ORIGINAL ARTICLE

# VALIDITY AND INTER-RATER RELIABILITY OF POSTURAL ANALYSIS AMONG NEW RATERS

**Widyanti A**

*Department of Industrial Engineering, Institut Teknologi Bandung, Labtek 3 Ganesa 10 Bandung, Indonesia*

*** Corresponding author: Widyanti A**
**Email: widyanti@mail.ti.itb.ac.id**

## ABSTRACT

*Work posture analysis is crucial in observing and reducing work-related musculoskeletal symptoms in the workplace. However, in a developing country, new raters are commonly assigned to conduct postural analysis to save on cost. This study aims to observe the validity and inter-rater reliability (defined as the degree of agreement among different raters) among new raters of three different commonly used work posture analysis methods: Rapid Upper Limb Assessment (RULA), Rapid Entire Body Assessment (REBA), and Ovako Workload Assessment System (OWAS). Fifty industrial engineering students, divided into five groups, who received prior training about the use of the methods, participated voluntarily in this study by observing ten different working postures in five different industries: the tofu, military equipment manufacturing, automotive maintenance and service, cracker, and milk-processing industries. One ergonomics expert also observed the working postures. Validity was observed based on the correlation between new raters' ratings and the rating of the ergonomics expert. Inter-rater reliability within one group was calculated using the percentage of agreement and kappa value. The result shows high validity of RULA, REBA, and OWAS among new raters. There are insignificant differences in the inter-rater reliability of new raters among RULA, REBA, and OWAS. The implications of the result are discussed.*

***Keywords:*** *New raters, validity, inter-rater reliability, RULA, REBA, OWAS*

## INTRODUCTION

Musculoskeletal symptoms (MSSs) and musculoskeletal disorders (MSDs) are among the most widespread occupational health issues. MSSs and MSDs have significant consequences both socially[1] and economically due to reduced quality and productivity[2]. These symptoms or disorders have been found in developing and industrialized countries[1-5].

In developing countries, similar pictures of MSD or symptom prevalence have been observed, particularly in recent decades. In Indonesia for example, MSSs or disorders have been reported in several studies. In the health sector, Phedy and Gatam found MSDs among Indonesian dentists[6]. In the agriculture sector, Widyanti reported MSDs among rice farmers[7]. Muslim and Nusbaum reported MSDs among posterior load carrier workers[8], while Widanarko et al. reported MSDs among Indonesian coal mining workers[9].

Postural analysis is crucial in providing a safe workplace, because improper work postures are considered a cause of work-related MSDs or MSSs. Therefore, various studies have analyzed work posture in different industries[10-12], mainly to minimize hazards and accidents related to awkward or hazardous postures[13].

The use of postural analysis has been found in many countries, both in developed and developing countries. Interestingly, these studies have been conducted not only by experts (including ergonomics practitioners, workers, and employers[14]), but also by non-experts. The latter refers to newly trained people with limited experience in work posture analysis (i.e., less than 1 year of experience). Particularly in developing countries such as Indonesia, new raters such as internship students or fresh graduates from bachelor degrees are involved in work posture analysis because of the limitations of certified ergonomics practitioners and to save on cost.

There are three postural analysis methods commonly used in industries: direct measurement, self-reporting, and posture observation[15]. The direct measurement method usually uses specific tools to record motion, such as electro goniometers, potentiometers[16], and Kinect[17]. The direct method's advantage is that it provides valid and reliable data. However, this method requires much effort and time to prepare for measurement and collect data. Special competence is also required to analyze the data.

The second method – the self-reporting method – is conducted using questionnaires filled out by workers or interviews about workers' postures. The advantage of this method is that it is relatively cost effective and can be applied to a large sample. Although the reliability of the self-reporting method is generally low[16], this method has been used in many studies in limited numbers[18]. One example of the self-reported method of analyzing work MSDs is the NORDIC

body map[19], a map of the body parts. Assessment is conducted by the worker marking the level of discomfort in the body parts.

The posture observation method, conducted by estimating posture angles both on site and from recorded videos, has been widely used[11,20]. Takala stated that this method is still the most commonly used by practitioners as it is easy to use and cost effective, and it offers flexibility in data collection[21]. Furthermore, David reviewed the increased number of published methods for observation in recent years[22].

However, the issue of validity (i.e., the extent to which the estimated posture angle is representative of the real angle) must be highlighted in determining the quality of the posture observation method, aside from other issues such as the ease of use[23]. Studies have reported various degrees of validity and reliability of this method. The problem of validity seems to be solved by special precautions in the measurement process[15]. Extensive research has been done on the validity of work posture analysis.

Another crucial problem in using the posture observation method that has recently gained attention is inter-rater reliability (IRR, also called inter-rater agreement). The issue is related to how much different raters agree on a posture angle. As stated by Hallgree[24], IRR is crucial particularly if the data are collected by several raters. IRR provides information about the degree of agreement among raters. Various different results as well as different statistical methods have been found in determining the IRR of posture observation. Several common statistical methods have been used in IRR analysis, including percentage of agreement among raters, kappa statistics, and intraclass correlation coefficient (ICC). The disadvantage of the proportion of agreement is the absence of chance agreement[25]. This is overcome by kappa statistics, which accounts for chance. Kappa can be calculated based on the observed proportion of agreement minus the proportion of agreement that could be expected by chance. In this manner, kappa can avoid erroneous conclusions where agreement is judged to be good, whereas it may simply be due to chance[25]. ICC's disadvantage is complexity, due to the large number of ICC formulas that can be used and the fact that the formulas themselves are complex. The main reason for the complexity is that the ICC is very flexible and can be adjusted for inconsistent raters.

The most used methods of work posture analysis are Ovako Working Posture Analysis System (OWAS[26]), Rapid Entire Body Assessment (REBA[20]), and Rapid Upper Limb Assessment (RULA[27]). OWAS observes the frequency of and time spent on tasks involving the back (4 postures), legs (7 postures), arms (3 postures), and load handled (3 categories). These four factors imply 252 possible combinations of observation results, with a 4-digit code. The next steps of OWAS are the categorization of risk and corrective actions[28]. As stated by Hignett[20], OWAS is commonly used but low in sensitivity.

RULA is used for work-related upper limb symptoms. An assessment can be conducted with the help of several tables indicating various positions of the neck, trunk, and upper limbs. The final value of the RULA assessment shows the level of intervention required to reduce the risks of injury.

Similar to RULA, REBA applies several tables to indicate different loads on parts of the body. Whereas RULA focuses on the upper limbs, REBA emphasizes the entire body. Unpredictable working postures must accommodate entire body posture and load. REBA also proposes action according to risk level.

This study aims to observe the validity and IRR of RULA, REBA, and OWAS among new raters, as an option to overcome the high cost of postural analysis and the limited number of certified ergonomics practitioners, particularly in developing countries such as Indonesia. Although previous studies have compared the three methods (e.g., Bao et al.[15]), this study's novelty is the addition of new raters. Postural analysis is conducted in five industries. Validity is analyzed by comparing results between new raters and an expert. Reliability is analyzed using kappa and percentage of agreement.

## METHODS

Fifty industrial engineering students (mean age = 20,5 years, SD = 1,2 years, 16 female) from the Bandung Institute of Technology participated voluntarily in this study. They were taking ergonomics classes and acted as new raters in the study. The students were trained in the use of RULA, REBA, and OWAS. They were instructed to use Indonesian versions of RULA, REBA, and OWAS to assess work postures in 5 different industries: the tofu, military equipment manufacturing, automotive maintenance and service, cracker, and milk-processing industries. Permits were obtained from these industries to take pictures or videos for postural analysis purposes. Ten postures were chosen in each industry. The choice of the postures was based on group agreement considering the various workloads of workers in each industry.

In addition, one ergonomics expert was involved in this study and conducted postural analysis for the 50 work postures in the five industries. This expert has a PhD in human factors and

ergonomics and more than 20 years of experience in applying ergonomics in various industries.

Validity was analyzed by comparing the median value of the new observers' ratings using RULA, REBA, and OWAS, with the value of the expert's ratings, which were set as a gold standard. IRR was calculated based on percentage of agreement and kappa value. The agreement percentage was calculated by dividing the number of scores of agreement by the total number of scores. Kappa was calculated using the Fleiss kappa formula[29].

## RESULTS

The postural analysis result for each industry was evaluated by each student group. Spearman correlation was used to test the correlation between the observation results of the new raters and the expert in order to test the validity of RULA, REBA, and OWAS among the new raters. Results showed significant correlations between the ratings of the new raters and the expert for RULA ($r = 0.799$, $\rho < 0.01$), REBA ($r = 0.790$, $\rho < 0.01$), and OWAS ($r = 0.802$, $\rho < 0.01$). These results revealed high concurrent validity of RULA, REBA, and OWAS among new raters.

The Wilcoxon signed rank test was applied to observe the difference in ratings between the new raters and the expert. No significant differences were found in RULA, REBA, or OWAS (all $\rho > 005$). In addition, the Wilcoxon signed rank test showed that in general, the new raters gave slightly lower ratings than the ergonomics expert for RULA, REBA, and OWAS.

The percentages of agreement and kappa values of RULA, REBA, and OWAS are shown in Table 1.

***Table 1**: Inter-rater reliability of RULA, REBA, and OWAS*

| Industries | RULA | | REBA | | OWAS | |
|---|---|---|---|---|---|---|
| | % agreement | kappa | % agreement | kappa | % agreement | kappa |
| 1 | 69.40 | 0.30 | 49.40 | 0.23 | 54.50 | 0.38 |
| 2 | 56.74 | 0.42 | 65.23 | 0.57 | 60.34 | 0.47 |
| 3 | 46.03 | 0.11 | 57.03 | 0.19 | 64.00 | 0.40 |
| 4 | 71.58 | 0.00 | 15.86 | 0.04 | 31.40 | 0.20 |
| 5 | 47.50 | 0.17 | 63.20 | 0.27 | 75.10 | 0.53 |
| mean | 58.25 | 0.20 | 50.14 | 0.26 | 57.07 | 0.39 |

ANOVA showed no significant differences among the three methods both in percentage of agreement ($F(2,14) = 0.355$, $\rho > 0.05$, $MSE = 3431.69$) and kappa ($F(2,14) = 1.891$, $\rho > 0.05$, $MSE = 0.421$).

## DISCUSSION

The purpose of this study is to analyze the validity and IRR of the three different work posture analyses RULA, REBA, and OWAS among new raters. Results showed high validity (observed by correlating the ratings of new raters with those of an ergonomics expert) of RULA, REBA, and OWAS. In addition, there were no significant differences in IRR among RULA, REBA, and OWAS. However, slight differences in overall IRR were found among the three methods, with IRR decreasing among RULA, OWAS, and REBA consecutively.

The RULA, REBA, and OWAS results showed high validity in terms of concurrent validity – testing how well a method corresponds with other more valid methods. In this study, the expert's rating was set as the gold standard. Although a common method of testing concurrent validity is comparing the results of observation with direct measurement as the gold standard, the direct method could not be conducted in this study due to restricted permissions from the industries. Another method of predictive validity, defined as the degree of risk estimation of the method related to the MSSs, was also impossible to conduct. This was due to the study being a cross-sectional study performed at one time and not a longitudinal study, where risk estimation is possible.

The fact that RULA is slightly better in terms of IRR than REBA and OWAS is not surprising. RULA has been widely used in different settings, such as in pediatric evaluation[30]. Plantard et al. reported on the validation process of RULA using objective measures (i.e., Kinect) and found valid RULA results compared with Kinect[31]. RULA is also used to observe MSSs when working with computers[32]. The vast usage of RULA in various settings reveals that RULA might be applied with

a similar level of understanding among new raters in the study.

The advantage of OWAS is that it is applied using a simple table. Interviews with the new raters in this study showed that OWAS was preferred over RULA and REBA due to its simplicity. In line with this result, Gomez-Galan et al. found that OWAS was preferable in many industries, ranging from manufacturing to healthcare and social assistance activities[28], for similar reasons.

As stated by Bao et al., consensus about which statistical methods are best for analyzing IRR is very limited[15]. For categorical posture data, however, percentage of agreement as well as kappa statistics are commonly used. In terms of the former, RULA is slightly better than REBA and OWAS. This result is in line with that of a previous study related to new raters, where Chen et al. found that RULA can be applied after a short training session, regardless of the assessor's experience in postural risk assessments[33]. It should be noted that raw percentage agreement is a simple and descriptive method, providing a straightforward interpretation of agreement between raters. The disadvantage is that it is an inadequate measure of agreement, because it does not account for chance agreement[25].

However, kappa is a measure of agreement that accounts for chance. Regarding the kappa value found, OWAS is slightly better than RULA and REBA, although all three methods revealed a kappa value of 0.21 – 0.40, which falls into the category of fair agreement.

This study has several limitations to be noted. First, only three work posture methods were analyzed in this study: RULA, REBA, and OWAS. These methods were chosen because they are the most commonly used and are found in many studies in developing countries, for example in Indonesia[7] and Korea[5]. Further research considering other work posture methods, such as the job strain index method and the OCRA method, are suggested to enrich the results[14]. Second, the postural analysis was only conducted in 5 different industries. Involving more industries will broaden the generalizability of the results. Third, only the observation method was used in this study. Comparing the results of the observation method with the other two methods – the self-reporting and objective methods – using specific tools, such as goniometers or Kinect, will give in-depth results[34,35]. Finally, this study applied percentage of agreement and kappa value to calculate IRR. The application of other methods of calculating IRR in future studies is suggested.

Despite its limitations, this study provides an important contribution to the body of research on work posture analysis. The results of this study show that among new raters there is high validity and no significant differences in IRR among RULA, REBA, and OWAS. However, the slight differences found in validity and IRR values of three analyzed postures reveal that RULA, OWAS, and REBA are consecutively ranked in terms of validity and IRR for new raters. For developing countries which struggle with the cost of hiring an ergonomics expert in postural analysis in order to observe and reduce work-related MSSs, the results of this study reveal that new raters can be as valid and reliable as an expert. The implications of this study when applied to an industry context are that industries can hire new raters to conduct postural analysis cost effectively and with valid and reliable results.

**ACKNOWLEDGEMENTS**
Author thanks K-02 of Ergonomics class, Industrial Engineering students 2017 for helping the data collection and analysis.

**COMPETING INTERESTS**
There is no conflict of interest.

**REFERENCES**

1. Punnett L, Wegman DH. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. J Electromyogr Kinesiol 2004; 14: 13-23.

2. Karwowski W, Marras WS. Occupational ergonomics: principles of work design. Boca Raton, FL, 2003.

3. Yelin EH. Cost of musculoskeletal diseases: impact of work disability and functional decline. J Rheumato 2003; 30: 8-11.

4. Buckle PW, Devereux JJ. The nature of work-related neck and upper limb musculoskeletal disorders. Appl Ergon 2003; 33: 207-217.

5. Kee D, Karwowski W. A comparison of three observational techniques for assessing postural loads in industry. Int J Occup Saf Ergo 2007; 13: 3-14.

6. Phedy P, Gatam L. Prevalence and Associated Factors of Musculoskeletal Disorders among Young Dentists in Indonesia. Malays Orthop J 2016; 10: 1-5.

7. Widyanti A. Ergonomic Checkpoint in Agriculture, Postural Analysis, and Prevalence of Work Musculoskeletal Symptoms among Indonesian Farmers: Road to Safety and Health in Agriculture. Jurnal Teknik Industri 2018; 20: 1- 10.

8. Muslim K, Nussbaum MA. Musculoskeletal symptoms associated with posterior load carriage: An assessment of manual material handling workers in Indonesia. Work 2015; 51: 205-213.

9. Widanarko B, Leg S, Devereux J, et al. Interaction between physical and psychosocial work risk factors for low back symptoms and its consequences amongst Indonesian coal mining workers. App Ergon 2016; 46: 158–167.

10. Juul-Kristensen B, Falletin N, Hansson GA, et al. Physical workload during manual and mechanical deboning of poultry. Int J Ind Ergon 2002; 29: 107–115.

11. Punnett L, Fine LJ, Keyserling WM, et al. Shoulder symptoms and postural stress in automobile assembly work. Scand J Work Environ Health 2002; 26: 283–291.

12. Silverstein BA, Viikari-Juntura E, Fan ZJ, et al. Natural course of nontraumatic rotator cuff tendinitis and shoulder symptoms in a working population. Scand J Work Environ Health 2006; 32: 99–108.

13. Cole DC, Wells RP, Frazer MB, et al. Methodological issues in evaluating workplace interventions to reduce work related musculoskeletal disorders through mechanical exposure reduction. Scand. J Work Environ Health 2003; 29: 396-405.

14. Chiasson M, Imbeau D, Aubry K, Delisle A. Comparing the results of eight methods used to evaluate risk factors associated with musculoskeletal disorders. Int J Ind Ergon 2012; 42: 478-488.

15. Bao S, Howard N, Spielholz P, et al. Interrater Reliability of Posture Observations. Hum Factors 2009; 51: 292 – 309.

16. Hansson GA, Balogh I, Bystrom JU, et al. Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands. Scand J Work Environ Health 2001; 27: 30–40.

17. Choppin S, Wheat J. The potential of the Microsoft Kinect in sports analysis and biomechanics. Sports Technology 2013; 6: 78-85.

18. Balogh I, Orbaek P, Winkel J, et al. Questionnaire-based mechanical exposure indices for large population studies: Reliability, internal consistency and predictive validity. Scand J Work Environ Health 2001; 27: 41–48.

19. Kourinka I, Jonsson B, Kilbom A, et al. Standardised Nordic Questionnaires for the Analysis of Musculoskeletal Symptoms. Appl Ergon 1987; 18: 233-237.

20. Hignett S, McAtamney L. Rapid Entire Body Assessment. Appl Ergon 2000; 31: 201-205.

21. Takala EP, Pehkonen I, Forsman M, et al. Systematic evaluation of observational methods assessing biomechanical exposures at work. Scand J Work Environ Health 2010; 36: 3-24.

22. David GC. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. Occup Med 2005; 55: 190-199.

23. Roman-Liu D. Comparison of concepts in easy-to-use methods for MSD risk assessment. Appl Ergon 2014; 45: 420-427.

24. Hallgree KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol 2012; 8: 23–34.

25. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. New York: Wiley 2003.

26. Karhu O, Kansi P, Kuorinka I. Correcting working postures in industry: a practical method for analysis. Appl Ergon 1997; 8: 199-201.

27. McAtamney L, Corlett EN. RULA: a survey method for the investigation of work-related upper limb symptoms. Appl Ergon 1993; 24: 91-99.

28. Gomez-Galan M, Perez-Alonso J, Callejon-Ferre A, Lopez-Martinez J. Musculoskeletal symptoms: OWAS review. Ind Health 2017; 55: 314-337.

29. Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin 1971; 378-382.

30. Dockrell S, O'Grady E, Bennett K, et al. An investigation of the reliability of Rapid Upper Limb Assessment (RULA) as a method of assessment of children's computing posture. Appl Ergon 2012; 43: 632-636.

31. Plantard P, Shum HPH, Le Pierres A, Multon F. Validation of an ergonomic assessment method using Kinect data in real workplace conditions. Appl Ergon 2017; 65: 562-569.

32. Levanon Y, Lerman Y, Gefen A, Ratzon NZ. Validity of the modified RULA for computer workers and reliability of one observation compared to six. Ergonomics 2014; 57: 1856-63.

33. Chen JD, Falkmer T, Parsons R, et al. Impact of experience when using the Rapid Upper Limb Assessment to assess postural risk in children using information and communication technologies. Appl Ergon 2014; 45: 398-405.

34. Widyanti A, Mahachandra M, Soetisna HR, Sutalaksana IZ. Anthropometry of Indonesian Sundanese children and the development of clothing size system for Indonesian Sundanese children aged 6-10 years. IJIE 2017; 61: 37-46.

35. Barriera-Viruet H, Sobeih TM, Daraiseh N, Salem S. Questionnaires vs observational and direct measurements: a systematic review. TIES 2006; 7: 261-284.