Study of the Reliability of Rapid Upper Limb Assessment (RULA) as Method of Assessment of Sculptor's Posture

Patricia Eugenia Sortillón-González,
Aidé Aracely Maldonado-Macías, Juan Luis Hernández-Arellano,
David Saénz-Zamarron and Enrique Javier De la Vega-Bustillos

December 2, 2024

Study of the reliability of rapid upper limb assessment (RULA) as method of assessment of sculptor's posture

Patricia Eugenia Sortillón González[1][0000-0002-9716-6693], Aidé Aracely Maldonado Macías[2][0000-0002-4959-161X], Juan Luis Hernández Arellano[1][0000-0002-8612-5132], David Saénz Zamarrón[2][0000-0002-5045-4997], Enrique Javier de la Vega Bustillos[3][0000-0002-0761-6858]

[1] Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Chihuahua, 32310, México
[2] Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Chihuahua, 32310, México
[3]TECNM/Instituto Tecnológico de Hermosillo.
a216733@alumnos.uacj.mx,amaldona@uacj.mx,luis.hernadez@uacj.mx, dsaenz@itcdcuauhtemoc.edu.mx, e_delavega_mx@yahoo.mx

Abstract. Observational methods have been applied to assess work-related musculoskeletal disorders (WMSDs). The reliability of their results is a challenge. WMSDs remain a significant concern for workers, emphasizing the need for reliable observation tools. This study aimed to investigate the inter-rater reliability (IRR) of the Rapid Upper Limb Assessment (RULA) in assessing the posture of sculptors performing three tasks (sifting, kneading, and modeling). A fully crossed random effects model was applied to analyze IRR across thirty-nine subjects, three tasks, and three raters. 2340 postures were scored individually by each rater. The IRR for each pair of raters was calculated using the weighted Cohen's Kappa statistic. The highest IRR values among the rater pairs were found in the kneading task when assessing the trunk (1-2: IRR=0.875; 95% CI [0.849-0.900], 1-3: IRR=0.887; 95% CI [0.878-0.896], and 2-3: IRR=0.934; 95% CI [0.923-0.945]). The worst IRRs were obtained for the wrist assessment (1-2: IRR=0.168; 95% CI [0.159-0.177], 1-3: IRR=0.145; 95% CI [0.138-0.152], and 2-3: IRR=0.158; 95% CI [0.148-0.168]). The IRRs differed among tasks. Raters demonstrated higher IRR when scored upper arm, trunk, and leg postures, but fair reliability was observed for neck and wrist postures. Raters showed better IRR values when assessing postures for the sifting and modeling tasks than for the kneading task. This underscores the need for proper and adequate training before using RULA. In addition, task complexity seems to be an important aspect of IRR in this study. Further studies are needed to analyze how reliability is affected when analyzing different tasks.

Keywords: Observation tools, agreement, inter-rater.

1 Introduction

Work-related musculoskeletal disorders (WMSDs) are primary contributors to injuries and long-term disabilities [1]. Organizations and governments have invested considerable resources in occupational health claims and research to develop solutions to the problem [2]. WMSDs are multifactorial in nature and are associated with exposure to specific risk factors [3,4]. A major challenge in the assessment of risk factors is their reliability [5]. The Rapid Upper Limb Assessment (RULA) is a widely used observational method for assessing risk factors associated with WMSDs and has been shown to be highly reliable.[6]. The Center for Disease Control and Prevention in the USA defines reliability as the consistency of measurements for a given variable [7]. One of the most popular coefficients to evaluate inter-rater reliability (IRR) when using categorical data is Weighted Cohen's kappa (WCK) [8]. The following guidelines have been suggested for interpreting the strength of agreement based on the kappa coefficient: 0.00 indicates poor agreement, 0.01 to 0.20 signifies slight agreement, 0.21 to 0.40 represents fair agreement, 0.41 to 0.60 suggests moderate agreement, 0.61 to 0.80 indicates substantial agreement, and 0.81 to 1 reflects almost perfect agreement. [9]. When assessing the level of agreement between two raters on ordered categorial variables with three or more categories, quadratic weighted Kappa is a commonly used [10]. For example, a study examining the reliability of the RULA in bakers reported excellent IRR. A study evaluated the reliability of the RULA in bank employees and showed an excellent IRR [11], similarly, excellent IRR results have been observed in studies evaluating manufacturing tasks [12]. WMSDs continue to be a major concern due to their prevalence and impact on worker health and productivity. Sculpting tasks, like manual activities, often involve prolonged or repetitive postures that can increase the risk of WMSDs. Despite the widespread use of the RULA method in various occupational settings, there is limited research on its reliability when applied to creative occupations such as sculpture, where postures are dynamic and task-specific. While RULA has been validated in other fields such as banking and manufacturing, its reliability in assessing sculpting tasks remains unexplored. Sculpting often presents unique ergonomic challenges due to prolonged awkward postures and fine motor movements. The consistency of rating between raters is critical to ensure reliable and actionable data. Assessing the IRR using weighted Cohen's Kappa provides insight into the robustness of the method for analyzing specific postural risks in sculpting tasks. Accurate assessment of risk is essential for designing ergonomic interventions and reducing the likelihood of WMSDs in sculptors. This study could provide a basis for the development of tailored solutions to improve occupational health in creative industries. This research will fill a critical gap in the literature by evaluating the applicability and reliability of the RULA method in creative occupational context, thereby supporting the development of evidence-based ergonomic practices. The objective of this research was to evaluate the IRR of the RULA method when used by three raters to assess postural risk factors in three specific sculpting tasks: kneading, sifting and modeling. The study aimed to determine the consistency of RULA assessments across raters and to identify potential challenges and limitations in applying this method in the context of sculpting, a profession characterized by dynamic and tasks-specific postures.

2. Methods

Three ergonomists evaluated six postures using the RULA method [6], including four limb-specific postures (upper arm, lower arm, wrist and leg, on both the left and right sides) and two core postures (neck and trunk), across three sculpting tasks: kneading, sifting, and modeling.

Thirty-nine sculptors from northern México participated in this study. Participants were selected based on specific inclusion criteria: 1) they had to be over eighteen years of age, 2) they had to have at least one year of experience in their current profession, 3) they had to work at least six hours a day, and 4) they could not be pregnant or undergoing medical treatment. Ethical approval (CEI-2022-01) was granted by the Art Association Committee. The selected sculptors (n=39) provided their consent to participate by signing an informed consent form. Demographic information was collected from participants. Sculpting tasks were recorded using video and digital photography resources (4K camcorder 10X optical zoom camera Ultra HD Wi-Fi video camera). Raters were provided with a posture assessment training program. One hundred and seventeen video frames from the three tasks were individually rated by each rater using the RULA method. A total of 2340 ratings were made by each rater. All data was securely stored in computer files protected by passwords. A fully crossed random effects model was applied to analyze the IRR across thirty-nine subjects, three tasks, and three raters. Data management and statistical analysis were conducted using the IBM SPSS software (v24, SPSS Inc., Chicago, USA). The WCK test was performed to assess the IRR between each pair of raters. A 5% significance level was used.

3. Results

Among the thirty-nine sculptors studied, 70% were male, all were right-handed, and their ages varied from 24 to 62 years with and average age of 37.23 ± 11.10 years. All raters were male, the average age was 35.8 +- 0.33 years, and they had ten years of an experience in ergonomics. The WCK IRR coefficients for the kneading, sifting, and modeling tasks are shown in Table 1 with corresponding standard deviations, confidence intervals (95%), and p-value<0.05.

According to Table 1, the highest levels of IRR between pair of raters on the sifting task were found for the leg, lower arm, and upper arm, except for the wrist, with values ranging from 0.711 to 0.899. In the modeling task, the highest IRRs among pairs of raters (2-3) for the neck (both right and left sides) were as follows IRR=0.895; 95% CI [0.855-0.935] for the right side and IRR=0.881; 95% CI [0.855-0.907] for the left side, reflecting almost perfect in these ratings.

Table 1. Weighted Kappa Coefficients by Task and Body Posture.

Task	Pair of raters	Body Posture	Right side				Left side			
			K	SD	CI (95%)	p-value	K	SD	CI (95%)	p-value
Kneading	1-2	Leg	0.753	0.052	[0.676 , 0.830]	*	0.852	0.023	[0.814 , 0.890]	*
	1-3	Leg	0.821	0.017	[0.794 , 0.848]	*	0.798	0.019	[0.768 , 0.828]	
	2-3	Leg	0.857	0.018	[0.827 , 0.887]	*	0.789	0.025	[0.750 , 0.828]	*
	1-2	Lower arm	0.561	0.018	[0.541 , 0.581]	*	0.658	0.025	[0.626 , 0.690]	*
	1-3	Lower arm	0.654	0.018	[0.631 , 0.677]		0.701	0.021	[0.672 , 0.730]	
	2-3	Lower arm	0.598	0.013	[0.583 , 0.613]	*	0.699	0.012	[0.683 , 0.715]	*
	1-2	Neck	0.245	0.021	[0.235 , 0.255]	*	0.168	0.021	[0.161 , 0.175]	
	1-3	Neck	0.248	0.012	[0.242 , 0.254]	*	0.179	0.012	[0.175 , 0.183]	*
	2-3	Neck	0.248	0.011	[0.243 , 0.253]		0.183	0.011	[0.179 , 0.187]	
	1-2	Trunk	0.875	0.015	[0.849 , 0.901]	*	0.845	0.015	[0.820 , 0.870]	
	1-3	Trunk	0.887	0.005	[0.878 , 0.896]	*	0.833	0.005	[0.825 , 0.841]	
	2-3	Trunk	0.934	0.006	[0.923 , 0.945]	*	0.888	0.006	[0.878 , 0.898]	*
	1-2	Upper arm	0.698	0.003	[0.694 , 0.702]		0.586	0.017	[0.566 , 0.606]	
	1-3	Upper arm	0.754	0.012	[0.736 , 0.772]	*	0.812	0.015	[0.788 , 0.836]	*
	2-3	Upper arm	0.598	0.013	[0.583 , 0.613]		0.645	0.023	[0.616 , 0.674]	
	1-2	Wrist	0.168	0.026	[0.159 , 0.177]	*	0.258	0.025	[0.245 , 0.271]	*
	1-3	Wrist	0.145	0.024	[0.138 , 0.152]	*	0.356	0.017	[0.344 , 0.368]	*
	2-3	Wrist	0.158	0.033	[0.148 , 0.168]		0.285	0.045	[0.260 , 0.310]	
Sifting	1-2	Leg	0.854	0.052	[0.767 , 0.941]	*	0.752	0.025	[0.715 , 0.789]	*
	1-3	Leg	0.796	0.017	[0.769 , 0.823]	*	0.699	0.022	[0.669 , 0.729]	
	2-3	Leg	0.899	0.017	[0.869 , 0.929]	*	0.852	0.018	[0.822 , 0.882]	*
	1-2	Lower arm	0.856	0.036	[0.796 , 0.916]	*	0.758	0.013	[0.739 , 0.777]	*
	1-3	Lower arm	0.896	0.028	[0.847 , 0.945]		0.759	0.025	[0.722 , 0.796]	
	2-3	Lower arm	0.852	0.029	[0.804 , 0.900]	*	0.759	0.025	[0.722 , 0.796]	*
	1-2	Neck	0.759	0.017	[0.734 , 0.784]		0.725	0.017	[0.701 , 0.749]	*
	1-3	Neck	0.786	0.019	[0.757 , 0.815]	*	0.765	0.019	[0.737 , 0.793]	
	2-3	Neck	0.774	0.027	[0.733 , 0.815]	*	0.785	0.027	[0.743 , 0.827]	*
	1-2	Trunk	0.785	0.017	[0.759 , 0.811]		0.774	0.017	[0.748 , 0.800]	*
	1-3	Trunk	0.711	0.015	[0.690 , 0.732]	*	0.715	0.015	[0.694 , 0.736]	*
	2-3	Trunk	0.725	0.018	[0.699 , 0.751]	*	0.716	0.018	[0.691 , 0.741]	
	1-2	Upper arm	0.874	0.018	[0.843 , 0.905]	*	0.698	0.015	[0.677 , 0.719]	
	1-3	Upper arm	0.789	0.022	[0.755 , 0.823]		0.785	0.014	[0.763 , 0.807]	*
	2-3	Upper arm	0.854	0.022	[0.817 , 0.891]	*	0.856	0.016	[0.829 , 0.883]	*
	1-2	Wrist	0.152	0.025	[0.145 , 0.159]		0.147	0.032	[0.138 , 0.156]	*
	1-3	Wrist	0.125	0.023	[0.119 , 0.131]	*	0.155	0.025	[0.147 , 0.163]	
	2-3	Wrist	0.123	0.027	[0.116 , 0.130]	*	0.166	0.014	[0.161 , 0.171]	*
Modeling	1-2	Leg	0.795	0.045	[0.725 , 0.865]	*	0.852	0.023	[0.814 , 0.890]	
	1-3	Leg	0.896	0.017	[0.866 , 0.926]		0.758	0.025	[0.721 , 0.795]	
	2-3	Leg	0.875	0.066	[0.762 , 0.988]	*	0.752	0.028	[0.711 , 0.793]	*
	1-2	Lower arm	0.897	0.032	[0.841 , 0.953]	*	0.852	0.025	[0.810 , 0.894]	*
	1-3	Lower arm	0.785	0.017	[0.759 , 0.811]	*	0.667	0.025	[0.634 , 0.700]	
	2-3	Lower arm	0.894	0.024	[0.852 , 0.936]		0.725	0.036	[0.674 , 0.776]	
	1-2	Neck	0.895	0.023	[0.855 , 0.935]	*	0.785	0.023	[0.750 , 0.820]	*
	1-3	Neck	0.862	0.051	[0.776 , 0.948]		0.799	0.051	[0.719 , 0.879]	
	2-3	Neck	0.887	0.015	[0.861 , 0.913]	*	0.881	0.015	[0.855 , 0.907]	*
	1-2	Trunk	0.789	0.034	[0.736 , 0.842]		0.699	0.034	[0.652 , 0.746]	
	1-3	Trunk	0.759	0.029	[0.716 , 0.802]	*	0.723	0.029	[0.682 , 0.764]	
	2-3	Trunk	0.698	0.029	[0.658 , 0.738]	*	0.675	0.029	[0.637 , 0.713]	*
	1-2	Upper arm	0.881	0.025	[0.838 , 0.924]	*	0.725	0.026	[0.688 , 0.762]	
	1-3	Upper arm	0.845	0.026	[0.802 , 0.888]		0.715	0.017	[0.691 , 0.739]	*
	2-3	Upper arm	0.822	0.037	[0.762 , 0.882]	*	0.736	0.018	[0.710 , 0.762]	*
	1-2	Wrist	0.168	0.052	[0.151 , 0.185]	*	0.215	0.047	[0.195 , 0.235]	*
	1-3	Wrist	0.158	0.017	[0.153 , 0.163]		0.189	0.015	[0.183 , 0.195]	*
	2-3	Wrist	0.154	0.06	[0.136 , 0.172]	*	0.178	0.008	[0.175 , 0.181]	*

* p-value < 0.05

On the other hand, the lowest IRR values were observed for the kneading task, where IRR values of 0.245 (95% CI [0.235-0.255] and 0.248 95% CI [0.242-0.254] indicated fair reliability. These lower IRR values suggest that there may have been more variability in the scores for this task, possibly due to the subjective nature of assessing kneading techniques or differences in rater interpretation. This contrast in reliability between tasks highlights the importance of task-specific training and clearer guidelines for raters to improve consistency in scoring.

4. Discussion

According to results, the kneading task showed the lowest IRR values for the wrist and neck postures. The assessment of these two specific postures had two limitations: one related to clothing obstructions and the other related to hair. Fair reliability due to smaller joint movements has been discussed as a contributing factor in other studies [13,14]. A study examining posture in children using a computer [6] found that the IRR was higher for the trunk and legs. Another study reported RULA had high reliability when used by therapists to assess adults [12]. Another study examining the interobserver repeatability found moderate to good reliability [15]. These reliability values were significantly lower than those found in the current study, although a direct comparison is difficult due to differences in the types of data analyzed (ordinal versus continuous). Additional factors may include differences in the observational methods used; this study used video recordings, whereas other research has relied on direct observation of work tasks. In this study, potential sources of error could contribute to variability in the data. The design of the study, such as camera angles and lighting conditions in which the sculptors worked, could affect the accuracy of posture assessments. Poor visibility or obstructions could bias measurements. Finally, certain modeling tasks may inherently lead to more difficult or less consistent judgements. For example, dynamics movements such as wrist and neck postures may be more difficult to reliably observe than static postures, which could lead to errors. The results suggest that when using the RULA method to assess sculpting tasks, inter-rater reliability between pairs of raters differs depending on the task assessed.

5. Conclusions

This study evaluated the IRR of the RULA method for assessing postures during sculpting tasks. Challenges in assessing wrist and neck postures, influenced by factors such as clothing obstructions and hair, highlight the complexity of assessing small joint movements and dynamic postures. The results of this study are consistent with previous research showing that the trunk and leg postures are easier to assess than upper limb postures. While RULA demonstrated moderate reliability, the study highlights the importance of methodological considerations, such as camera positioning and task characteristics, in achieving accurate assessments. This study provides a foundation for improving observational methods in posture assessment, particularly for tasks involving complex or constrained postural variations. Given the challenges in assessing postures, future research should explore complementary methods, such as wearable sensor or three-dimensional motion capture, to enhance the accuracy of posture assessments, especially for small joint movements. Further research should explore the applicability of the RULA method across various populations and tasks involving complex or constrained movements.

References

1. Punnett L, Wegman DH.: Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. Journal of Electromyography and Kinesiology. Feb;14(1),13–23 (2004).
2. Soto M. IMF engagement on health spending issues in surveillance and program work. Technical Notes and Manuals. Dec 2023(008) (2023).
3. Yang F, Di N, Guo W, Ding W, Jia N, Zhang H.: The prevalence and risk factors of work-related musculoskeletal disorders among electronics manufacturing workers: A cross-sectional analytical study in China. BMC Public Health. Jan 3;23(1) (2023).
4. Das D, Kumar A, Sharma: M. A systematic review of work-related musculoskeletal disorders among handicraft workers. International Journal of Occupational Safety and Ergonomics. May 31;26(1),55–70 (2018).
5. Graben PR, Schall MC, Gallagher S, Sesek R, Acosta-Sojo Y.: Reliability Analysis of observation-based exposure assessment tools for the Upper Extremities: A systematic review. International Journal of Environmental Research and Public Health. Aug 25;19(17):10595 (2022).
6. McAtamney, L., Corlett, E.N., RULA: a survey method for the investigation of work-related upper limb disorders. Applied Ergonomics 24 (2), 91–99 (1993).
7. Centers for Disease Control and Prevention, 2001. Guide to Evaluating the Effectiveness of Strategies for reventing Work Injuries: How to ShowWhether a Safety Intervention Really Works. Department of Health and Human Services, Centers for Disease Control and Prevention, and National Institute for Occupational Safety and Health. DHHS (NIOSH) Publication No.2001-119.
8. Li, M., Gao, Q. & Yu, T.: Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. *BMC Cancer* 23, 799 (2023).
9. Landis, J. R., & Koch, G. G.: The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174 (1977).
10. Vanbelle S.: A new interpretation of the weighted kappa coefficients. Psy-chometrika.81(2):399–410 (2016).
11. Kumar A, Kamath S.: Rapid upper limb assessment (RULA): Validity and reliability evidence in identifying workplace ergonomics among bank employee's using computers. Revista Pesquisa em Fisioterapia. May 28;9(2):194–203 (2019).
12. Dartt A, Rosecrane J, Gerr F, Chen P, Anton D, Merlino L.: Reliability of assessing upper limb postures among workers performing manufacturing tasks. Applied Ergonomics. May;40(3):371–8 (2009).
13. Keyserling, W.M.: Postural analysis of the trunk and shoulders in simulated real time. Ergonomics 29 (4), 569–583 (1986).
14. Stetson, D.S., Keyserling, W.M., Silverstein, B.A., Leonard, J.A.: Observational analysis of the hand and wrist: a pilot study. Applied Occupational Environmental Hygiene 6 (11), 927–937 (1991).
15. Ketola, R., Toivonen, R., Viikari-Juntura, E.: Interobserver repeatability and validity of an observation method to assess physical loads imposed on the upper extremities. Ergonomics 44 (2), 119–131 (2001).