



Published in final edited form as:

*Int J Ind Ergon.* 2014 January ; 44(1): 32–38.

## Inter-rater reliability of cyclic and non-cyclic task assessment using the hand activity level in appliance manufacturing

Robert Paulsen<sup>a</sup>, Natalie Schwatka<sup>a</sup>, Jennifer Gober<sup>a,1</sup>, David Gilkey<sup>a</sup>, Dan Anton<sup>b,2</sup>, Fred Gerr<sup>b</sup>, and John Rosecrance<sup>a,\*</sup>

Robert Paulsen: rob.paulsen@rams.colostate.edu; Natalie Schwatka: natalie.schwatka@colostate.edu; Jennifer Gober: jennie.gober@gmail.com; David Gilkey: david.gilkey@colostate.edu; Dan Anton: dan.anton@ewu.edu; Fred Gerr: fred-gerr@uiowa.edu; John Rosecrance: john.rosecrance@colostate.edu

<sup>a</sup>Colorado State University, Environmental and Radiological Health Sciences, 1681 Campus Delivery, Fort Collins, CO 80523, USA

<sup>b</sup>University of Iowa, College of Public Health, 105 River St., Iowa City, IA 52242, USA

### Abstract

This study evaluated the inter-rater reliability of the American Conference of Governmental Industrial Hygienists (ACGIH<sup>®</sup>) hand activity level (HAL), an observational ergonomic assessment method used to estimate physical exposure to repetitive exertions during task performance. Video recordings of 858 cyclic and non-cyclic appliance manufacturing tasks were assessed by sixteen pairs of raters using the HAL visual-analog scale. A weighted Pearson Product Moment-Correlation Coefficient was used to evaluate the agreement between the HAL scores recorded by each rater pair, and the mean weighted correlation coefficients for cyclic and non-cyclic tasks were calculated. Results indicated that the HAL is a reliable exposure assessment method for cyclic ( $r\text{-}\bar{w} = 0.69$ ) and non-cyclic work tasks ( $r\text{-}\bar{w} = 0.68$ ). When the two reliability scores were compared using a two-sample Student's *t*-test, no significant difference in reliability ( $p = 0.63$ ) between these work task categories was found. This study demonstrated that the HAL may be a useful measure of exposure to repetitive exertions during cyclic and non-cyclic tasks.

**Relevance to industry**—Exposure to hazardous levels of repetitive action during non-cyclic task completion has traditionally been difficult to assess using simple observational techniques. The present study suggests that ergonomists could use the HAL to reliably and easily evaluate exposures associated with some non-cyclic work tasks.

### Keywords

HAL; Hand activity level; Reliability; Exposure assessment; Non-cyclic work; Repetitive exertions

\* Corresponding author. Tel.: +1 970 491 1405; fax: +1 970 491 2940.

<sup>1</sup>Present address: Humantech Inc., 1161 Oak Valley Drive, Ann Arbor, MI 48108, USA.

<sup>2</sup>Present address: Eastern Washington University, Department of Physical Therapy, 310 N, Riverpoint Blvd., Spokane, WA 99202, USA.

## 1. Introduction

Musculoskeletal disorders (MSDs) continue to be one of the leading sources of impairment and lost work time in the United States and elsewhere. In 2011, occupationally-related MSDs in the United States accounted for 32.8% of all cases of injuries and illnesses requiring time away from work and resulted in a median of 11 lost work days (Bureau of Labor Statistics, 2012). The development of MSDs is linked to a variety of physical work exposures, such as awkward postures, excessive forces, prolonged vibration, and high repetition (Bernard, 1997; NRC/IOM, 2001). In particular, repetitive hand activity has been identified as one of the primary occupational risk factors associated with upper extremity MSDs (Bernard, 1997; Latko et al., 1999; Silverstein et al., 1986, 1987). Exposure assessment tools, such as the American Conference of Governmental Industrial Hygienists (ACGIH®) Hand Activity Level (HAL) Threshold Limit Value (TLV®) (ACGIH, 2005) have been developed to quantify these physical risk factors (Latko et al., 1997). In 2001, the National Research Council and the Institute of Medicine (NRC/IOM) reported that additional occupational risk factor exposure assessment tools should be developed or improved (NRC/IOM, 2001). As a measure of physical exposure to repetitive exertion, the utility of the HAL would be improved if it could be used to assess non-cyclic tasks. Non-cyclic tasks in a manufacturing, construction, agriculture, healthcare, service, and general office/administrative industries may expose workers to repetitive exertions that repeatedly stress their musculoskeletal systems, and the associated MSD hazard exposure should be assessed (Fethke et al., 2012; Paquet et al., 2005; Punnett and Wegman, 2004).

Exposure assessment tools are used to quantify physical exposure and estimate the risk of developing a work-related MSD. Ergonomists investigating exposures that may increase MSD risk use a variety of metrics, including those based on self-report (e.g., work diaries), expert observation (e.g., HAL or Strain Index), and direct measurement (e.g., push/pull force sensors, electrogoniometry, or surface electromyography) (David, 2005; Dempsey et al., 2005; Kilbom, 1994). The choice of assessment tools depends on the characteristics of the work task, but may also depend on training, familiarity, practicality, cost, and time required to use the tool (Dempsey et al., 2005; Li and Buckle, 1999).

Some investigators have quantified repetitive hand activity in the field using direct measures of muscle activity or wrist deviation frequencies (Chen et al., 2010; Fethke et al., 2012; Hansson et al., 1996; Jones and Kumar, 2007; Spielholz et al., 2001). These data intensive methods produce quantitative estimates with better accuracy than observational or self-report assessment tools (David, 2005; Spielholz et al., 2001). However, analyzing and interpreting direct measurement results is time intensive and requires considerable technical expertise. Furthermore, the cost of instrumentation and software required to perform direct measures can be prohibitively expensive (Anton et al., 2003; David, 2005). Observational methods are frequently employed in industry because they cost less and are more time efficient than direct measures, and are generally more accurate and reliable than self-reports (Ebersole and Armstrong, 2002; Garg and Kapellusch, 2011; Kilbom, 1994; Takala et al., 2010).

In performing a HAL assessment, an ergonomist typically uses a standard scale to judge the magnitude of worker exposure to repetitive and forceful exertions. Because the estimation of HAL values is based on observer judgment, establishing the reliability of the HAL method is important for interpreting HAL results, whether the aim is for research, hazard mapping, or intervention evaluation (Kilbom, 1994; Streiner and Norman, 2008). Multiple studies report that the HAL inter-rater reliability ranges from moderate to good when assessing cyclic tasks (Ebersole and Armstrong, 2006; Spielholz et al., 2008; Takala et al., 2010). However, the inter-rater reliability of non-cyclic task assessment has not been estimated, in part because the HAL was designed to assess cyclic, mono-task jobs (Armstrong, 2006; Latko et al., 1997), but also because of the difficulty assessing non-cyclic tasks given the absence of an inherent task completion pattern (Punnett and Wegman, 2004).

In some of the earlier literature, the distinction between cyclic and repetitive tasks is unclear (Bao et al., 2009; Latko et al., 1997). This is primarily because some ergonomic researchers have used the concept of cycle-time to define tasks as repetitive or non-repetitive (Armstrong et al., 1987; Buchholz et al., 1996; Chiang et al., 1993; Colombini, 1998; Silverstein et al., 1986). In the present study, appliance assembly line tasks were evaluated regardless of whether they were expected to be classified as repetitive according to the HAL or any other type of exposure assessment. The aim of this was to determine if a repetitive task could be estimated reliably regardless of whether the work was cyclic or non-cyclic. Further, classification of tasks as cyclic or non-cyclic was entirely based on whether the work conformed to easily identifiable patterns of subtask or work element procedures lasting no more than 3 min.

Further confusion arises from the inconsistent usage of the terms “mono-task,” “single-exertion,” and “complex task” (Bao et al., 2009; Kapellusch et al., 2013). The HAL was designed to assess repetitive force exposures during mono-task work performance lasting at least 4 h (Armstrong, 2006). The developers of the assessment defined mono-task work as a predictable pattern of work elements (or subtasks) reoccurring throughout the work shift (ACGIH, 2005; Latko et al., 1997). This definition of mono-task work differs from the one presented by Moore and Garg (1995) during their description of a similar assessment tool, the Strain Index, where they equated mono-tasks with single exertion tasks (Moore and Garg, 1995). More often than not, tasks are comprised of subtasks requiring different levels of exertion rather than a single level of exertion, and these are called complex tasks (Bao et al., 2009; Garg and Kapellusch, 2011; Kapellusch et al., 2013). In the present study, the HAL was applied to single exertion and complex exertion tasks. Some of these tasks were characterized by unpredictable subtask performance patterns (i.e. non-cyclic tasks), so they would not be considered mono-tasks according to the HAL developers. Nonetheless, these non-cyclic tasks may still expose workers to predictable patterns of repetitive force exertions. The purpose of the present study was to compare the inter-rater reliability of the HAL assessments used to estimate worker exposure to repetitive hand exertions during cyclic and non-cyclic task performance in the appliance manufacturing industry.

## 2. Materials and methods

### 2.1. Study context

The present study obtained previously recorded videos of cyclic and non-cyclic work tasks performed by adult (18 years aged) workers in a household appliance manufacturing facility. The videos were recorded during a large prospective cohort study (Gerr et al., 2013) focused on associations between physical exposures and MSD incidence among manufacturing workers.

The appliance manufacturing facility employed approximately 2000 workers on multiple assembly lines. The research team observed manual tasks performed on multiple assembly lines representing all stages of appliance production—from materials fabrication to product assembly and packaging. For the present study, “tasks” were defined as assembly, inspection, or packaging procedures performed at a specific workstation, such as “assemble wire harness” or “install ice maker.” Tasks were categorized as cyclic if they were performed according to an identifiable work cycle lasting 3 min or less. Otherwise, tasks were categorized as non-cyclic. University faculty members in ergonomics determined *a priori* whether tasks were cyclic and non-cyclic. An appliance product quality inspection task is a good example of one that is non-cyclic. This task involved use of hand tools requiring various levels of grip strength to operate, manual handling of materials of varying weights, intermittent inspection of control panels, and the making of assembly line adjustments as needed. The subtasks or work elements comprising the quality inspection task did not proceed according to a clearly identifiable procedure, and inspections could last longer than 3 min.

Digital video cameras were arranged within the manufacturing facility to grossly record the frontal and sagittal planes of the workers' upper extremities during task completion. One video camera was mounted on a tripod for a consistent, stable viewing angle, while another researcher operated a hand-held camera. Camera views were continuously adjusted in an attempt to fill the frame with the worker's upper body. Dynamic control of the second camera improved tracking of the upper limbs when work materials or equipment obstructed the view of the workers. Workers were videotaped for a minimum of 30 min for each task that they performed. Prior to the HAL rating sessions the two video recordings were synchronized, providing raters with two simultaneous views of each worker.

In the present study, video recordings of 385 workers performing their standard assembly-line tasks were observed, and a total of 858 tasks were evaluated with the HAL. The mean worker age was 42.3 years ( $SD = 10.6$ ), and on average they had worked at the manufacturing facility for 14.7 years ( $SD = 11.4$ ). Workers were primarily non-Hispanic white (91.5%), and there were approximately equal numbers of males (48.7%) and females (51.3%). Nearly all (96.6%) had at least a high school diploma and 30.2% had received some post-secondary education or training. The majority were also right handed (88.3%).

Raters were on average 29.8 years ( $SD = 8.6$ ) of age and about half (54.5%) were female. Raters consisted of two university faculty members experienced at using the HAL and nine graduate students who were trained to use the HAL by one of the two faculty members.

Fifteen pairs of raters assessed the cyclic tasks and six pairs assessed the non-cyclic tasks. Between the two task categories, sixteen unique rater-pair combinations participated.

The study procedures were approved by the Institutional Review Board at the University of Iowa. Study participants were aware that their exposure to physical risk factors for MSDs was under observation. All participants provided their written consent.

## 2.2. Procedures

Video recordings of cyclic and non-cyclic work tasks were provided to graduate students and faculty in the field of ergonomics to conduct HAL ratings. Two raters assessed each video-recorded work task. Each rater-pair consisted of one rater from each of two universities (University of Iowa, Colorado State University). For all work tasks assessed, each member of the rater-pair recorded a HAL score independently of the other rater, and each task was rated by only one pair of raters. Raters estimated the HAL for all tasks using Latko's 10-cm visual-analog scale with verbal anchors (ACGIH, 2005; Latko et al., 1997) rather than using the ACGIH tabulation table (ACGIH, 2005). Using only the visual-analog scale reduced the time necessary to complete the 858 task ratings. Further, the visual-analog scale is easy to employ in industry (Ebersole and Armstrong, 2006; Wurzelbacher et al., 2010) and recent longitudinal studies of job physical exposure have all used the visual-analog HAL scale to assess task repetition, whereas only some have used the HAL tabulation table approach (Kapellusch et al., 2013).

One faculty member at each academic institution trained their respective graduate students on the use of the HAL. The training began with a didactic review of the HAL scale and its application, followed by a series of practice rating sessions to ensure complete familiarity with the verbal anchors of the current visual-analog scale (ACGIH, 2005). These practice sessions required students and faculty to independently rate video segments of manufacturing tasks that exhibited a range of hand activity levels. The ratings were compared for consistency, and tasks were analyzed until students and faculty members were able to reach consensus (i.e. consistently rate tasks within one unit of each other) for a minimum of five tasks. Additionally, pairs of students at each respective institution compared independent ratings of twenty work tasks until a consensus was reached. Students were considered competent as HAL raters upon completing this training.

All HAL ratings were completed based on the video of the worker's dominant limb as defined by their handedness with writing. For cyclic tasks, the HAL score determined by the rater was based on observation of three task cycles. Each cycle analyzed was chosen *a priori* and consisted of one cycle from the first 5 min, one cycle from 15 to 20 min and one cycle from 25 to 30 min of the recorded video sample. For non-cyclic tasks, three video samples were randomly chosen *a priori* to conduct the HAL rating and consisted of one from a 30-s interval during the first 5-min, a 30-s interval from 15 to 20 min, and a 30-s interval from 25 to 30 min. All of the video samples analyzed by the rater-pairs were selected *a priori* by a research team member who did not participate as a HAL rater.

After viewing the three video samples of the work task, each rater recorded a single HAL rating for the task into a computer spreadsheet. In some cases, a worker did not maintain a

consistent level of hand activity for the entire task duration. When this occurred, the raters were instructed to average the HAL scores for the three task samples to reach a single rating.

### 2.3. Statistical analysis

Statistical analysis was completed separately for cyclic and non-cyclic task categories, and analyses were performed using SAS/STAT<sup>®</sup> software version 9.3. For each rater pair, inter-rater reliability was measured as agreement between the two HAL scores through a Pearson Product Moment Correlation Coefficient ( $r$ ) (Streiner and Norman, 2008). In order to obtain a mean correlation value ( $\bar{r}$ ) for each of the task categories (i.e. cyclic and non-cyclic), each  $r$ -value underwent a Fisher  $z$ -score transformation. The mean  $z$ -score ( $\bar{z}$ ) was calculated and back-transformed to obtain a mean correlation value ( $\bar{r}$ ) (Steel and Torrie, 1980). To account for the variation in the number of tasks analyzed by each rater-pair, the  $z$ -scores were weighted ( $\bar{z}_w$ ) and then back-transformed to a weighted mean  $r$ -value ( $\bar{r}_w$ ) (Steel and Torrie, 1980), yielding an estimate of the overall inter-rater reliability of each task category. Confidence intervals (95%) were obtained for the weighted mean  $r$ -values based on the weighted mean  $z$ -scores and SAS software-generated weighted variance estimates.

To aid in the interpretation of results, the following decision criteria for weighted mean correlation coefficients were adopted: negligible reliability: 0.00–0.25; fair to moderate reliability: 0.25–0.50; moderate to good reliability: 0.50–0.75; good to excellent reliability: 0.75–1.0. The selection of these criteria was based on similar studies of rater reliability (Dartt et al., 2009; Ebersole and Armstrong, 2002; Stevens et al., 2004) as well as other reliability statistics, such as the kappa coefficient and the intra-class correlation (Fleiss, 1986; Streiner and Norman, 2008). To evaluate if inter-rater reliability differed depending on whether the tasks rated were cyclic or non-cyclic, a two-sample Student's  $t$ -test ( $\alpha = 0.05$ ) using Satterthwaite's method for unequal variance compared the weighted mean  $z$ -scores from both task categories (Ott and Longnecker, 2010).

## 3. Results

### 3.1. HAL assessments

A total of 1072 work tasks were initially recorded, but 214 were not included in the data analyses because the video was already used for HAL training purposes or because the task was only rated by one person. A total of 858 work tasks, consisting of 71 non-cyclic tasks and 787 cyclic tasks, were rated and used in the statistical analyses. Using the 0 to 10 point scale of the HAL rating system, cyclic ratings ranged between 2 and 9 with a mean rating of 5.3 (SD = 1.2). Non-cyclic ratings ranged between 1 and 8 with a mean rating of 4.9 (SD = 1.4).

### 3.2. Inter-rater reliability

The inter-rater reliability of both cyclic (Table 1) and non-cyclic work (Table 2) tasks was evaluated using a weighted mean Pearson's Product Moment Correlation Coefficient ( $\bar{r}_w$ ). Fifteen rater pairs rated 787 cyclic work tasks, rating an average of 52.5 work tasks each. The unweighted mean correlation between ratings for cyclic tasks among all rater-

pairs was  $r\text{-}\bar{r} = 0.79$ , and weighting by task produced a correlation value of  $r\text{-}\bar{r}_w = 0.69$  (95% CI: 0.61, 0.77). Six rater pairs rated 71 non-cyclic work tasks, rating an average of 11.8 work tasks each. The mean unweighted correlation between ratings for non-cyclic work tasks among all rater-pairs was  $r\text{-}\bar{r} = 0.73$  and the weighted correlation value was  $r\text{-}\bar{r}_w = 0.68$  (95% CI: 0.45, 0.82).

The mean weighted  $z$ -scores ( $z\text{-}\bar{z}_w$ ) for the cyclic and non-cyclic tasks were compared with a two-sample Student's  $t$ -test using Satterthwaite's method for unequal variance. No significant difference at the 95% confidence level was found between the mean inter-rater reliability scores for cyclic and non-cyclic tasks ( $df = 9.4$ ,  $t = 0.50$ ,  $p = 0.63$ ).

#### 4. Discussion

The present study is the first published study reporting on the inter-rater reliability of the HAL for non-cyclic work tasks. The results suggested that the HAL is a reliable measure of exposure to repetitive exertions regardless of whether the task was cyclic or non-cyclic. Given that the inter-rater reliability of non-cyclic task assessment was moderate to good, the application of the HAL may also be useful for non-cyclic work tasks. With further validation and study of HAL applications to non-cyclic task assessment, occupational health professionals may be able to identify ergonomic hazards among a greater variety of work tasks than previously expected. Additionally, ergonomists may seek to test the HAL as an intervention outcome measure of repetitive hand activity regardless of whether the task is cyclic or non-cyclic.

In an effort to ensure the validity of the HAL, Latko et al. (1997) created the visual-analog scale based on assessments of over 185 jobs in multiple industries with varying tasks. Since its development, the HAL and ACGIH<sup>®</sup> TLV<sup>®</sup> have been used to evaluate upper extremity MSD risk factor exposure in a variety of industries, although most evaluations have been made in manufacturing environments (Dempsey et al., 2005; Franzblau et al., 2005; Garg et al., 2012; Gerr et al., 2013; Kapellusch et al., 2013; Latko et al., 1999). It is typically used to evaluate cyclic tasks in which a well-defined set of work cycles or a series of forceful exertions are repeated on a regular basis. Examples of cyclic tasks involve assembly or disassembly work, such as those found in appliance manufacturing, automobile assembly, or meat processing. Previous studies have investigated the inter-rater reliability of the HAL for cyclic work tasks (Ebersole and Armstrong, 2002; Spielholz et al., 2008). Several investigators have used other observational measures to assess the physical risk associated with non-cyclic (or variable) tasks (Hoozemans et al., 2001; Paquet et al., 2005; Tak et al., 2009), but none have investigated the HAL scale reliability when applied to non-cyclic work tasks.

The findings from the present study support previous research indicating that the HAL is a reliable measure of repetition exposure from cyclic work tasks (Armstrong, 2006; Takala et al., 2010). Ebersole and Armstrong's (2002) evaluation of 410 on-line jobs at an automotive assembly plant using the HAL found a weighted kappa value of  $K = 0.52$ . According to their definition, this reliability estimate was considered "moderate" for inter-rater reliability. Ebersole and Armstrong later reported that HAL assessments of 848 cyclic automotive line

jobs were reliable, reporting an intraclass correlation coefficient (ICC) of 0.71 for pairs of raters (Ebersole and Armstrong, 2006). Spielholz et al. (2008) evaluated 125 mono-task manufacturing and healthcare tasks using the HAL. Inter-rater reliability was measured using Spearman correlations and unweighted kappa coefficients. Ratings were characterized by a Spearman value of  $r = 0.65$ , and the overall kappa value for rater pairs was  $K = 0.34$ . The authors considered the HAL scale to exhibit “fair to moderate” reliability. They also compared ratings between pairs of expert (Certified Professional Ergonomist) and novice (master's degree student) raters and found that expert–expert pairs exhibited a greater agreement ( $K = 0.40$ ) than expert–novice pairs ( $K = 0.25$ ). The study only included one novice rater and three expert raters, and therefore the results may not be generalizable to other rater populations. The present study did not examine the differences in ratings between experts and novices.

Because of the long latency period of many work-related MSDs, measuring health outcomes after implementing work process changes often requires observing and evaluating workers for at least 4–6 months, and preferably up to a year or more (Kennedy et al., 2010; Westgaard and Winkel, 1997). Yet, shorter outcome observation times are possible when measuring changes in physical risk factor exposure (Westgaard and Winkel, 1997; Zwierling et al., 1997). And the most comprehensive interventions often include outcome measures of risk factor exposure, regardless of the time allotted for follow-up observations (Denis et al., 2008). The results of the present study do not imply that HAL repetition exposure estimates would also be a reliable measure of ergonomic interventions. This study was not designed to test the HAL as an intervention tool. However, if the HAL were used as an ergonomic outcome measure in a manufacturing setting, the present study suggests that the inter-rater reliability would be similar for cyclic and non-cyclic tasks. Those interested in using the HAL as an intervention tool are encouraged to use caution when applying the instrument to cyclic and non-cyclic task assessments.

Several ergonomic intervention studies have reported outcome measures using observational assessments, such as the Rapid Upper Limb Assessment (RULA) (Choobineh et al., 2004; Kilroy and Dockrell, 2000; Massaccesi et al., 2003; Robertson et al., 2009) and Rapid Entire Body Assessment (REBA) (Pillastrini et al., 2010; Yanes Escalona et al., 2012). Peer-reviewed publications describing the use of the HAL or ACGIH® TLV® for HAL as intervention outcome measures were not found, but other investigators have used similar upper extremity assessments for this purpose, such as the Strain Index (Moore and Garg, 1997; Motamedzade et al., 2011) and the Occupational Repetitive Actions (OCRA) Checklist (Escalona and Yanes, 2012). There is little to no evidence that the reliability of these observational assessments is greater than the ACGIH® HAL (Takala et al., 2010). And when used to calculate the ACGIH® TLV® for repetitive hand activity, the HAL has demonstrated sensitivity to health outcomes (Bonfiglioli et al., 2012; Garg et al., 2012).

#### 4.1. Limitations and future study

The present study relied on data obtained during a large prospective cohort study that was focused on the relationship between exposures and health outcomes and not necessarily on the interrater reliability of the HAL scale (Gerr et al., 2013). If the *a priori* research question



was only an assessment of inter-rater reliability, then using an ICC measure of reliability, rather than Pearson's correlation, would have been a more robust and appropriate statistical measure. An ICC could identify whether variance in the mean ratings of a task from multiple rater-pairs contributes to measurement error whereas the Pearson cannot (Streiner and Norman, 2008). However, in the present study, because no more than one rater-pair evaluated any particular task, an ICC cannot be calculated.

The margin of error for the mean weighted agreement ( $r\text{-}\bar{w}$ ) value for non-cyclic tasks was about two times the size of the error margin for cyclic tasks. This could be due to the smaller sample size of 71 tasks rated compared to 787 cyclic tasks rated. However, greater variation in ratings might be reasonable given the wider variation in non-cyclic task performance and the lack of an inherent task cycle. Whatever the cause for the greater variance in the non-cyclic  $r\text{-}\bar{w}$ -value, the finding remains that its confidence interval spans more than one reliability category. While the mean is firmly situated in the moderate to good reliability range, the lower bound is 0.45, which is just into the fair to moderate reliability range.

The present study did not evaluate the reliability of peak force exposure estimates during cyclic or non-cyclic task performance. The peak force estimate is used in conjunction with the HAL rating to determine if a task is above or below the ACGIH<sup>®</sup> TLV<sup>®</sup> or Action Limit. While this study suggest that the inter-rater reliability of the HAL might be equivalent when applied to any repetitive manufacturing tasks, it would be preferable to know whether the full ACGIH<sup>®</sup> TLV<sup>®</sup> for hand activity can be successfully applied to non-cyclic tasks characterized by repetitive technical actions. Similarly, the research design did not allow for an assessment of the intra-rater reliability of the HAL scale. These limitations were due to resource constraints. Future research should focus on the test-rest reliability of both the HAL and peak force exposure estimates during non-cyclic task performance, and evaluating the validity of these estimates for non-cyclic tasks is essential. Further, the reliability of the HAL as applied by groups of researchers independently recording and observing the same cyclic and non-cyclic tasks should be studied. This would inform potential HAL-users of any differences between the reliability of rater-groups that reach consensus compared to single raters.

One of the challenges with any observational exposure assessment tool is ensuring that observations are consistent between raters. In the present study, video recordings captured upper body work activity in two different anatomical planes. The two video recording planes were used in an attempt to increase the visibility of the upper extremity of the worker. Unfortunately, the upper extremity was not always visible for the entire task duration, for instance, while the worker reached within the product to secure an attachment point. Other times, the upper extremities were obscured by machinery or materials moving in front of the cameras during the manufacturing process. In these cases, raters were told to rate what they could see, but there may be potential subjectivity in what the rater considered "visible." Additionally, the work tasks occasionally contained long pauses followed by activity. During pauses or obstructions, raters were instructed to "mentally average" the HAL ratings as described by Latko et al. (1997). Mental averaging could be a source of variability between raters, as there is some subjectivity with this method. In practice, this is somewhat

accounted for by reaching consensus with other raters as  $\pm 1$  on the rating scale (Armstrong, 2006; Latko et al., 1997). In the present study, reliability analyses were conducted before consensus was reached.

Another source of variability between raters is the interpretation of HAL verbal anchors. Raters may each have a slightly different interpretation of words such as “steady”, “frequent,” or “consistent.” Additionally, this study was not designed to evaluate the intra-rater reliability of the HAL applied to non-cyclic tasks. Although the purpose of training is to minimize intra-rater variability and error introduced by verbal anchor interpretation, some variability likely persisted. For those practitioners interested in applying the HAL as an intervention outcome measure, it is worth noting that test-retest reliability is generally greater than inter-rater reliability for observational ergonomic assessment tools (Takala et al., 2010).

## 5. Conclusion

The present study appears to be the first to assess the inter-rater reliability of the HAL for non-cyclic work tasks. Observational exposure assessment tools, such as the HAL, enable researchers and practitioners to evaluate large samples of workers with minimally invasive techniques and limited resources. The findings of the present study are consistent with previous research that has determined the HAL to be a reliable exposure assessment tool for cyclic work tasks. The findings suggest that the HAL is a reliable ergonomic exposure assessment tool for non-cyclic work tasks.

## Acknowledgments

The authors wish to thank Jim zumBrunnen from the Franklin A. Graybill Statistical Laboratory at Colorado State University for his assistance with the SAS software statistical analysis. The present study was supported by the National Institute for Occupational Safety & Health (NIOSH) (grant number: R01 OH007945) and the Center for Disease Control (CDC)/NIOSH Mountain and Plains Education and Research Center (grant number: T42OH009229-04). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or NIOSH.

## References

- American Conference of Governmental Industrial Hygienists (ACGIH). Documentation of the TLVs and BEIs with Their Worldwide Occupational Exposure Values. ACGIH Worldwide; Cincinnati, OH: 2005.
- Anton D, Cook TM, Rosecrance JC, Merlino LA. Method for quantitatively assessing physical risk factors during variable noncyclic work. *Scand J Work Environ Health*. 2003; 29(5):354–362. [PubMed: 14584516]
- Armstrong, TJ. The ACGIH TLV<sup>®</sup> for hand activity level. In: Marras, WS.; Karwowski, W., editors. *Fundamentals and Assessment Tools for Occupational Ergonomics*. CRC Press; Boca Raton, FL: 2006.
- Armstrong TJ, Fine LJ, Goldstein SA, Lifshitz YR, Silverstein BA. Ergonomics considerations in hand and wrist tendonitis. *J Hand Surg-Am*. 1987; 12A(5):830–837. [PubMed: 3655257]
- Bao S, Spielholz P, Howard N, Silverstein B. Application of the strain index in multiple task jobs. *Appl Ergon*. 2009; 40(1):56–68. <http://dx.doi.org/10.1016/j.apergo.2008.01.013>. [PubMed: 18336790]

- Bernard BP. Musculoskeletal Disorders and Workplace Factors: a Critical Review of Epidemiologic Evidence for Work-related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back. DHHS (NIOSH) Publication no. 97-141. 1997
- Bonfiglioli, R.; Mattioli, S.; Armstrong, T.; Graziosi, F.; Marinelli, F.; Farioli, A., et al. Validation of the ACGIH TLV for hand activity level in the OCTOPUS cohort: a two-year longitudinal study of carpal tunnel syndrome. *Scand J Work Environ Health*. 2012. <http://dx.doi.org/10.5271/sjweh.3312>
- Buchholz B, Paquet V, Punnett L, Lee D, Moir S. PATH: a work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work. *Appl Ergon*. 1996; 27(3): 177–187. [http://dx.doi.org/10.1016/0003-6870\(95\)00078-x](http://dx.doi.org/10.1016/0003-6870(95)00078-x). [PubMed: 15677058]
- Bureau of Labor Statistics (BLS). Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work. Nov 08. 2012 From: <http://www.bls.gov/news.release/osh2.toc.htm>
- Chen HC, Chang CM, Liu YP, Chen CY. Ergonomic risk factors for the wrists of hairdressers. *Appl Ergon*. 2010; 41(1):98–105. <http://dx.doi.org/10.1016/j.apergo.2009.05.001>. [PubMed: 19497556]
- Chiang HC, Ko YC, Chen SS, Yu HS, Wu TN, Chang PY. Prevalence of shoulder and upper-limb disorders among workers in the fish-processing industry. *Scand J Work Environ Health*. 1993; 19(2):126–131. [PubMed: 8316780]
- Chooibneh A, Tosian R, Alhamdi Z, Davarzanie M. Ergonomic intervention in carpet mending operation. *Appl Ergon*. 2004; 35(5):493–496. [PubMed: 15246888]
- Colombini D. An observational method for classifying exposure to repetitive movements of the upper limbs. *Ergonomics*. 1998; 41(9):1261–1289. <http://dx.doi.org/10.1080/001401398186306>. [PubMed: 9754031]
- Dartt A, Rosecrance J, Gerr F, Chen P, Anton D, Merlino L. Reliability of assessing upper limb postures among workers performing manufacturing tasks. *Appl Ergon*. 2009; 40(3):371–378. <http://dx.doi.org/10.1016/j.apergo.2008.11.008>. [PubMed: 19103432]
- David GC. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occup Med-Ox*. 2005; 55(3):190–199. <http://dx.doi.org/10.1093/occmed/kqi082>.
- Dempsey PG, McGorry RW, Maynard WS. A survey of tools and methods used by certified professional ergonomists. *Appl Ergon*. 2005; 36(4):489–503. <http://dx.doi.org/10.1016/j.apergo.2005.01.007>. [PubMed: 15892943]
- Denis D, St-Vincent M, Imbeau D, Jette C, Nastasia I. Intervention practices in musculoskeletal disorder prevention: a critical literature review. *Appl Ergon*. 2008; 39(1):1–14. [PubMed: 17506975]
- Ebersole, ML.; Armstrong, TJ. Inter-rater reliability for hand activity level (HAL) and force metrics. Paper presented at the Human Factors and Ergonomics Society 46th Annual Meeting; Baltimore, MD. 2002.
- Ebersole ML, Armstrong TJ. Analysis of an observational rating scale for repetition, posture, and force in selected manufacturing settings. *Hum Factors*. 2006; 48(3):487–498. <http://dx.doi.org/10.1518/001872006778606912>. [PubMed: 17063964]
- Escalona E, Yanes L. The reality of the women who make our lives easier: experience in a company that assembles electric motors in Venezuela. *Work: J Prevent Assess Rehabil*. 2012; 41:1775–1777.
- Fethke NB, Gerr F, Anton D, Cavanaugh JE, Quickel MT. Variability in muscle activity and wrist motion measurements among workers performing non-cyclic work. *J Occup Environ Hyg*. 2012; 9(1):25–35. <http://dx.doi.org/10.1080/15459624.2012.634361>. [PubMed: 22150404]
- Fleiss, JL. *The Design and Analysis of Clinical Experiments*. Wiley; New York: 1986.
- Franzblau A, Armstrong TJ, Werner RA, Ulin SS. A cross-sectional assessment of the ACGIH TLV for hand activity level. *J Occup Rehabil*. 2005; 15(1):57–67. <http://dx.doi.org/10.1007/s10926-005-0874-z>. [PubMed: 15794497]
- Garg A, Kapellusch J. Job analysis techniques for distal upper extremity disorders. *Rev Hum Fact Ergonom*. 2011; 7(1):149–196. <http://dx.doi.org/10.1177/1557234x11410386>.
- Garg A, Kapellusch J, Hegmann K, Wertsch J, Merryweather A, Deckow-Schaefer G, et al. The strain index (SI) and threshold limit value (TLV) for hand activity level (HAL): risk of carpal tunnel syndrome (CTS) in a prospective cohort. *Ergonomics*. 2012; 55(4):396–414. <http://dx.doi.org/10.1080/00140139.2011.644328>. [PubMed: 22397385]

- Gerr, F.; Fethke, N.; Merlino, L.; Anton, D.; Rosecrance, J.; Jones, MP., et al. A prospective study of musculoskeletal outcomes among manufacturing Workers: I. Effects of physical risk factors. *Hum Fact J Hum Fact Ergonom Soc.* 2013. <http://dx.doi.org/10.1177/0018720813491114>
- Hansson GA, Balogh I, Ohlsson K, Rylander L, Skerfving S. Goniometer measurement and computer analysis of wrist angles and movements applied to occupational repetitive work. *J Electromyogr Kinesiol.* 1996; 6(1):23–35. [PubMed: 20719660]
- Hoozemans MJM, Burdorf A, van der Beek AJ, Frings-Dresen MHW, Mathiassen SE. Group-based measurement strategies in exposure assessment explored by bootstrapping. *Scand J Work Environ Health.* 2001; 27(2):125–132. [PubMed: 11409595]
- Jones T, Kumar S. Comparison of ergonomic risk assessments in a repetitive high-risk sawmill occupation: saw-filer. *Int J Ind Ergon.* 2007; 37(9–10):744–753. <http://dx.doi.org/10.1016/j.ergon.2007.05.005>.
- Kapellusch JM, Garg A, Bao SS, Silverstein BA, Burt SE, Dale AM, et al. Pooling job physical exposure data from multiple independent studies in a consortium study of carpal tunnel syndrome. *Ergonomics.* 2013; 56(6):1021–1037. <http://dx.doi.org/10.1080/00140139.2013.797112>. [PubMed: 23697792]
- Kennedy CA, Amick BC III, Dennerlein JT, Brewer S, Catli S, Williams R, et al. Systematic review of the role of occupational health and safety interventions in the prevention of upper extremity musculoskeletal symptoms, signs, disorders, injuries, claims and lost time. *J Occup Rehabil.* 2010; 20(2):127–162. [PubMed: 19885644]
- Kilbom A. Assessment of physical exposure in relation to work-related musculoskeletal disorders – what information can be obtained from systematic observations. *Scand J Work Environ Health.* 1994; 20:30–45. [PubMed: 7846490]
- Kilroy N, Dockrell S. Ergonomic intervention: its effect on working posture and musculoskeletal symptoms in female biomedical scientists. *Br J Biomed Sci.* 2000; 57(3):199. [PubMed: 11050770]
- Latko WA, Armstrong TJ, Foulke JA, Herrin GD, Radbourn RA, Ulin SS. Development and evaluation of an observational method for assessing repetition in hand tasks. *Am Ind Hyg Assoc J.* 1997; 58:278–285. [PubMed: 9115085]
- Latko WA, Armstrong TJ, Franzblau A, Ulin SS, Werner RA, Albers JW. Crosssectional study of the relationship between repetitive work and the prevalence of upper limb musculoskeletal disorders. *Am J Ind Med.* 1999; 36:248–259. [PubMed: 10398933]
- Li G, Buckle P. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posturebased methods. *Ergonomics.* 1999; 42(5):674–695. <http://dx.doi.org/10.1080/001401399185388>. [PubMed: 10327891]
- Massaccesi M, Pagnotta A, Soccetti A, Masali M, Masiero C, Greco F. Investigation of work-related disorders in truck drivers using RULA method. *Appl Ergon.* 2003; 34(4):303–307. [PubMed: 12880740]
- Moore JS, Garg A. The strain index – a proposed method to analyze jobs for risk of distal upper extremity disorders. *Am Ind Hyg Assoc J.* 1995; 56(5):443–458. [PubMed: 7754975]
- Moore JS, Garg A. Participatory ergonomics in a red meat packing plant, part I: evidence of long-term effectiveness. *Am Ind Hyg Assoc J.* 1997; 58(2):127–131.
- Motamedzade M, Mohseni M, Golmohammadi R, Mahjoob H. Ergonomics intervention in an Iranian television manufacturing industry. *Work: J Prevent Assess Rehabil.* 2011; 38(3):257–263.
- National Research Council and the Institute of Medicine (NRC/IOM). *Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities.* National Academy Press; Washington, D.C: 2001.
- Ott, L.; Longnecker, M. *An Introduction to Statistical Methods and Data Analysis.* Brooks/Cole Cengage Learning; Belmont, CA: 2010.
- Paquet V, Punnett L, Woskie S, Buchholz B. Reliable exposure assessment strategies for physical ergonomics stressors in construction and other non-routinized work. *Ergonomics.* 2005; 48(9): 1200–1219. <http://dx.doi.org/10.1080/00140130500197302>. [PubMed: 16251156]

- Pillastrini P, Mugnai R, Bertozzi L, Costi S, Curti S, Guccione A, et al. Effectiveness of an ergonomic intervention on work-related posture and low back pain in video display terminal operators: a 3 year cross-over trial. *Appl Ergon*. 2010; 41(3):436–443. [PubMed: 19853837]
- Punnett L, Wegman DH. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *J Electromyogr Kinesiol*. 2004; 14(1):13–23. <http://dx.doi.org/10.1016/j.jelekin.2003.09.015>. [PubMed: 14759746]
- Robertson M, Amick BC III, DeRango K, Rooney T, Bazzani L, Harrist R, et al. The effects of an office ergonomics training and chair intervention on worker knowledge, behavior and musculoskeletal risk. *Appl Ergon*. 2009; 40(1):124–135. [PubMed: 18336791]
- Silverstein BA, Fine LJ, Armstrong TJ. Hand wrist cumulative trauma disorders in industry. *Br J Ind Med*. 1986; 43:779–784. [PubMed: 3790459]
- Silverstein BA, Fine LJ, Armstrong TJ. Occupational factors and carpal tunnel syndrome. *Am J Ind Med*. 1987; 11:343–358. [PubMed: 3578290]
- Spielholz P, Bao S, Howard N, Silverstein B, Fan J, Smith C, et al. Reliability and validity assessment of the hand activity level Threshold Limit value and strain index using expert ratings of mono-task jobs. *J Occup Environ Hyg*. 2008; 5(4):250–257. <http://dx.doi.org/10.1080/15459620801922211>. [PubMed: 18286422]
- Spielholz P, Silverstein B, Morgan M, Checkoway H, Kaufman J. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics*. 2001; 44(6):588–613. <http://dx.doi.org/10.1080/00140130118050>. [PubMed: 11373023]
- Steel, RG.; Torrie, JH. *Principles and Procedures of Statistics: a Biometrical Approach*. second. McGraw-Hill; New York: 1980.
- Stevens EM, Vos GA, Stephens JP, Moore JS. Inter-rater reliability of the strain index. *J Occup Environ Hyg*. 2004; 1(11):745–751. [PubMed: 15673095]
- Streiner, DL.; Norman, GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press; Oxford: 2008.
- Tak S, Paquet V, Woskie S, Buchholz B, Punnett L. Variability in risk factors for knee injury in construction. *J Occup Environ Hyg*. 2009; 6(2):113–120. <http://dx.doi.org/10.1080/15459620802615822>. [PubMed: 19085603]
- Takala EP, Pehkonen I, Forsman M, Hansson GA, Mathiassen SE, Neumann WP, et al. Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scand J Work Environ Health*. 2010; 36(1):3–24. [PubMed: 19953213]
- Westgaard R, Winkel J. Ergonomic intervention research for improved musculoskeletal health: a critical review. *Int J Ind Ergon*. 1997; 20(6):463–500.
- Wurzelbacher S, Burt S, Crombie K, Ramsey J, Luo L, Allee S, et al. A comparison of assessment methods of hand activity and force for use in calculating the ACGIH<sup>®</sup> hand activity level (HAL) TLV<sup>®</sup>. *J Occup Environ Hyg*. 2010; 7(7):407–416. <http://dx.doi.org/10.1080/15459624.2010.481171>. [PubMed: 20446152]
- Yanes Escalona L, Sandia Venot R, Escalona E, Yanes L. The reality of the women who make our lives easier: experience in a company that assembles electric motors in Venezuela. *Work: J Prevent Assess Rehabil*. 2012; 41:1775–1777.
- Zwerling C, Daltroy LH, Fine LJ, Johnston JJ, Melius J, Silverstein BA. Design and conduct of occupational injury intervention studies: a review of evaluation strategies. *Am J Ind Med*. 1997; 32(2):164–179. [PubMed: 9215438]

**Table 1**

Inter-rater reliability for cyclic work tasks.

Rater-pair <sup>a</sup>	Number of tasks rated	<i>r</i>
Rater A & Rater 1	108	0.79
Rater A & Rater 2	153	0.66
Rater A & Rater 3	12	0.88
Rater A & Rater 4	7	0.97
Rater B & Rater 2	103	0.58
Rater C & Rater 1	6	0.64
Rater C & Rater 2	79	0.61
Rater D & Rater 2	140	0.68
Rater E & Rater 1	7	0.99
Rater E & Rater 2	121	0.70
Rater E & Rater 3	4	0.94
Rater E & Rater 4	9	0.46
Rater F & Rater 1	10	0.78
Rater F & Rater 2	22	0.44
Rater G & Rater 1	6	0.57
<i>r</i> -bar	0.79	
<i>r</i> -bar <sub>w</sub>	0.69 (95% CI = 0.61, 0.77)	

Note. CI = Confidence Interval.

<sup>a</sup>Raters A–G were from the University of Iowa and raters 1–4 were from Colorado State University. Raters F and 3 were faculty.

**Table 2**

Inter-rater reliability for non-cyclic work tasks.

Rater-pair <sup>a</sup>	Number of tasks rated	<i>r</i>
Rater A & Rater 1	7	0.84
Rater A & Rater 2	10	0.47
Rater A & Rater 3	6	0.89
Rater E & Rater 2	38	0.60
Rater E & Rater 4	7	0.83
Rater G & Rater 3	3	0.5
<i>r</i> -bar	0.73	
<i>r</i> -bar <sub>w</sub>	0.68 (95% CI = 0.45, 0.82)	

Note. CI = Confidence Interval.

<sup>a</sup>Raters A–G were from the University of Iowa and raters 1–4 were from Colorado State University. Raters F and 3 were faculty.