and wheel tasks (20.1%, 22.1%, 25.8%, and 25.1% MVC, respectively).

Although these interaction effects were statistically significant, neither appears to heavily influence the main training effect. These effects reflect small differences in the magnitude of the main effect and are not of practical importance.

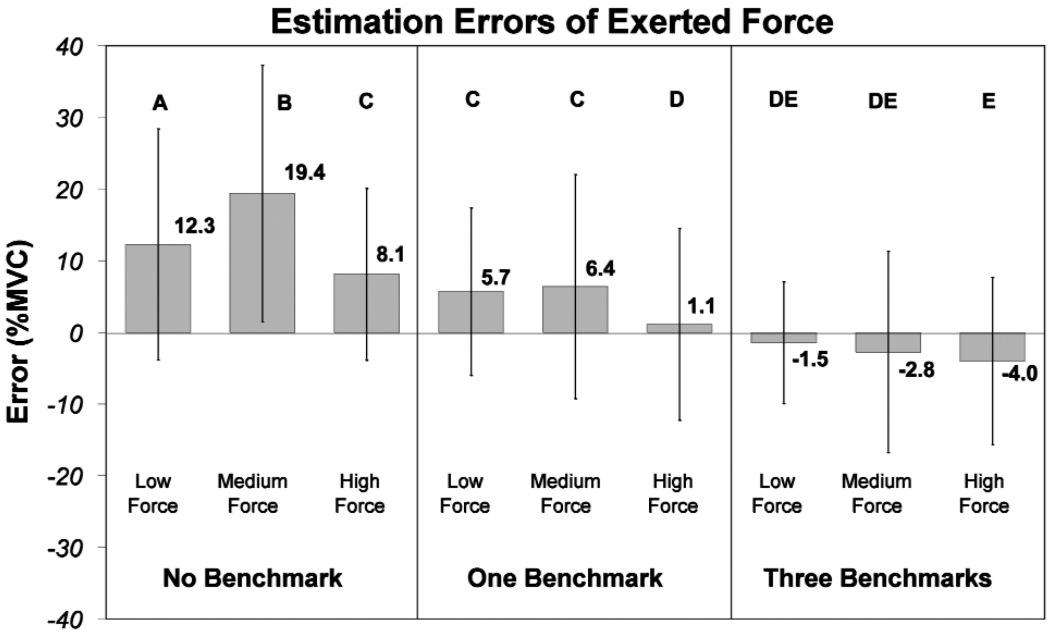

Figure 2. Mean and standard deviation of estimation error for all combinations of force and training. The letters represent significant differences ($p < .05$) among conditions.

In terms of the main effects of training, the no-benchmark group overestimated the actual force by 14.0% MVC. The one-benchmark participants still tended to overestimate the force, but to a lesser extent (4.5% MVC on average). Finally, the three-benchmark group tended to slightly underestimate the force (–2.8% MVC on average). The mean error from each training group was significantly different ($p < .0001$) from those of the other two (no benchmark > one benchmark > three benchmark).

The precision of force estimation also improved significantly when participants were exposed to physical benchmarks, as Bartlett's and Levene's tests each revealed a statistically significant ($p < .0001$) decrease in the standard deviation of estimation. The standard deviation decreased from 16.6% MVC for the no-benchmark group to 13.8% MVC for the one-benchmark group, and it further decreased to 11.6% MVC for the three-benchmark group.

Effect of Force on Estimation Error

The main effect of force was significant $F(2, 90) = 19.99$, $p < .0001$, but several interaction effects were also statistically significant. First, although the main effect of order was not significant, $F(2, 90) = 0.21$, $p = .81$, the Order × Force interaction effect was significant, $F(4, 179) = 4.01$, $p = .0038$, and is represented by the coding in Figure 3. For the low force, exertions performed first produced the highest error (6.8% MVC), whereas exertions performed third produced the lowest (4.5% MVC), although these differences were not statistically significant. The same trend was observed for medium force, and again, these differences were not statistically significant. However, a statistically significant order effect was observed for high force ($p < .01$), which accounts for the interaction effect. Specifically, estimates made third had a significantly higher error (3.5% MVC) as compared with those performed first (–2.1% MVC).

The Task × Force interaction effect was also significant, $F(22, 830) = 2.21$, $p = .001$. Again, these interaction effects are somewhat difficult to interpret, given all the combinations of factors. For some tasks differences among force levels were statistically significant, and for others there was no significant difference, which accounts for the interaction effect. Using grip

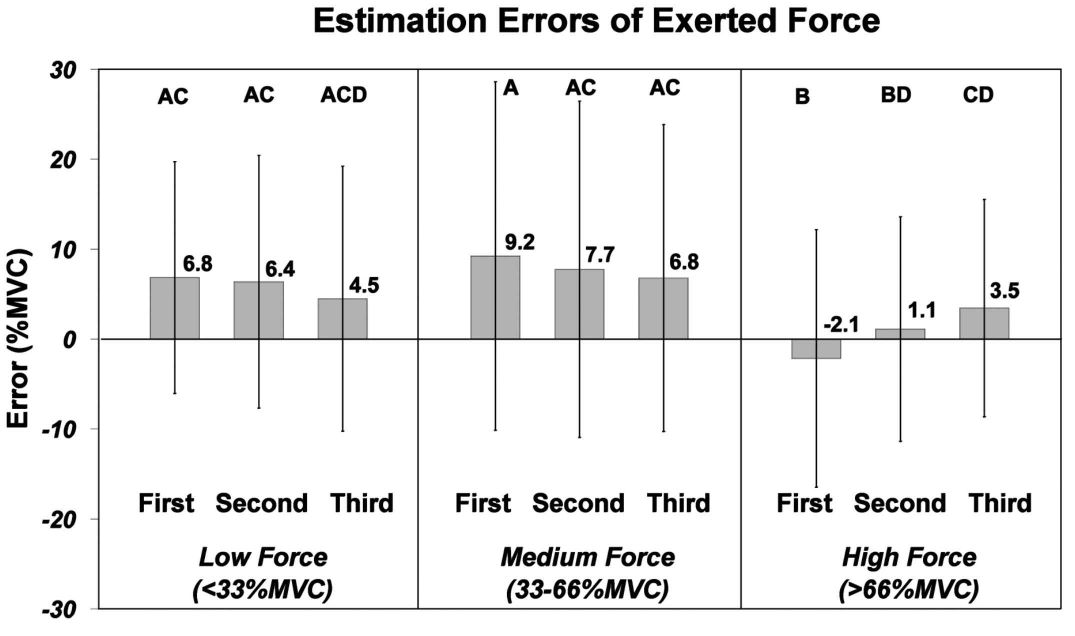

Figure 3. Mean and standard deviation of estimation error for all combinations of force and order. The letters represent significant differences ($p < .05$) among conditions.

as an example, the error for medium force was significantly larger than that for low ($p = .02$) and high ($p = .02$) force, although the difference between low and high force was not significant ($p = .55$). In contrast, for the knob and pinch tasks, no statistically significant difference was observed between the error of the medium force level and the error for low force.

Both the Order × Force and Task × Force interaction effects again appear to reflect small differences in the main force effect and are of little practical significance. In terms of the main effect of force, the overall average errors for low and medium force levels were 5.9% and 7.9% MVC, respectively, each of which was significantly greater ($p < .05$) than the average error for high force (0.6% MVC). Furthermore, the error for medium force was significantly larger than the error for low force ($p = .02$).

The force level also affected the precision of the estimate, as Bartlett's and Levene's tests each revealed that the standard deviations for both low (13.9% MVC) and high (13.3% MVC) force were significantly lower ($p < .0001$) than the standard deviation for medium force (18.3% MVC). There was no statistically significant difference between the standard deviations for low and high forces.

Effect of Task on Estimation Error

Figure 4 presents the mean and standard deviation of estimation error for each of the 12 tasks. The significant interaction effects involving task (Task × Force, Task × Training) have been described previously. The main task effect was statistically significant, $F(11, 494) = 18.91$, $p < .0001$, and these differences are represented by the letter coding in Figure 4. The tasks are listed from left to right in approximately ascending order of the degrees of freedom required to perform the exertion. In terms of main task effects, the overhead and wheel tasks resulted in average errors (13.0% and 12.3% MVC, respectively) that were significantly larger ($p < .01$) than the errors for 8 of the other 10 tasks, which otherwise ranged from –4.1% to 8.7% MVC.

Relationship Between Perceived and Actual Exertions

Because there is no consensus on the mathematical relationship between perceived and actual force, three regression models (linear,

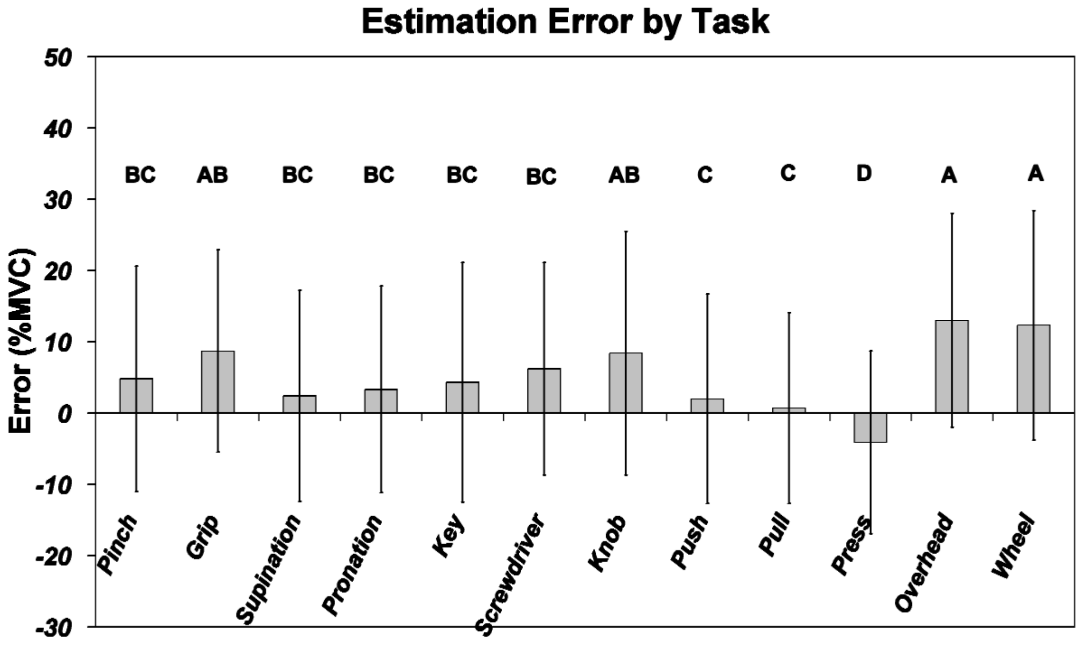

Figure 4. Mean and standard deviation of estimation error for each task (see text for descriptions). The letters represent statistically significant differences between mean values ($p < .05$).

power, and logarithmic) were prepared for each group of participants and are listed in Table 1. These models are based on data pooled across all tasks, force levels, and participants within each training group. All regression models were highly significant ($p < .0001$). With the exception of the no-benchmark group, the first-order (linear) regressions had the highest R^2 values of all three models evaluated. However, the range in individual R^2 values was quite small for all groups, and no tests were performed to determine whether the differences in R^2 values between models were statistically significant. The remainder of the analysis will focus on the

TABLE 1: Regression Equations Describing the Relationship Between Actual Force (x) and Estimated Force (y)

	Linear	Power	Logarithmic
No benchmark			
y	$0.96x + 15.9$	$2.40x^{0.84}$	$33.3\text{Ln}(x) – 60.9$
R^2	.67	.64	.68
$F(1, 574)$	1120.5	967.1	1188.6
p	<.0001	<.0001	<.0001
One benchmark			
y	$0.91x + 9.0$	$1.86x^{0.86}$	$33.9\text{Ln}(x) – 72.7$
R^2	.73	.69	.70
$F(1, 574)$	1509.9	1253.3	1290.9
p	<.0001	<.0001	<.0001
Three benchmarks			
y	$0.94x + 0.4$	$0.86x^{1.01}$	$34.7\text{Ln}(x) – 82.2$
R^2	.81	.76	.74
$F(1, 574)$	2466.5	1798.7	1636.5
p	<.0001	<.0001	<.0001

linear models because the difference in R^2 values is not of practical significance.

Individual regression lines for the 16 participants within each training group are shown in Figures 5a through 5c. For the no-benchmark group (Figure 5a), slopes ranged from 0.7 to 1.2 for individual participants (0.96 pooled data), intercepts ranged from –0.4 to 35.6 (15.9 pooled data), and all regression lines were located above the ideal line of $y = x$, indicating a strong tendency to overestimate the actual exertion level. For the one-benchmark group (Figure 5b), five individual regression lines fell below the ideal line. Slopes ranged from 0.6 to 1.1 (0.91 pooled data), and intercepts ranged from –5.4 to 25.8 (9.0 pooled data). Individual regression lines for the three-benchmark group (Figure 5c) were tightly clustered about the $y = x$ line. Slopes ranged from 0.7 to 1.1 (0.94 pooled data), and intercepts ranged from –9.5 to 5.8. The overall pooled intercept (0.4) was not significantly greater than zero.

DISCUSSION

Effects of Training on Estimation Error

This study found that the accuracy and precision with which participants estimate the magnitude of upper extremity forceful exertions improves by a statistically significant amount when participants are first exposed to physical benchmarks. These results enable practitioners and researchers to quantify error when using verbal estimation to assess forceful exertion, thereby allowing them to determine with some confidence the approximate true force, given a participant's estimate.

For example, suppose a worker estimates that the grip force required to engage a hand tool is 35% MVC. Assuming a naive participant (no benchmark), and ignoring interaction effects for a moment, the results of this study indicate that the actual force would be 21% MVC because participants, on average, overestimate the force by 14% MVC. Using a conservative confidence level of 75%, the range of expected actual force in this case would be 2% to 40% MVC. If the participant first produces a maximum grip prior to estimating, these results indicate that the estimate of 35% MVC corresponds to an actual force, on average, of 30% MVC, with a 75% confidence range of 15% to 46% MVC. Finally, if it is possible to expose the participant to the two additional benchmarks of 25% and 75% MVC, the estimate of 35% MVC would correspond to an average actual force of 38% MVC, with a range of 24% to 51% MVC.

In terms of estimation bias (overestimation/underestimation), the no-benchmark and one-benchmark participants tended to overestimate the intensity of exertions, and the three-benchmark participants tended to slightly underestimate the intensity. We are aware of no studies with which we could compare of these effects of training on estimation bias.

The use of physical benchmarks as reference points may easily be extended to occupational settings. In this study the work simulator was used for training both benchmark groups, but it was necessary only for the three-benchmark group. No special equipment is required to perform the maximum exertion used by the one-benchmark group. For example, to evaluate the exertion involved in squeezing the handle of a power tool, a worker would first squeeze the tool handle as hard as possible to obtain the maximal benchmark. Similarly, to evaluate the exertion required to pull a lever, one would have the worker first perform a maximum pull on a rigid object at approximately the same height and orientation as the actual lever. Essentially, the goal should be to simulate the task of interest and allow the worker to perform a maximum exertion prior to estimating the force of the actual task.

Effects of Force on Estimation Error

This study also demonstrates that estimation error is dependent on the magnitude of the test force, as indicated by Figure 2. In general, participants tended to produce larger and more variable estimation errors for the medium levels of force than for the high and low levels. A possible explanation for this is that for the medium force levels, there is more opportunity for error. Low and high force levels are constrained by 0% and 100% MVC, respectively, thereby limiting the extent to which a high force can be overestimated and a low force underestimated.

In terms of the estimation bias, participants tended to overestimate across all levels of force, although the average error for high force (0.6% MVC) was not statistically different from zero

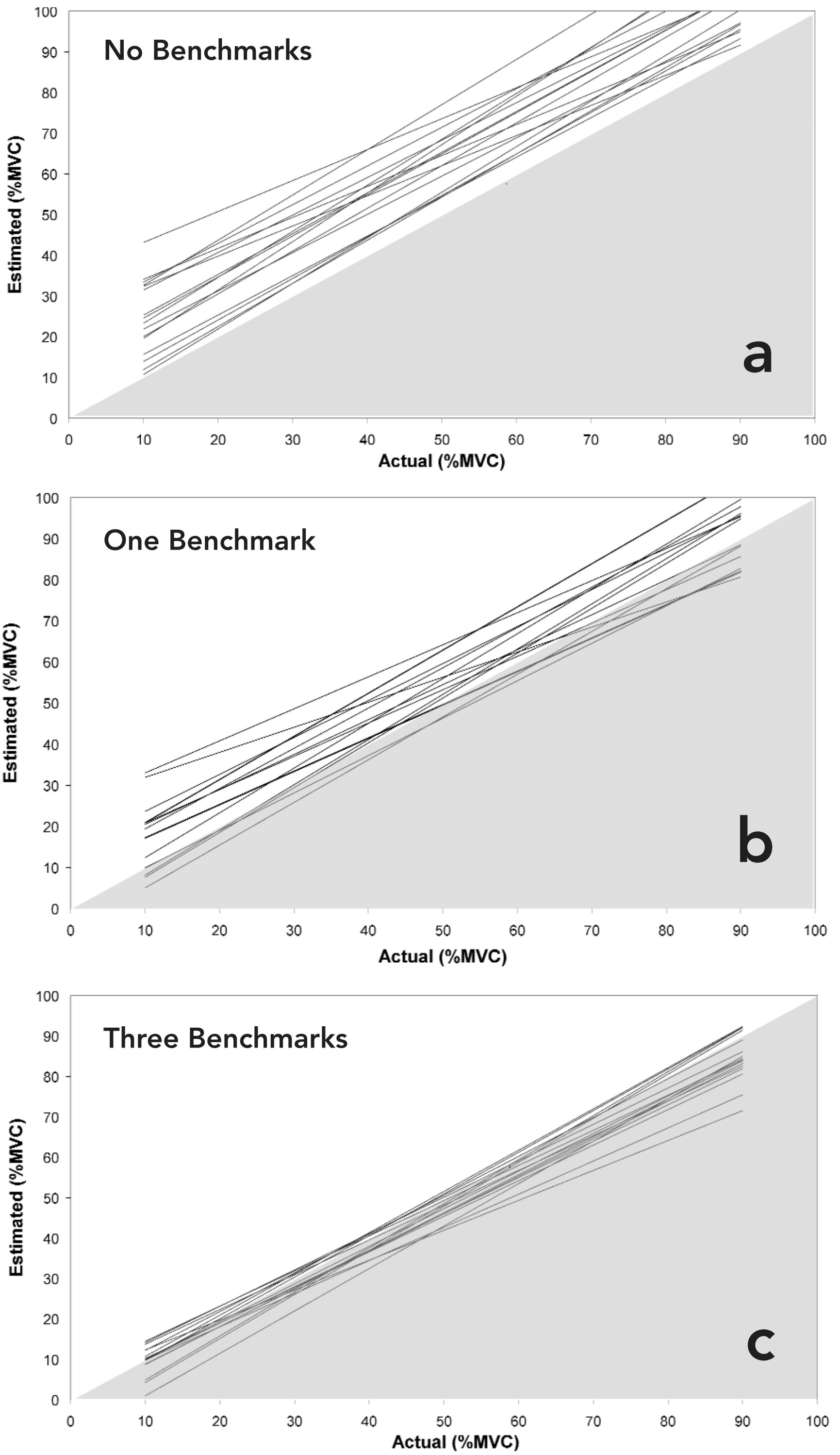

Figure 5. The relationship between actual and estimated exertion for each of the 16 participants in each training group. A perfect relationship is described by the line $y = x$, represented by the interface of the shaded and white regions. As the number of benchmarks increased from zero (a) to one (b) to three (c), the regression lines become more tightly clustered and centered about $y = x$.

($p < .05$). The literature has produced mixed results with respect to estimation bias and force level, but several studies support these findings. Among the studies that investigated magnitude estimation, Cooper et al. (1979) found that participants overestimated the actual exertion level by 10% to 20% (not %MVC) across all force levels. In a study that investigated the accuracy of magnitude production (participants were asked to produce a level of force rather than to estimate), Krombholz (1985) found results that were very similar to ours. Participants overestimated low levels of exertion (30%–50% MVC) by up to 100% and slightly underestimated (approximately 5%) high levels of exertion (90% MVC).

Likewise, Wiktorin, Selin, Ekenvall, Kilbom, and Alfredsson (1996) found that participants produced push-pull forces (in newtons) significantly larger than those that were requested. For very low levels of force (10–50 N), participants overproduced by as much as 700%. As the requested force increased, the overproduction decreased until, at a very high force (300 N), participants produced forces that were less than that requested, by up to 50%. In contrast to those reports and this study, Chin et al. (1995) found that participants consistently underestimated exertion intensity, with the underestimation increasing as the exertion level increased. At low levels of exertion (20% MVC), underestimation was less than 10%, increasing to approximately 20% at high levels of exertion (80% MVC). Although magnitude production elicits force perception differently than does magnitude estimation, Cooper et al. (1979) demonstrated that no statistically significant difference existed between the errors produced for the two techniques.

Effects of Task on Estimation Error

The tasks that produced the highest estimation error (the overhead press and steering wheel tasks) also involve the most and largest muscles of the upper extremity (Figure 4). Whereas the other tasks use predominantly the forearm muscles and/or upper arm muscles, the overhead press and steering wheel tasks additionally utilize muscles in the shoulder and back. As the number of joints and muscles increases, the participant integrates more information to estimate the exertion, and the higher errors observed here suggest that the participants were less successful in doing so.

Modeling Estimated Versus Actual Force

The results of this experiment show that a linear model describes the relationship between perceived and actual exertion as well as, or better than, a power or logarithmic function does (Table 1). Considerable disagreement exists in the literature concerning the mathematical relationship between perceived and actual exertion. Of the previous studies, those by Cooper et al. (1979) and Chin et al. (1995) are the most consistent with the results obtained here, in that they both concluded that the difference between a linear model and alternative regression models was not of practical significance.

Based on the results of this experiment and those of others described previously, the relationship between estimated and actual exertion appears to be affected by the technique with which perceived quantities are elicited. Whereas many studies have required participants to report a number corresponding to the intensity of the physical exertion, the type of reference on which these scales are based appears to affect the relationship between perceived and actual exertion. Several studies (Gamberale et al., 1987; Ljungberg et al., 1982; J. C. Stevens & Mack, 1959) asked participants to estimate force by specifying a number proportional (e.g., half or double) to a submaximal reference force. Stevens and Mack, for example, requested that participants "assign numbers proportional to the apparent magnitude" of a reference pull force of 17 pounds (~6.3 kg). In such a scenario, if the test exertion was 34 pounds (~12.7 kg), the correct estimation would be two, or double the reference. Almost invariably, this approach has yielded a power relationship between actual and perceived force.

In contrast, other studies, including this one, required participants to estimate force as a percentage of maximum using one of two approaches. Some studies (Chin et al., 1995; Deeb, 1999) have used the Borg CR-10 scale (Borg, 1990), which is a 10-point scale that uses simple verbal anchor points such as *moderate* and *extremely strong.* Similarly, other studies (Cooper et al., 1979; Krombholz, 1985), including this one, used a "percentage of maximum" (%MVC) scale, which is essentially a 100-point scale (0%–100% MVC). The %MVC scale used here is different from the Borg scale in that it is administered verbally, is based on a 100-point scale, and does not utilize descriptors as anchor points. In essence, a force estimate of 10% MVC is equivalent to a rating of 1 on the Borg CR-10 scale. What makes the two scales very similar is that both are anchored at the low end by zero and at the high end by "maximum possible exertion." Contrary to the studies that used an arbitrary reference, studies that have used the participant's maximum contraction as the reference have found the relationship between perceived and actual force to be linear.

An important distinction between these references (maximum vs. arbitrary reference) is the participant's familiarity (or lack thereof) with the reference stimulus. In both cases the participant compares the stimulus with an internal reference. In studies that use an arbitrary reference, the benchmark is typically set by the experimenter and may lack an inherent meaning. For example, J. C. Stevens and Mack (1959) used a reference pull force of 17 pounds (~6.3 kg), which, based on this experiment, corresponds to 71% and 45% MVC of average female and male strength, respectively. This magnitude of force (17 pounds) had no innate meaning to their participants because they had not been told what %MVC the force represented and thus had not been "calibrated." In estimating the force, therefore, the participant had to recall the magnitude of the reference stimulus, mentally "weigh" it with the test stimulus, and calculate the correct ratio of the reference to the test stimulus.

In contrast to the arbitrary reference, a benchmark of "maximum possible" is more innate to a person than is a quantity such as "17 pounds." Unlike the arbitrary reference force, which might stay fresh in the participant's mind for only a moment, the concept of strength is to some extent ingrained. Although a person's strength may change over time and is affected by transient factors such as fatigue, people are bound in their ability to produce force by their maximum strength, and they have the sensorimotor mechanisms with which to monitor how close an exertion is to this maximum, regardless of whether they first performed reference exertions.

Borg (1990) maintained that psychophysical units, like objective physical scales such as kilograms and meters, should be well defined, constant, and public. The familiarity of human participants with physical units of force such as pounds, kilograms, and newtons is highly variable. Research concerning the perception and judgment of force using these quantities has found a poor relationship between perceived and actual quantities. For example, Wiktorin et al. (1996) had participants estimate forces using kilograms and found, as compared with this study, a very weak relationship (R^2: .04–.48) between actual and self-reported push-pull exertions. Physical units of force are well defined and constant, but they appear to be less useful for subjective techniques such as psychophysics because humans are not calibrated or equipped to measure force in the same way that mechanical systems are. The "instrument" humans have to assess the magnitude of force consists of their body's sensorimotor feedback mechanisms, which are not calibrated with physical force units. Furthermore, in the context of studying and preventing work-related musculoskeletal disorders, units of %MVC offer an advantage over physical units of force because %MVC is normalized and provides an indication of how taxing the exertion is to the individual. A task requiring "75% of maximum" for a person provides more information about its demand than does a task specifying the pounds of force required.

Limitations

A factor that may affect these results and the use of magnitude estimation in the field is worker fatigue. In this experiment the potential effects of muscular fatigue were controlled through rest allowances and through randomization of the experimental trials. However, the application of these methods in the field would not take place in such a controlled environment. For example, workers may accumulate fatigue throughout the course of the workday. Fatigue affects the worker's strength and, therefore, the relative percentage (%MVC) a given exertion level represents. An exertion that represents 20% MVC at the start of a work shift may represent 30% MVC at the end of the shift because of diminished muscle strength. The effects of fatigue on the accuracy and precision of force estimation have not been studied extensively. J. C. Stevens and Cain (1970) found that an increase in the duration of a static exertion resulted in an increase in perceived force when, in fact, the force had not actually changed. For example, after 30 s of performing a continuous grip exertion, the participants' magnitude estimation nearly tripled regardless of the initial force magnitude. However, the study did not address the effects of fatigue that accumulate throughout the course of a workday.

This experiment also did not investigate the effect of the time lapse between exposure to the benchmarks and when the estimation trials took place. For the one- and three-benchmark groups, this delay was approximately 30 min, during which some of the effect of the reference exertions may have diminished. It is reasonable to assert that the sooner a participant estimates the exertion intensity after performing the reference exertion, the more accurate the estimation will be, as the reference will be fresh in the participant's mind. It is reasonable to expect an even higher level of accuracy than that obtained here if the maximal reference exertion immediately precedes the estimation. Conversely, a very long delay between the reference and the estimation would probably have a negative effect on accuracy.

CONCLUSION

A limitation to measuring exertion intensity has been the lack of methods to accurately, precisely, and efficiently quantify the exertion. Although the most appropriate technique to quantify exertion is largely dependent on the objectives of the study and the resources available, psychophysical methods such as magnitude estimation, the focus of this study, are often used. Such methods generally require less time to implement and may be done so at much lower cost, as compared with instrument-based techniques such as direct measurement and EMG.

This study investigated the precision and accuracy with which individuals could estimate the magnitude of submaximal, upper extremity forceful exertions over a range of tasks typical of occupational and industrial settings. The significance of these results is that they provide practical data for determining what an expected actual exertion is, given a worker's estimate. The study revealed that a statistically significant improvement in estimation can easily be obtained when the participant first produces a maximum exertion before estimating the magnitude of the force, a technique that would not be difficult to implement. This information should encourage the use of verbal estimation for assessing forceful exertion in epidemiological studies and other types of field-based research. Practitioners can also use the data from this research to evaluate the physical demand of tasks within their facility.

ACKNOWLEDGMENTS

This work was funded, in part, by the National Institute on Disability and Rehabilitation Research of the United States Department of Education under Grant #H133E980007. Matthew Marshall was supported by Grant T42-CCT510428-08 from the National Institute for Occupational Safety and Health.

REFERENCES

Baltimore Therapeutic Equipment Company. (1992). *Operator's manual.* Hanover, MD: Author.

Barnes, R. M. (1980). *Motion and time study: Design and measurement of work.* New York: Wiley.

Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian Journal of Work and Environmental Health, 16*(Suppl. 1), 55–58.

Chaffin, D. B. (1975). Ergonomics guide for the assessment of human strength. *American Industrial Hygiene Journal, 36,* 505–510.

Chin, A., Bishu, R. R., & Hallbeck, S. (1995). Psychophysical measures of exertion: Are they muscle group dependent? In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 694–698). Santa Monica, CA: Human Factors and Ergonomics Society.

Cooper, D. F., Grimby, G., Jones, D. A., & Edwards, R. H. T. (1979). Perception of effort in isometric and dynamic muscular contraction. *European Journal of Applied Physiology and Occupational Physiology, 41,* 181–185.

Deeb, J. M. (1999). Muscular fatigue and its effects on weight perception. *International Journal of Industrial Ergonomics, 24,* 223–233.

Fechner, G. (1966). *Elements of psychophysics* (D. H. Howes & E. C. Boring, Eds.) New York: Holt, Rinehart and Winston. (Original work published 1860)

Gamberale, F., Ljungberg, A. S., Annwall, G., & Kilbom, Å. (1987). An experimental evaluation of psychophysical criteria for repetitive lifting work. *Applied Ergonomics, 18,* 311–321.

Krombholz, H. (1985). On the association of effort and force of handgrip. *Perception and Motor Skills, 60,* 161–162.

Ljungberg, A. S., Gamberale, F., & Kilbom, Å. (1982). Horizontal lifting – Physiological and psychological responses. *Ergonomics, 25,* 741–757.

Powell, D. M., Zimmer, C. A., Antoine, M. M., Baruch, L. D., Bellian, K. T., Morgan, R. F., et al. (1991). Computer analysis of the performance of the BTE work simulator. *Journal of Burn Care and Rehabilitation, 12,* 250–256.

Sinclair, M. A. (1995). Subjective assessment. In J. R. Wilson & E. N. Corlett (Eds.), *Evaluation of human work* (pp. 69–99). London: Taylor & Francis.

Solomonow, M. (2000). Important factors in the recording, processing, and interpreting EMG. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 5.537–5.540). Santa Monica, CA: Human Factors and Ergonomics Society.

Stevens, J. C., & Cain, W. S. (1970). Effort in isometric muscular contractions related to force level and duration. *Perception and Psychophysics, 8,* 240–244.

Stevens, J. C., & Mack, J. D. (1959). Scales of apparent force. *Journal of Experimental Psychology, 58,* 405–413.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153–181.

Weiss, D. J. (1981). The impossible dream of Fechner and Stevens. *Perception, 10,* 431–434.

Wiktorin, C., Selin, K., Ekenvall, L., Kilbom, Å., & Alfredsson, L. (1996). Evaluation of perceived and self-reported manual forces exerted in occupational materials handling. *Applied Ergonomics, 27,* 231–239.

Matthew M. Marshall is an assistant professor of industrial and systems engineering at the Rochester Institute of Technology in Rochester, New York. He completed his Ph.D. in industrial and operations engineering at the University of Michigan in 2002.

Thomas J. Armstrong is director of the University of Michigan Center for Ergonomics and a professor of industrial and operations engineering and biomedical engineering at the University of Michigan, where he received his Ph.D. in industrial engineering physiology in 1976.

Marissa L. Ebersole is a doctoral student at the University of Michigan, where she received her M.S. in industrial and operations engineering in 2001.

Date received: March 21, 2002
Date accepted: March 22, 2004