Physical Ergonomics and Human Factors, Vol. 185, 2025, 69–76
https://doi.org/10.54941/ahfe1006465

Influence of Workstation Design on Intra-Rater Reliability in Sculpting Kneading Task Assessments

Patricia Eugenia Sortillón González[1], Aide Aracely Maldonado Macías[1], David Saénz Zamarrón[2], Juan Luis Hernández Arellano[1], and Enrique Javier De la Vega Bustillos[3]

[1]Department of Electrical Engineering and Computing, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Chihuahua 32310, México
[2]Electronics Department, Instituto Tecnológico de México, Ciudad Madero, Chihuahua, Chihuahua 31500, México
[3]Department of Industrial Engineering, Instituto Tecnológico de México, Hermosillo, Sonora, 83100, México

ABSTRACT

Intra-rater reliability is a crucial metric for evaluating the consistency of ratings made by the same individual across repeated trials. This study examines intra-rater reliability for posture assessment during a kneading sculpting task on worktables titled at 0° and 25°, and analyzes how table tilt and other factors affect the evaluation consistency. Thirty-nine experienced sculptors performed the tasks under both conditions in a randomized sequence. One rater assessed upper arm, lower arm and wrist postures across three trials. Postures were assessed using the Rapid Upper Limb Assessment method. Reliability and performance differences using the two table tilts were analyzed using Weighted Cohen's Kappa coefficient (WCK), Wilcoxon signed-rank test, and Bland-Altman plots. Results showed high intra-rater reliability, with WCK averages of 0.78 (0° tilt) and 0.85 (25° tilt). Bland-Altman plots indicated strong agreement, with mean differences near zero and slightly narrower limits of agreement for the 2° tilt, suggesting greater consistency. Minor systematic bias appeared under the 0° tilt, likely due to higher ergonomic demand. The Wilcoxon test revealed significant differences in posture scores between trials for both tilts ($p<0.05$). These findings underscore the impact of workstation design on intra-rater reliability, highlighting the value of ergonomic integration in manual tasks. Reliability remained high across tilts, with slight variations linked to ergonomic factors. Future studies should include diverse participants, inter-rater comparisons, and varied task types to strengthen generalizability across manual work scenarios.

Keywords: Intra-rater, Reliability, Sculpting tasks, Workstation design

INTRODUCTION

Workstation design plays a critical role in ensuring task efficiency and worker well-being, particularly in manual tasks requiring repetitive movements (Kahn and Pope, 2015). Proper workstation design has been shown to

promote natural postures, decrease unnecessary muscle exertion and improve task accuracy (Bai, Kamarudin and Alli, 2024). Furthermore, consistency in task execution is closely tied to ergonomic design, as stable and comfortable postures enable workers to maintain focus and repeat movements (El-Sherbeeny et al., 2023). Ergonomic assessments are essential tools for identifying risk factors associated with prolonged postures, repetitive movements, and force exertion (Hoe et al., 2018). In addition, in the field of sculpting, ergonomic assessments and workstation interventions ensure safer work, reducing injury risks and improving productivity.

Intra-rater reliability, which measures the consistency of assessments made by the same individual across multiple trials, is a key factor in evaluating the accuracy and reproducibility of ergonomic assessments. By analysing how intra-rater reliability might vary across different workstation setups, researchers can identify ergonomic conditions that minimize postural variability and improve assessment consistency. Investigating intra-rater reliability variations can help determine whether certain workstation configurations lead to assessment inconsistencies, revealing potential limitations in ergonomic evaluation methods. Studying how workstation design affects intra-rater reliability can help validate the generalizability of ergonomic evaluation techniques across diverse work environments. This is particularly important in workplace ergonomics, where reliable data is needed to identify risk factors and design effective interventions.

While previous research (Ghesmaty, Gustafson and Cavuoto, 2016; Commissaris et al., 2016) has explored the impact of workstation design on task performance and ergonomics, there remains a gap in investigating intra-rater reliability in ergonomic assessments, particularly in tasks involving repetitive movements. While most studies have focused on inter-rater reliability (Dart et al., 2009, Cann et al., 2008), there has been little attention paid to how a single evaluator's assessments remain consistent across repeated trials under different ergonomic conditions. Variations in worktable tilt can alter joint angles, visibility of body segments, and perceived postural alignment, factors that may subtly influence how raters interpret and score posture during repeated evaluations. This study examines intra-rater reliability for kneading sculpting tasks performed on worktables with two different tilt angles: 0° and 25°. By analyzing the impact of workstation inclination on assessment consistency, this research aims to provide insights into the role of ergonomic factors in shaping reliability. This study utilizes the Rapid Upper Limb Assessment (RULA) method to conduct the ergonomic assessment and employs the Weighted Cohen's Kappa (WCK) coefficient, the Wilcoxon signed-rank test, and the Bland-Altman plots to assess the intra-rater reliability and elucidate the relationship between workstation design and reliability in ergonomic evaluations.

METHODOLOGY

This study employed a repeated-measures design, with professional sculptors performing a standardized kneading sculpting task under two workstation conditions. A trained rater assessed upper arm, lower arm, and wrist,

postures using the RULA method across three separate trials for each condition. The following section provides a detailed overview of the participants, study design, assessment procedures, and statistical methods employed to ensure reliability analysis.

Participants

A total of 39 sculptors participated in this study. To ensure consistency in skill level, participants were required to have at least one year of experience in sculpting. Additionally, all participants should be over 18 years old. Individuals with diagnosed musculoskeletal disorders or prior injuries were excluded from the study. The study was approved by the Sculptors Association Ethics Committee (CE-2024-001). Prior to participation, all individuals provided informed consent, acknowledging their voluntary participation. The participants were informed about the study's objectives, including the sculpting task, the workstation conditions, and the posture assessment process.

Study Design and Workstation Conditions

This study employed a repeated-measures design to assess the intra-rater reliability when assessing kneading sculpting postures under two distinct workstation conditions (worktable inclinations of 0° and 25° degrees). In this design, each participant completed the kneading sculpting task under both conditions. The participants completed three trials of kneading tasks (40 s) in both workstation settings. The order in which participants performed the tasks was randomized, ensuring that the results were not biased by the sequence of tasks.

Rater and Posture Assessment

One rater with training in ergonomic assessment methodologies evaluated participants' postures during the kneading task. The rater was selected based on their expertise in ergonomic posture evaluation, ensuring that he was familiar with the application of the RULA scoring system. The rater scored the participant's postures across three separate trials conducted under both workstation conditions. The sculpting task was performed first under one condition, followed by a brief rest period to reduce fatigue, after which participants completed the same task under the alternate workstation condition.

Data Collection and Analysis

The RULA posture scores were recorded in a data base for further analysis. The WCK was employed to quantify the intra-rater reliability. To assess the level of agreement between repeated measurements and detect any systematic biases, the Bland-Altman plots were generated. In addition, the Wilcoxon signed-rank test was performed to determine the presence of statistically significant differences of the postural scores on paired trials.

RESULTS

The study applied statistical techniques to examine variations in assessment agreement and potential biases introduced by worktable inclination. The results offer insights into the extent to which intra-rater reliability is affected by workstation design. According to Table 1, the overall WCK value ranged from 0.721 to 0.966. However, a slight improvement in agreement was observed for the 25° worktable tilt condition (0.756–0.966) compared to the 0° worktable tilt condition (0.721–0.879), suggesting that ergonomic adjustments may have positively influenced posture consistency.

Table 1: Weighted Cohen's Kappa Intra-rater reliability.

Body Posture	0° Worktable Tilt			25° Worktable Tilt		
	K	SD	CI (95%)	K	SD	CI (95%)
Upper arm	0.879*	0.027	[0.827, 0.931]	0.966*	0.017	[0.933, 0.999]
Lower arm	0.742*	0.036	[0.742, 0.813]	0.829*	0.026	[0.778, 0.879]
Wrist	0.721	0.017	[0.688, 0.754]	0.756*	0.025	[0.707, 0.805]

*P-value<0.05

The Bland-Altman plots demonstrated that the mean difference between repeated measurements was nearly zero in both workstation conditions, indicating strong intra-rater reliability for all evaluated postures. However, the limits of agreement (LoA) were slightly narrower for the 25° tilt worktable in all postures, suggesting more consistent assessments under the 25° tilt worktable condition (see Figure 1 for upper arm posture).

Despite this minor systematic bias, particular to the 0° tilt worktable condition, where participants exhibited a tendency to adjust their posture more frequently, likely due to increased physical strain, it was found better agreement levels in posture assessments compared to the 25° tilt worktable. To determine whether there were statistically significant differences in postural scores in the two table inclinations, a Wilcoxon signed-rank test was conducted for all postures in both worktable tilt conditions (see Table 2).

Table 2: Wilcoxon signed-rank test for posture scores across trials and worktable tilts.

Paired Comparations	0° Worktable Tilt			25° Worktable Tilt		
	First Md	Second Md	W	First Md	Second Md	W
Trials 1–2	2.5	3	45.691*	1	2	98.336*
Trials 1–3	2	4	57.132*	1.5	2	125.45*
Trials 2–3	1.5	3	98.361*	1.5	2	113.54*

*P-value<0.05

All pairwise comparisons were significant, regardless of worktable tilt. Median RULA scores differed significantly between trials 1, 2, and 3. Although the results show significance within each worktable tilt, absolute median scores in the 25° worktable tilt produced lower scores than those in the 0° worktable tilt.

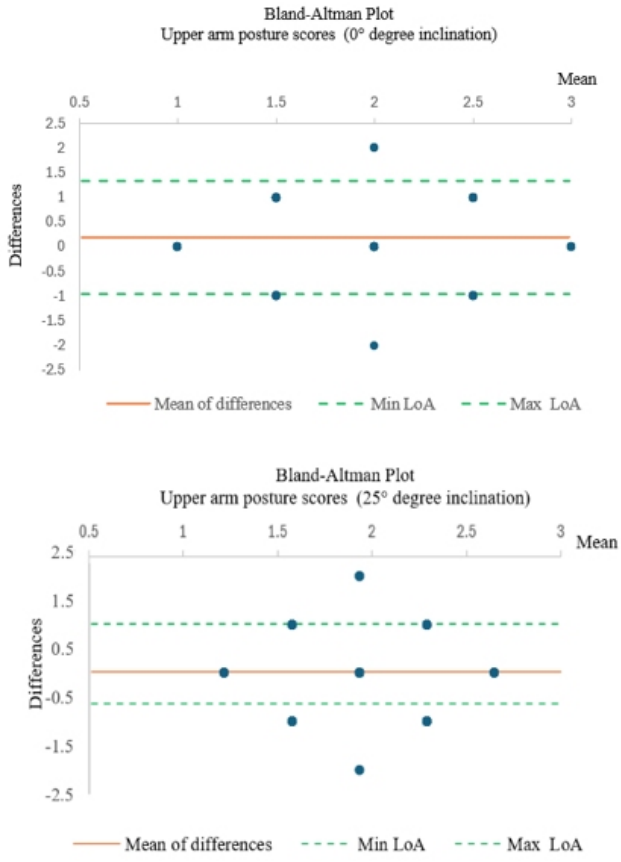

Figure 1: Upper arm posture scores Bland-Altman Plots for 0° and 25° worktable tilts.

DISCUSSION

The results of this study underscore the impact of workstation design on the intra-rater reliability during ergonomic assessments of kneading tasks. The findings reveal a high degree of intra-rater reliability observed across both table tilts (an average WCK = 0.78 for 0° worktable tilt and an average WCK = 0.85 for 25° worktable tilt). This indicates that ergonomic evaluations using RULA were consistent within the rater. However, the slight increase in reliability for the 25° worktable tilt suggests that workstation ergonomics can impact assessment consistency, likely due to improved posture stabilization and reduced upper limb strain.

The presence of minor systematic biases in the 0° worktable tilt condition suggests that the 0° worktable tilt may pose higher ergonomic challenges, potentially leading to greater variability in posture scoring across trials.

The significant differences found in Wilcoxon signed-rank test underscore the importance of task variation and workstation adjustments. These findings provide evidence that both the specific trial tasks (Trials 1, 2, and 3) and the worktable tilt influence postural outcomes. Further research could examine the magnitude of these differences and their practical significance

for ergonomic interventions. Conversely, the slightly better reliability scores in the 25° tilt worktable indicate that minor ergonomic modifications can enhance task consistency.

Despite the valuable insights gained from this study, several limitations should be acknowledged: The study was conducted with thirty-nine participants. While this helped control for skill-related variability, the findings may not be fully generalizable to individuals with different levels of expertise. Future research should explore a broader participant pool. The study focused exclusively on a kneading sculpting task, representing only one aspect of manual sculpting work. Other sculpting activities, such as carving, molding, may involve different ergonomic demands and could influence intra-rater reliability differently. Posture assessments were conducted during a single experimental session, which limits the ability to assess the long-term effects of workstation design. The assessments were performed in a controlled environment, ensuring consistent lighting, camera angles, and observer positioning. While this approach minimized external variability, it may not fully reflect real-world sculpting conditions, where environmental factors such as workspace constraints, and tool variations, could influence posture and assessment reliability. While the rater assessed postures across repeated trials, unintentional biases may still have influenced scoring, despite efforts to maintain consistency. The study only examined two specific worktable tilts (0° and 25°). While these conditions provided meaningful comparisons, the results may not apply to workstations with different tilt angles. To address these limitations, future studies should: a) include a more diverse participant sample, covering different experience levels; b) examine additional sculpting techniques to assess how different tasks influence ergonomic assessments; c) conduct longitudinal studies to evaluate long-term effects of workstation design; d) integrate advanced biomechanical tools, such as motion tracking or muscle activity measurements, to complement posture scoring methods. By addressing these limitations, future research can further enhance the understanding of how workstation design influences assessment reliability in manual sculpting and other precision-based tasks.

While intra-rater reliability has been examined in previous research (Zumana et al., 2019; Smith et al., 2013; Dart et al., 2009), the contexts of these studies differ from the present work. Many prior studies (Morgan et al., 2023; Duits and Kempes, 2022; Ramsey et al., 2022; Zhang et al., 2022,) have focused on clinical settings, industrial work, or general occupational tasks, whereas this study specifically evaluates sculpting-related postures, which involve unique ergonomic demands.

CONCLUSION

This study highlights the importance of workstation design in shaping intra-rater reliability during ergonomic assessments of kneading tasks. The high intra-rater reliability observed across both worktable tilts suggests that, within a single rater, RULA-based ergonomic evaluations remain consistent. The improved reliability at a 25° tilt worktable suggests that workstation modifications may enhance postural stability and reduce variability in

assessments by minimizing upper limb strain. The minor systematic biases observed under the 0° tilt condition suggest greater ergonomic demands that could affect assessment consistency. However, this study's findings should be interpreted with caution due to its use of only one rater. While the results provide useful insights into how workstation design may influence intra-rater reliability, the lack of inter-rater data limits the broader applicability of these conclusions. Future research should include multiple raters to evaluate the consistency of findings across observers and further validate the impact of ergonomic factors on assessment reliability. Moreover, integrating objective tools such as motion-tracking technologies may complement observational methods like RULA and reduce potential subjective biases. Expanding the participant pool, incorporating different sculpting techniques, and testing in naturalistic settings will also help improve the generalizability of results and provide a deeper understanding of how workstation design affects ergonomic evaluation intra-rater reliability.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the rater for their dedication in conducting the posture evaluations. We would also like to extend our appreciation to the sculptors for their time and effort. Additionally, we acknowledge the support of our colleagues and institutions for providing the necessary resources and facilities for this study.

REFERENCES

Bai, Y., Kamarudin, K. M. and Alli, H. (2024) 'A systematic review of research on sitting and working furniture ergonomic from 2012 to 2022: Analysis of Assessment Approaches', *Heliyon*, 10(7). doi: 10.1016/j.heliyon. 2024.e28384.

Cann, A. P. et al. (2008) 'Inter-rater reliability of output measures for a posture matching assessment approach: A pilot study with food service workers', *Ergonomics*, 51(4), pp. 556–572. doi:10.1080/00140130701711455.

Commissaris, D. et al. (2014) 'Effects of a standing and three dynamic workstations on computer task performance and Cognitive Function tests', *Applied Ergonomics*, 45(6), pp. 1570–1578. doi: 10.1016/j.apergo.2014.05.003.

Dartt, A. et al. (2009) 'Reliability of assessing upper limb postures among workers performing manufacturing tasks', *Applied Ergonomics*, 40(3), pp. 371–378. doi: 10.1016/j.apergo.2008.11.008.

Duits, N. and Kempes, M. (2023) 'Interrater and intra-rater reliability of the vera-2r tool', *Frontiers in Psychiatry*, 14. doi: 10.3389/fpsyt.2023.1236295.

El-Sherbeeny, A. M. et al. (2023) 'How is job performance affected by ergonomics in the tourism and hospitality industry? mediating roles of work engagement and talent retention', *Sustainability*, 15(20), p. 14947. doi: 10.3390/su152014947.

Ghesmaty Sangachin, M., Gustafson, W. W. and Cavuoto, L. A. (2016) 'Effect of active workstation use on workload, task performance, and postural and physiological responses', *IIE Transactions on Occupational Ergonomics and Human Factors*, 4(1), pp. 67–81. doi: 10.1080/21577323.2016.1184196.

Hoe, V. C. et al. (2018) 'Ergonomic interventions for preventing work-related musculoskeletal disorders of the upper limb and neck among office workers', *Cochrane Database of Systematic Reviews*, 2018(10). doi: 10.1002/14651858.cd008570.pub3.

Khan, M. and Pope-Ford, R. (2015) 'Improving and modifying the design of workstations within a manufacturing environment', *Procedia Manufacturing*, 3, pp. 4927–4934. doi: 10.1016/j.promfg.2015.07.631.

Morgan, R. et al. (2023) 'The functional movement screen: Exploring interrater reliability between raters in the updated version', *International Journal of Sports Physical Therapy*, 18(3). doi: 10.26603/001c.74724.

Ramsey, D. et al. (2022) 'Revised hammersmith scale for spinal muscular atrophy: Inter and intra-rater reliability and agreement', *PLOS ONE*, 17(12). doi: 10.1371/journal.pone.0278996.

Smith, C. A. et al. (2013) 'Interrater and intra-rater reliability of the functional movement screen', *Journal of Strength and Conditioning Research*, 27(4), pp. 982–987. doi: 10.1519/jsc.0b013e3182606df2.

Zhang, Y. H. et al. (2022) 'Intra-rater and inter-rater reliability of pressure pain threshold assessment in stroke patients', *European Journal of Physical and Rehabilitation Medicine*, 58(4). doi: 10.23736/s1973-9087.22.07378–6.

Zumana, N. et al. (2019) 'Intra-rater and inter-rater reliability of six musculoskeletal preparticipatory screening tests', *South African Journal of Physiotherapy*, 75(1). doi:10.4102/sajp. v75i1.469.