Impact of task design on task performance and injury risk: case study of a simulated drilling task

Saad Alabdulkarim, Maury A. Nussbaum, Ehsan Rashedi, Sunwook Kim, Michael Agnew & Richard Gardner

To cite this article: Saad Alabdulkarim, Maury A. Nussbaum, Ehsan Rashedi, Sunwook Kim, Michael Agnew & Richard Gardner (2016): Impact of task design on task performance and injury risk: case study of a simulated drilling task, Ergonomics, DOI: 10.1080/00140139.2016.1217354

To link to this article: http://dx.doi.org/10.1080/00140139.2016.1217354

Accepted author version posted online: 26 Jul 2016. Published online: 31 Aug 2016.

Submit your article to this journal

Article views: 43

View related articles

View Crossmark data
Impact of task design on task performance and injury risk: case study of a simulated drilling task

Saad Alabdulkarim\textsuperscript{a,b}, Maury A. Nussbaum\textsuperscript{b}, Ehsan Rashedi\textsuperscript{b}, Sunwook Kim\textsuperscript{b}, Michael Agnew\textsuperscript{b} and Richard Gardner\textsuperscript{c}

\textsuperscript{a}Industrial Engineering Department, College of Engineering, King Saud University, Riyadh, Saudi Arabia; \textsuperscript{b}Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA; \textsuperscript{c}Advanced Ergonomics Technologies, Boeing Research & Technology, Everett, WA, USA

ABSTRACT

Existing evidence is limited regarding the influence of task design on performance and ergonomic risk, or the association between these two outcomes. In a controlled experiment, we constructed a mock fuselage to simulate a drilling task common in aircraft manufacturing, and examined the effect of three levels of workstation adjustability on performance as measured by productivity (e.g. fuselage completion time) and quality (e.g. fuselage defective holes), and ergonomic risk as quantified using two common methods (rapid upper limb assessment and the strain index). The primary finding was that both productivity and quality significantly improved with increased adjustability, yet this occurred only when that adjustability succeeded in reducing ergonomic risk. Supporting the inverse association between ergonomic risk and performance, the condition with highest adjustability created the lowest ergonomic risk and the best performance while there was not a substantial difference in ergonomic risk between the other two conditions, in which performance was also comparable.

Practitioner Summary: Findings of this study supported a causal relationship between task design and both ergonomic risk and performance, and that ergonomic risk and performance are inversely associated. While future work is needed under more realistic conditions and a broader population, these results may be useful for task (re)design and to help cost-justify some ergonomic interventions.

ARTICLE HISTORY

Received 15 April 2016
Accepted 18 July 2016

KEYWORDS

Ergonomic risk; performance; quality; adjustability; task design

Introduction

Work-related musculoskeletal disorders (WMSDs) reached approximately 365,000 cases in the United States in 2014, and the manufacturing sector, in contrast to most other sectors, had a 5% increase in WMSD cases compared to 2013 (BLS 2015). These disorders involve a substantial personal and societal burden (e.g. NRC 2001). WMSDs continue to occur despite an increasing understanding of the risk factors that initiate these disorders, and the reasons why such understanding has not consistently translated into a reduction in WMSDs are likely diverse and as yet not completely known. One likely reason is that operations management research is traditionally separated from that in physical ergonomics (Neumann and Dul 2010), along with increasing management disinterest in issues with ethical implications and a more exclusive focus on profitability (Walsh, Weber, and Margolis 2003). An important consequence of the former separation is the (mis)conception that ergonomics is strictly a health and safety tool that has no influence on the return of investment (Hägg 2003; Zare et al. 2016). Such a belief makes it more difficult to justify ergonomic interventions from a direct cost-savings standpoint.

WMSDs can have both tangible and intangible costs. While tangible costs, such those related to worker’s compensation costs and replacing (and training) injured workers, are relatively easy to track, it is more challenging to quantify intangible costs, such as reduced performance due to WMSD symptoms. Although the former may be easier to quantify, it can be misleading to measure the effect of (poor) ergonomic conditions based solely on these (e.g. the number of lost workdays). Musculoskeletal complaints also reduce productive time (Ricci et al. 2006), contribute to lower work ability (Neupane et al. 2011), and cause early retirement (van den Berg, Elders, and Burdorf 2010). Such intangible costs can be amplified by poor ergonomic conditions. In a strategy developed for the future of human factors/ergonomics discipline, Dul et al. (2012) indicated that common applications that do not highlight the dual influence of ergonomics on performance and human well-being, and in a broader sense do not use a holistic
approach in design, play a role in under-exploiting the potential benefits of ergonomics.

Ergonomic interventions, particularly those done proactively, typically require economic justification, yet there are factors that make this challenging. For example, a common focus is on injury-related outcomes, yet these tend to be lagging (not leading) indicators. Further, workers are variable in both their exposures and responses to given tasks. Focusing solely on potential injury reduction may also miss potential positive impacts of ergonomic interventions on worker performance (i.e. productivity and quality).

In a review of 250 case studies, Goggins, Spielholz, and Nothstein (2008) showed that positive effects, including increased productivity and quality and reduced turnover and absenteeism, are subsequent to ergonomic interventions across a variety of industrial sectors. In a recent review of the influence of ergonomics on quality, 12 studies were summarised as demonstrating a strong relationship between physical ergonomics risk and quality (Zare et al. 2016). However, it is important to note that many of these studies (e.g. Eklund 1995; Falck, Örtengren, and Högberg 2010; Falck, Örtengren, and Rosenqvist 2014; Fritzshe et al. 2014) were done to quantify the association between predefined risk groups (e.g. high, moderate and low risk) and quality metrics, for several tasks in targeted manufacturing plants. Given that task design variables, and the subsequent risk levels, were not ‘manipulated’ in the same task, such studies provide only limited support for causal relationships between task (re)design and both performance improvement and risk reduction.

Other studies (e.g. González, Adenso-Díaz, and González Torre 2003; Yeow and Sen 2006; Das, Shikdar, and Winters 2007) involved ‘interventions’, in which risk levels were manipulated through some modifications to a task. These studies, however, typically included only two levels of ergonomic risk and associated quality outcomes (i.e. before and after an intervention). While providing stronger support for the noted causality, these studies also provide incomplete support for ergonomic interventions, for two reasons. First, two-level studies do not necessarily capture the shape (or, functional form) of the relationship between task design factors (e.g. posture) or ergonomic risk and performance. Understanding such relationships would be valuable for purposes of intervention justification, particularly for tasks involving low or moderate levels of risks (for which existing evidence is more limited in the context of worker performance). Second, in many of the noted studies the intervention substantially changed the nature of the task, often such that human involvement was minimised (e.g. adding a conveyor instead of a worker manually handling boxes). In these cases, performance improvement could not clearly be tied to reduced task-related ergonomic risk.

Despite these limitations, however, existing evidence does support at least an association between an ergonomic intervention and a subsequent positive effect on both ergonomic risk and performance (i.e. productivity and quality). From a proactive perspective, such relationships are of value, as they can facilitate cost/benefit analyses of an intervention prior to any capital investments. The purposes of the current study were twofold, and were addressed within the context of a specific repetitive drilling task that is common in aircraft manufacturing and was simulated in a laboratory environment. First, to evaluate if causal relationships existed between workstation adjustability, as a task variable amendable to (re)design, and both performance and ergonomic risk. Second, to assess whether ergonomic risk was associated with performance. We hypothesised that the addition of workstation adjustability will improve performance and reduce ergonomic risk and that performance and risk will be inversely associated.

Methods

Participants

A total of 34 individuals (18 males, 16 females) were recruited from the local community and University to participate in the study. All completed an informed consent procedure approved by the Virginia Tech Institutional Review Board. Of these, 16 (1 male, 15 females) were removed during an initial screening session, as they could not consistently perform all of the experimental tasks (i.e. unable to perform the higher drilling force condition as described in the subsequent section). The remaining 18 participants (17 males, 1 female) completed the study. All but one male participant reported being right handed, and all reported no recent (past year) or current musculoskeletal problems and having normal or corrected-to-normal vision. Their mean (SD) age, stature and body mass were 23.4 (2.6) years, 176.5 (7.8) cm and 78.2 (17.5) kg, respectively.

Task description

A repetitive drilling task was simulated in the laboratory, and was chosen because it approximates a realistic manufacturing task, is repetitive, involves both precision and strength demands, and it was feasible to implement different levels of adjustability. A mock cylindrical fuselage was constructed for use in the study (Figure 1(a)), the configuration of which was considered representative of small aircraft. An 80-cm-diameter disc was at either end of the cylinder, and these were connected by six rungs to simulate longitudinal stringers (5.08 x 5.08-cm hollow rectangular stock, 121.9 cm long) spaced evenly around the circumference of the discs. During the experiment, the
initial fuselage configuration was such that there were two rungs each in the ‘upper’, ‘middle’ and ‘lower’ positions, and two rails at the height of the central axis of the cylinder. Each rung had six evenly space holes that were 1.3 cm in diameter, for a total of 36 holes. The structure was covered with heavy canvas fabric, to simulate the fuselage skin, with cut-outs at each hole. The fuselage was attached to vertical structures that allowed for changes in height and rotation about the central axis.

A commercial pneumatic drill (Figure 1(b)) was employed (mass = 1.10 kg; model #A2801447, Atlas Copco, Hungary). A simulated drill bit was used (Figure 1(c)), with four components. At the base, a hexagonal portion was ‘chucked’ in the drill, and connected to a uniaxial load cell (Interface, SML-100, Scottsdale, AZ). The load cell, in turn, was connected to a ‘probe’, consisting of sequential portions of steel and nylon with respective lengths of 3.3 and 2.5 cm. The diameter of the probe was 1 cm, and was used to mimic the quality requirements from an aircraft manufacturer’s specifications for drilling fuselage fastener holes. Specifically, drilling for subsequent riveting requires that holes be within an angularity tolerance of ±2° of normal to the surface. Holes drilled outside this tolerance range are considered errors or defects as excessive hole angularity adversely impacts rivet quality. Given the diameters of the rung holes, the probe diameter and the rung dimensions, any deviation beyond 2° of normal led to contact between the probe and a rung, and was captured as an ‘error’ as described below.

The experimental task involved completing a sequence of discrete simulated drilling actions. Each action involved three steps. The probe was first inserted into a hole and centred. Participants then needed to generate sufficient force for a set duration to ‘complete’ a hole. Exertions were isometric because the probe did not make any displacement through the rung. Finally, the probe was removed and the participant moved to the next hole. Two target forces used were 66.7 N and 111.2 N. These levels were estimated among expert workers at an aircraft manufacturer, and represent forces required in practice to complete ‘pilot’ and ‘full size’ drilling in a typical aluminium fuselage skin structure. Pilot drilling was completed first, for all holes in the fuselage, followed by final drilling after a rest break (~10 min). The same sequence of holes was completed throughout the experiment, and by all participants, with all holes numbered to facilitate this.

The accumulated duration for a complete hole was set to 2.5 s. Forces being generated needed to exceed the relevant target for this duration, and were monitored using the load cell (at 1000 Hz). When forces were below the target, the duration of this did not count towards the 2.5 s required total for a hole. No maximal force level was specified. For each hole, participants were provided with three auditory feedback types indicating when the generated force was above the target (tone 1) and when the 2.5 s duration was completed (tone 2). Feedback was also generated when an ‘error’ was made (tone 3), consisting of contact of the probe with a rung. Both the probe and the fuselage were wired together; contact closed the circuit and this was used within the data collection system to generate the error tone when the noted contact was made. A custom Labview™ program (National Instruments, Austin, TX, USA) was developed to manage the experiment.

**Independent variable**

A single independent variable was manipulated in the experiment, which was the extent of ‘workstation adjustability’. This variable and the levels used were selected partly...
because this variable is feasibly modifiable in actual settings. Three adjustability conditions were created:

- No adjustability (None): the midline of the fuselage was set at mean elbow height, or 108.2 cm (Marras and Kim 1993). Participants were allowed to choose any posture in this (and any) condition; they often squatted, or sat on the floor, for the lower rungs in this condition (Figure 2(d) and (e)).
- Some adjustability (Some): the midline of fuselage was set at mean shoulder height, or 144.0 cm (Marras and Kim 1993). This condition made it easier to access the lower rungs, and a two-step stool was available on each side of the fuselage to aid in reaching the higher rungs. When completing the lower rungs, many participants sat on the stool (Figure 2(a)–(c)).
- High adjustability (High): the fuselage height was set to participant preference, following initial practice. The mean (SD) of chosen heights was 69.4 (5.5) % of individual stature, or 14.2 (9.9) cm above the noted mean elbow height. In this condition, the fuselage could also be rotated by the participant about the long axis. The fuselage was locked to prevent rotation during drilling. All participants rotated the fuselage for each rung except the first.

### Procedures and data collection

Participants completed both training and testing sessions, separated by at least two days to minimise any effect from residual muscle fatigue. In the training session (~2.5 h), participants were asked to practice drilling the entire fuselage at least 10 times. Practice was performed for the three conditions and at the two force levels. More practice was provided as needed, until participants reported being comfortable and proficient with the procedures. Participants were also instructed during practice (and again during the testing session) to work as quickly but as accurately as possible. Participants were allowed to use one or both hands during the tasks, though not allowed

![Figure 2. Examples of adopted postures for the Some (a–c) and None (d, e) conditions.](image-url)
to grab/hold the probe during drilling, and freely chose their postures. When practising the Some condition, participants were instructed to make sure they know when and how to move the stools.

In the testing session (~2 h), participants were reminded about the procedures and completed additional practice (~15 min). The order of exposure to the three adjustability conditions was completely counterbalanced. Rest breaks of at least 10 min were provided after completing each condition. Generated forces were collected (at 1000 Hz) throughout the testing sessions, and participants’ postures were monitored (at 60 Hz) using a wearable inertial motion capture system (MVN, Xsens technologies B.V., Enschede, the Netherlands).

**Dependent measures**

To evaluate productivity and quality, several measures were obtained. Productivity was quantified by two metrics, both based on the time series of recorded probe forces. First, fuselage completion time (FCT) defined as the time between the first hole to ending the last hole in the fuselage. Second, the time between rungs (TBR), or the time spent transitioning between rungs, and calculated by subtracting rung completion times from a FCT, for a given fuselage. Quality was assessed using two metrics. First, the number of defective holes for the fuselage (FDH) was obtained. A hole was considered defective if at least one error was made. Given that defective holes can (and did) have drastically different numbers of errors, the total number of errors for the fuselage (FTE) was obtained as a second quality metric.

Ergonomic risk can be assessed using diverse approaches that can be broadly classified into three categories. They are, in order of increasing precision of the collected data and invasiveness, self-reports, observational methods and direct measurements (David 2005). Despite the relative ease of use, self-reports (e.g. questionnaires and interviews) can have low validity (Wiktorin, Karlqvist, and Winkel 1993) and reliability (Burdorf and Laan 1991). In observational methods, workers are observed and a systematic approach used to classify pre-specified risk factors (e.g. posture or force). These approaches are different in the factors they evaluate, and different in the validity and intra-observer and inter-observer repeatability (for recent reviews on observational methods, see NIOSH [2014] and Takala et al. [2010]). Direct measurements require more expertise and appropriate technology, yet generally yield more accurate and valid measures.

The purpose of an investigation and the nature of the evaluated work determine the appropriate assessment method(s). Here, two observational methods were used, the rapid upper limb assessment (RULA; McAtamney and Nigel Corlett 1993) and the strain index (SI; Moore and Garg 1995). Direct measures were used, where possible (i.e. direct measures of posture using the inertial motion capture system). Because the two methods differ in their approach and emphasis (Drinkaus et al. 2002), both were used to provide a broader assessment. Two cross-sectional field studies have shown an association between higher RULA scores and self-reported musculoskeletal disorders (Shuval and Donchin 2005; Breen et al. 2007). Also, two laboratory studies found an association between RULA scores and discomfort (McAtamney and Nigel Corlett 1993; Fountain 2003). The inter-observer repeatability of RULA was found to be good (McAtamney and Nigel Corlett 1993), though there is insufficient evidence concerning its intra-observer repeatability (Takala et al. 2010). Several retrospective studies have clearly demonstrated an association between SI scores and upper limb disorders (Moore and Garg 1995; Moore, Rucker, and Knox 2001; Rucker and Moore 2002; Spielholz et al. 2008). The inter-observer and intra-observer repeatability of SI have been reported to be moderate to good (Stevens et al. 2004; Stephens et al. 2006).

RULA and SI risk scores at both the rung level and fuselage levels were obtained as estimates of ergonomic risk (note, fuselage-level risk was the sum of risk scores at the rung level). The following discussion is a brief summary of the two methods, and the reader is referred to the original papers cited above for more complete details.

RULA was developed to evaluate musculoskeletal loads resulting from posture, forces exerted and muscle use. It yields a ‘grand score’ from combining estimated risk in two body regions, specifically neck/trunk/legs (score D, abbreviated as NTLS) and upper arm/lower arm/wrist (score C, abbreviated as ULS standing for upper limb score). For each body part, there are different risk classifications. For example, a risk value of 4 is added to the upper arm score when the upper arm is flexed more than 90° from neutral. And, this risk is increased by one when the upper arm is abducted. ULS and NTLS are achieved using tables that combine the risk values for the specified body parts and by adding the muscle use and force scores (discussed below). Because this method evaluates each side of the body separately, we analysed the side that was used in the task.

In the original RULA, there are risk factors described but often with little specification. For example, the method requires increasing the risk on the upper arm if it was abducted, without specifying a threshold for the magnitude (e.g. degrees) of abduction. We attempted to increase the specificity of the method by adopting the following missing thresholds:

1. Wrist ulnar and radial deviations were considered ‘risky’ when they exceed 14.5° and 21.8°, respectively, from neutral. An earlier study indicated that these angles were set to protect 75%
of some healthy participants from reaching a critical pressure (30 mmHg) for nerve injury in the carpal tunnel (Keir et al. 2007).

(2) Wrist pronation and supination were considered ‘risky’ when the deviation exceeded 45° in either direction from neutral. This was also used because carpal tunnel pressure exceeds the noted critical pressure beyond these thresholds (Rempel et al. 1998).

(3) In a prospective cohort study of electronics assembly workers, it was found that the percentage of a work cycle in which the shoulder is abducted more than 30° was an indicator for symptoms of cervicobrachial disorders (Kilbom and Persson 1987). Therefore, we considered abduction ‘risky’ when this threshold was exceeded.

Concerning the muscle use and force scores, a score of 1 needs to be added to both ULS and/or NTLS when the action was repeated more than four times per minute to account for risk from repetition (McAtamney and Nigel Corlett 1993). This was the case for both ULS and NTLS in the simulated task, because both body regions repeated actions more than the specified threshold. For force scores, respective scores of 1 and 3 were added to the ULS for the 66.7 and 111.2 N conditions, to include risk from forceful actions, and these particular scores were consistent with descriptions of the RULA method.

The SI is a semi-quantitative method that was developed from principles of biomechanics, physiology and epidemiology, and focuses on the risk of distal upper extremity (elbow, forearm, wrist and hand) disorders. Demands of an analysed task are evaluated by six task variables: intensity of exertion, duration of exertion, efforts per minute, hand/wrist posture, speed of work and duration of task per day. Each of the six variables is assigned a multiplier based on the associated risk. The SI ‘grand risk’ score results from multiplying the six multipliers. Intensity of exertion is the most influential variable.

Participants rated physical effort in the hand/wrist area at each of the three rung levels after each trial using the Borg 10-point scale (Borg 1990). The intensity multiplier was derived from these ratings, as described in the original SI paper (Moore and Garg 1995). For the duration of exertion, the multiplier was derived based on the mean duty cycle in each rung (duty cycle = exertion duration divided by cycle time). It should be noted that participants varied here in their response speed to the completion tone. Exertion duration extends as long as the target force level was exceeded. The efforts per minute (or frequency) multiplier were assigned based on the mean cycle time in each rung. The hand/wrist posture (of the dominant side that was used in the task) multiplier was assigned according to ranges specified in the method for wrist extension/flexion angle and ulnar deviation angle. These angles were obtained from the motion capture system. Based on the original SI method description, the task speed of work was not considered to be a risk-amplifying factor. For the duration per day variable, we assumed the task was performed for a consistent 4 h per day.

When evaluating tasks with a variable physical exposure, the authors of the two methods proposed assessing ‘peak’ stresses, as they are potentially responsible for exceeding injury thresholds. Based on this, we evaluated the 90th percentile angles within each rung for RULA variables and the hand/wrist posture variable in SI; this approach avoided peak values which were likely influenced by noise.

Perceptual measures were obtained using a questionnaire administered after each condition was completed. Global ratings of perceived exertion (RPE) were obtained using the 20-point Borg scale (Borg 1982), and participants were asked to use this scale to provide an overall evaluation of the condition just completed. Ratings of perceived discomfort (RPD) experienced in the hand/wrist, upper arm and shoulder were also obtained, using Borg’s 10-point scale (Borg 1990). To help conceptualise the latter scale, as well as to normalise ratings over the full range, participants performed a static endurance task in the training session, and provided RPDs every 5 s until they reach maximum discomfort (Sood, Nussbaum, and Hager 2007; Rashedi et al. 2014). Using the same scale, participants rated the level of physical effort required in hand/wrist area for each of the three rung levels (high, middle and low).

After completing a fuselage, participants were asked to rate the extent to which they believed that the condition they just completed (i.e. the adjustability and force levels) supported their efficiency (ratings of perceived efficiency, RPEF) and accuracy (ratings of perceived accuracy, RPA). These ratings were done using two 10-cm visual analogue scales (VASs). RPEF and RPA were obtained as the location (in cm) of the lines drawn between the left (minimal extent) and right (labelled: maximal extent) verbal anchors. These ratings were used to explore the relationship between perceived and actual performance in the examined conditions. Note, acronyms and their meanings are summarised in Table 1.

**Statistical analyses**

Separate $2 \times 3$ repeated measures analyses of variance (ANOVA) were used for each dependent variable to assess the effects of drilling force and adjustability. Presentation order for the latter was included as a blocking variable. All statistical tests were considered significant when $p < 0.05$. Post hoc comparisons were performed using
Tukey’s HSD test. Partial $\eta^2$ was used to determine effect sizes. To meet parametric model assumptions, appropriate transformations of the dependent variables were used. The association between ergonomic risk and performance was assessed both across and within conditions. For the former, simple linear regression models were developed to assess the correspondence between each performance metric (e.g. FCT) and each risk metric (e.g. RULA score). Coefficients of determination ($r^2$) were used to evaluate the level of correspondence. Within conditions, a qualitative comparison was done between mean performance and mean ergonomic risk in each condition (across the two force levels).

**Results**

Statistical results are summarised in Table 2, and subsections below expand on these.

**Time-related metrics**

FCT was affected by Condition and Force, with no significant interactive effects (Figure 3(a)). High adjustability led to the shortest FCTs, with the longest times required in the Some condition. The higher Force condition led to longer FCTs, and this effect was fairly consistent between conditions (i.e. no interaction effect). A similar pattern of results was found for TBR, but without a significant effect of Force (Figure 3(b)).

**Quality-related metrics**

Both the total number of defective holes (FDH) and errors (FTE) significantly differed between conditions, but were not affected by Force or the interaction (Figure 3(c) and (d)). Both FDH and FTE were lowest in the High condition.

**Risk metrics**

For RULA scores, there was a significant Condition × Force interaction effect. RULA scores were lowest in the High condition and with lower force, with the influence of force being most substantial in the High vs. the other two conditions (Figure 4(a)). SI scores differed between Conditions and Force levels (Figure 4(b)). SI scores were lowest in the High condition and lower at the 66.7 N level.

**Perceived exertion and discomfort**

All measures of perceived exertion and discomfort were significantly affected only by the main effects of Conditions and Force, with no significant interactive effects. Each was smallest in the High condition and smaller at the lower force (Figure 5).

**Relationships between risk and performance**

Results from fitting simple regression models to the risk and performance metrics are summarised in Table 3. Figure 6 shows graphical representations of the significant models for RULA and SI risk metrics.

Graphical representations of the relationships between mean performance and RULA and SI scores (across the two force levels) are provided in Figures 7 and 8, respectively.

**Perceived versus measured performance**

Both RPA and RPEF were largest in the High condition, followed by None and Some, and both were significantly higher when exerting 66.7 N vs. 111.2 N. Simple linear
The purposes of this study were to (1) evaluate if causal relationships exist between task design, specifically work-station adjustability here, and both performance and ergonomic risk in the simulated drilling task; and (2) to investigate the association between these outcomes (i.e. regression models of the relationship between measured and perceived performance are summarised in Table 4 and illustrated in Figure 9. While there were significant relationships between perceived and measured performance, the level of correspondence was relatively low.

**Discussion**

The purposes of this study were to (1) evaluate if causal relationships exist between task design, specifically work-station adjustability here, and both performance and ergonomic risk in the simulated drilling task; and (2) to investigate the association between these outcomes (i.e. **Figure 3.** Time-related (a and b) and quality-related performance results (c and d).

Notes: Light and dark bars indicate the 66.7 and 111.2 N force levels, respectively. Capital letters indicate groupings obtained from pairwise comparisons between adjustability conditions, and error bars indicate 95% confidence intervals.

**Figure 4.** RULA (a) and strain index (b) risk results.

Notes: Light and dark bars indicate the 66.7 and 111.2 N force levels, respectively. Capital letters indicate groupings obtained from pairwise comparisons between adjustability conditions, and error bars indicate 95% confidence intervals.
have a significant influence on either of the quality metrics (although there was evidence to support a possible deterioration in FDH between the Some and None conditions).

There are several sources that may have contributed to the reduced productivity with the Some adjustability. In particular, participants were required to manually move the stools whenever and wherever they decided to use them. That movement time included time spent optimising the stool's location (i.e. horizontal and vertical distance in relation to the fuselage). Also, participants sometimes needed to re-locate the stool after attempting to drill a hole. While we attempted to minimise such effects, through practice and reminders to focus on stool placement, it might have been more effective to more precisely guide stool movement through, for example, lines on the workstation floor.

Figure 5. Ratings of perceived exertion (RPE: a) and discomfort (RPD: b–d). Light and dark bars indicate the 66.7 and 111.2 N force levels, respectively.
Notes: Capital letters indicate groupings obtained from pairwise comparisons between adjustability conditions, and error bars indicate 95% confidence intervals.

Table 3. p Values, $r^2$, and slopes of fitted lines for the simple linear regression models of risk-performance relationships.

<table>
<thead>
<tr>
<th></th>
<th>RULA</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p Value ($r^2$)</td>
<td>Slope (95% CI)</td>
</tr>
<tr>
<td>FCT</td>
<td>&lt;0.0001 (0.251)</td>
<td>0.0236 (0.0157–0.0314)</td>
</tr>
<tr>
<td>TBR</td>
<td>0.0004 (0.113)</td>
<td>0.0.0196 (0.0090–0.0301)</td>
</tr>
<tr>
<td>FDH</td>
<td>0.0095 (0.062)</td>
<td>0.3739 (0.0934–0.6543)</td>
</tr>
<tr>
<td>FTE</td>
<td>0.0211 (0.049)</td>
<td>0.1696 (0.0259–0.3132)</td>
</tr>
</tbody>
</table>

Notes: Bold font indicates significant slopes ($p < 0.05$). Note, data from all three conditions.

ergonomic risk and performance). As discussed below, the primary finding from this study was that both productivity and quality significantly improved when increasing adjustability, but only when that adjustability succeeded in reducing ergonomic risk.

Regarding the relationship between task design and performance, it was hypothesised that increasing workstation adjustability will improve both productivity and quality. However, the results suggest that the influence of increasing workstation adjustability depends on the level of that added adjustability. The use of High adjustability significantly and substantially improved both quality and productivity. Yet, despite the expectation that Some adjustability would reduce postural extremes, the Some condition reduced productivity (both metrics) and did not have a significant influence on either of the quality metrics (although there was evidence to support a possible deterioration in FDH between the Some and None conditions). There are several sources that may have contributed to the reduced productivity with the Some adjustability. In particular, participants were required to manually move the stools whenever and wherever they decided to use them. That movement time included time spent optimising the stool’s location (i.e. horizontal and vertical distance in relation to the fuselage). Also, participants sometimes needed to re-locate the stool after attempting to drill a hole. While we attempted to minimise such effects, through practice and reminders to focus on stool placement, it might have been more effective to more precisely guide stool movement through, for example, lines on the workstation floor.
In addition, several participants felt less stable on the stool when completing the upper rungs (Figure 2(c)) and that this instability was a hindrance to performance in that condition.

Quality appeared to be independent of Force, though completion time increased with Force. Supporting the former, Finneran and O’Sullivan (2013) found that force did not affect precision performance, unlike posture, in a precision task (Fitts-type tapping task). In a repetitive gripping task, increasing force was found to reduce self-selected duty cycle (Finneran and O’Sullivan 2010, 2014), consistent with the noted main effect of Force on FCT found here.
not sufficiently sensitive to detect the (small) magnitude of differences in ergonomic risk between the Some and None conditions. Differences in exposure in the High condition were likely detected because the impacts were more substantial.

Despite the noted lack of differences in ergonomic risk between the Some and None conditions, there is evidence that adjustability in the Some condition improved certain risk factors pertaining to distal upper extremity, as quantified by the SI. In particular, the Some condition reduced the specific risk (SI multiplier) related to hand/wrist posture, repetition and duration of exertion by 6.2, 8.6 and 2.4%, respectively. In contrast, risk related to the intensity of exertion in the hand/wrist increased by ~25% compared to the None condition. The former improvements were offset by the latter increase in risk, since the SI method is most influenced by the intensity of exertions, as discussed earlier. Given that this critical variable was measured subjectively, though, it is possible that participants overestimated their perceived exertion because of, for example, the repeated need to move the stool and/or the feeling of instability when using the stools. Note that the intensity of exertion captures the proportion of maximum capacity (strength) required to perform a task (Moore and Garg 1995). It was quantified subjectively here, because use of objective methods, such as directly measuring muscles activation levels through electromyography, was considered infeasible in the current study. Bao et al. (2006), however, found a poor correlation between self-reports, ergonomist-estimates and direct measurements (using instrumentation) of exertion intensity. While we believe that the calibration task used here improved the validity of our measures (using Borg’s 10-point scale), the possibility

Figure 8. Mean (95% confidence interval) values of performance measures and SI scores.

<p>| Table 4. Results from linear regression modelling of the relationship between perceived and measured performance. |
|--------------------------------------------------|--------------------------------------------------|------------------|------------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>p Value ($r^2$)</th>
<th>Slope (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDH</td>
<td>0.0221 (0.048)</td>
<td>$-0.7163$</td>
</tr>
<tr>
<td>FTE</td>
<td>0.0305 (0.043)</td>
<td>$-0.34902$</td>
</tr>
<tr>
<td>RPA</td>
<td>0.0005 (0.110)</td>
<td>$-7.633$</td>
</tr>
<tr>
<td>FCT</td>
<td>0.0468 (0.110)</td>
<td>$-0.9058$</td>
</tr>
</tbody>
</table>

Note: FTE data were square-root transformed to meet model assumptions. Bold font indicates significant slopes ($p < 0.05$).

Also, Potvin (2011) showed a negative exponential relationship between maximum acceptable effort and duty cycle.

Regarding the relationship between task design and ergonomic risk, it was hypothesised that increasing workstation adjustability will reduce ergonomic risk. Similar to the earlier findings for performance, the effect of increasing workstation adjustability depended on the level of that added adjustability. For both RULA and SI risk assessment methods, the High condition resulted in the lowest ergonomic risk while non-significant differences were found between Some and None. As for performance, we expected that the Some condition would reduce ergonomic risk as a result of providing ‘better’, or less extreme, reaches to the holes. Although the validity of the risk assessment methods used here has been supported in earlier work (e.g. Moore, Rucker, and Knox 2001; Rucker and Moore 2002; Fountain 2003; Shuval and Donchin 2005; Spielholz et al. 2008), it is possible that these methods were
of overestimation remains present. Supporting this possibility, Spielholz et al. (2001) found that self-reported hand force, using a visual analogue scale, consistently overestimated force levels compared to both observational and direct measurement methods. For estimating gripping forces, McGorry et al. (2010) indicated that the correlation between the noted three methods is best for simple tasks. DiDomenico and Nussbaum (2008) found Borg CR10 ratings to be sensitive to changes in physical workload for lifting tasks, and this sensitivity was not altered by the addition of mental workload. This also suggests that the mental workload due to precision requirements in the current simulated task likely did not substantially affect perceived exertion.

A subsequent analysis on postural extremes (90th percentile angles) found inconsistent effects of the added adjustability in the Some condition, compared to None. For example, neck and trunk postures generally improved, but upper and lower arm postures had larger extremes. This inconsistency can be considered a manifestation of the potential trade-off between body parts in tasks (re)design (Marras 2006). Generally, it may not be feasible to reduce postural extremes for all body parts in a given (re)design, particularly for a relatively complex task.

The association between ergonomic risk and performance was assessed at two levels, first across conditions and second between the three conditions. For the former, the majority of the derived regression models (Table 3) showed a significant, indirect relationship between ergonomic risk and performance (Figure 6). Because participants responded differently to the different adjustability conditions (cf. Figure 2(a) and (b)), these models have the advantage of looking at each data point, irrespective of the adjustability condition. While RULA scores were inversely associated with both productivity and quality, SI scores were inversely associated only with quality metrics (FTE and FDH). These findings suggest that reducing risk in the distal upper extremity, as quantified by SI in the simulated task, should be associated particularly with quality improvement; reducing risk in the upper limb overall, as quantified by RULA, should be associated with productivity and quality improvements. It should be noted, though, that there was relatively large variability that was not explained by the risk metrics, highlighting the need to investigate other variables that influence performance, ergonomic risk and their inter-relationships.

Between the three conditions, there was some support for the hypothesised inverse relationship between ergonomic risk and performance (Figures 7 and 8). The lowest ergonomic risk in the High condition was associated with the highest productivity and quality, though the latter was only evident for FTE, and not for FDH. Further, the lack of difference in ergonomic risk between the None and Some conditions was associated with a lack of difference in quality (both metrics), though contrary to our hypothesis there was a deterioration in productivity (both metrics). These findings indicate a need to test the efficacy (in terms of risk reduction and performance improvement) of any (re)design prior to actual implementation.

Our findings are generally consistent with results reported in the literature. In a field study, Loo and Yeow (2015) tested two physical ergonomics interventions for a brazing task. An inverse association was evident between the three RULA scores (pre-intervention, intervention 1 and intervention 2) and both productivity and quality. However, it should be noted that RULA assessments were
done on only one operator in that study. In a metal factory, a physical ergonomic intervention was applied based on RULA risk evaluations and this reduced the rejected and reprocessed parts by 45 and 22%, respectively (González, Adeno-Díaz, and González Torre 2003). Subsequent to a redesign based on principles of work design and ergonomics, productivity and quality in a simulated drill press workstation were improved by 22 and 50%, respectively (Das, Shikdar, and Winters 2007). These studies support our findings that task (re)design can improve performance when it reduces ergonomic risk, as well as the inverse relationships between risk and performance. In assembly tasks, Ivarsson and Eek (2015) provided some support for the inverse relationship between ergonomic risk and quality. However, it should be noted that a company-specific assessment tool quantified ergonomic risk, and that its validity and reliability were not tested. Two studies in the automotive industry also showed that quality was worse for tasks that had higher ergonomic risk, as estimated by ergonomists (Falck, Örtengren, and Högborg 2010; Falck, Örtengren, and Rosenqvist 2014). However, these later three studies provided less support for the noted inverse association because comparisons were made across different tasks.

In addition to performance and injury risk, different levels of adjustability also affected perceptual responses (RPE and RPD). Such perceptual measures have been suggested as indicators of muscle fatigue (Hummel et al. 2005; Sood, Nussbaum, and Hager 2007), and thus the current results suggest that fatigue was minimised when the added adjustability reduced ergonomic risk (i.e. the High condition). This finding was consistent for the majority of RPD variables. Minimising discomfort may be an important goal, since its presence may indicate the risk of future development of a WMSD (Madeleine 2010). Across the three conditions, a significant positive linear association was observed between RPE and each of FCT, TBR, and FDH, and between RPDs (all three RPDs) and each of FCT and TBR (results not reported). A similar regression model was reported between discomfort and self-selected duty cycle for a repetitive gripping task (Finneran and O’Sullivan 2010). This is generally consistent with the concept that more rest is required for more fatiguing tasks (Rohmert 1973). Although studying the effect of fatigue on performance was beyond the scope of this study, these findings suggest an inverse association between fatigue and performance. However, mixed evidence has been reported for the effect of fatigue on performance (Mehta and Agnew 2010).

Although this was not a fundamental research question examined here, we found a significant association between perceived and measured performance. This association suggests that individual perception could be helpful to quantify performance, particularly when objective measures may not be feasible or practical. However, firm conclusions are not warranted here regarding this approach, especially given the large residual variability present (cf. Figure 9), until more evidence is available regarding validity and reliability (for more on self-reported productivity, see Prasad et al. [2004]).

There are a few limitations in this study that should be noted. While the use of a controlled experimental design ensured good internal validity, the level of external validity is unknown. For example, the current results may be specific to the tolerance level and forces employed. Vibration exposure is a known WMSD risk factor (Armstrong 1986) and present in actual drilling, yet it was not simulated here. Additionally, actual drilling operations produce torques that intensify as cutters break through the exit surface of a drilled material. These forces are reacted to by the operator in manual drilling tasks, but our simulation did not account for this effect. It is thus unclear regarding whether the current results will generalise to a wide range of actual drilling tasks or to other manual tasks. The lack of differences found between the None and Some conditions, along with the unforeseen limitations inherent in the latter condition noted above, limited our ability to address the ‘shape’ of the relationships between adjustability and performance/risk. Because recruiting experienced drilling workers was considered infeasible in the experiment, we trained novices on the task and used a repeated measures design. However, it is unknown if experienced workers would give the same pattern of major results.

We limited our definition of ergonomic risk to ‘physical’ exposure, which is an influential determinant of ergonomic risk (e.g. Winkel and Mathiassen 1994) and perhaps the most easily modifiable in practice. However, it should be noted that WMSDs are multi-factorial in nature (van der Beek and Frings-Dresen 1998). In addition to physical aspects, earlier research has identified important individual (Armstrong et al. 1993) and psychosocial/organisational (Devereux, Vlahonikolis, and Buckle 2002) occupational risk factors for WMSDs development. Consideration of the latter factors, and their possible interactive effect (e.g. Widanarko et al. 2014, 2015), however, was beyond the scope of this study.

In summary, increasing the adjustability in a simulated drilling task improved performance (quality and productivity), but only when that adjustability succeeded in also reducing ergonomic risk. Along with performance improvements, the ‘successful’ adjustability also reduced discomfort in several body parts and perceived exertion. Across the examined adjustability conditions, several significant inverse relationships were found between ergonomic risk and performance. These results provide support that ergonomic ‘improvements’ can have dual benefits in some cases, by enhancing performance in parallel with
reducing injury risk. Future work is needed, however, involving a wider range of tasks, task conditions, and participants, to assess the generality and applicability of these results.

Acknowledgement

The authors thank Ms Denean Kelson for her assistance in data collection and processing.

Funding

This work was supported by the Boeing Company. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of Boeing.

ORCID

Saad Alabdulkarim  http://orcid.org/0000-0002-2404-8207
Maury A. Nussbaum  http://orcid.org/0000-0002-1887-8431
Ehsan Rashedi  http://orcid.org/0000-0002-2998-4993
Sunwook Kim  http://orcid.org/0000-0003-3624-1781

References


