Reliability Analysis of Covid-19 Risk Management System in Malaysia Higher Learning Institution

Khairil Anuar Abdul Hamid
Senior Teaching Engineer, Centre for Design and Innovation of Technology (PRInT)
Industrial Technology Engineering Creativity Space
University Malaysia Pahang
Pekan, Pahang, Malaysia
khairil@ump.edu.my

Ezrin Hani Sukadarin and Mirta Widia
Occupational Safety and Health Program
Faculty of Industrial Sciences and Technology (FIST)
Universiti Malaysia Pahang
Kuantan, Pahang, Malaysia
ezrin@ump.edu.my, mirta@ump.edu.my

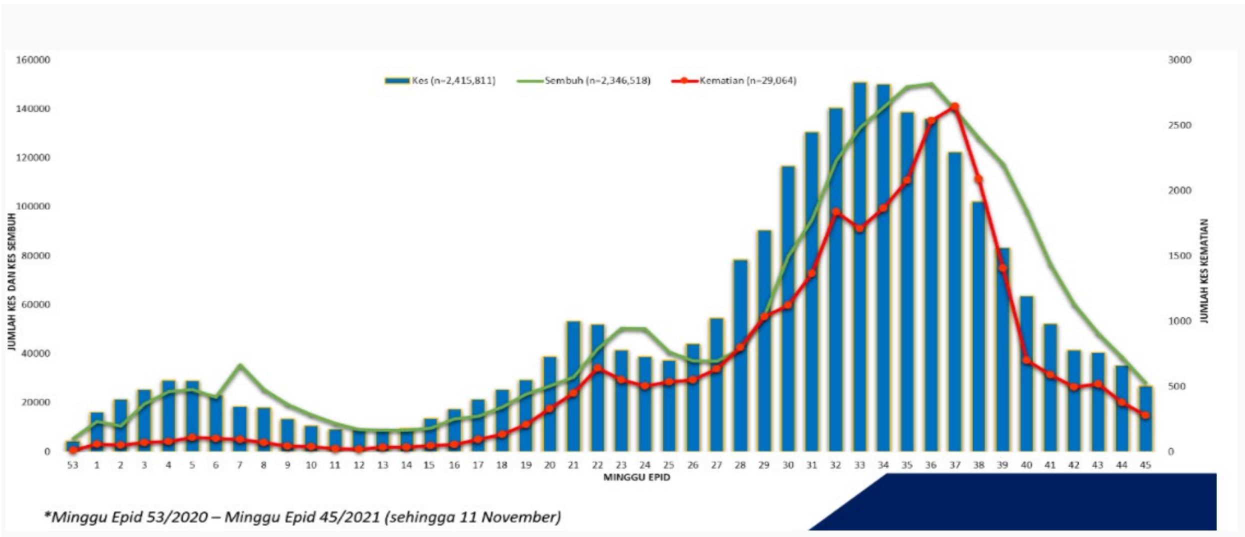
Abstract

Reliability concerns the extent to which a measurement of a phenomenon provides a stable and consistent result. Reliability is also concerned with repeatability. For example, a test will be reliable if measurement is repeated and conducted under constant conditions, hence will give the same result. This article attempts to present the result of a reliability study in the development of the Covid-19 risk management system in a Malaysia Institution of Higher Learning. The system is called C19-HIRAO. Two types of reliability study were explored for this purpose namely interrater reliability study and test-retest method. Interrater reliability study consists of Cohen's Kappa statistic and percent agreement (conventional method). Interrater Cohen's Kappa and Test-retest method were performed by the raters from Universiti Malaysia Pahang (UMP). The result proved that the developed C19-HIRAO has achieved an almost perfect agreement (0.92), moderate agreement (0.75) and very high relationship (1.00) for Interrater percentage agreement, Interrater Cohen's Kappa and test-retest method respectively. In conclusion, although this system is still newly developed, the risk rating value obtained is consistent among different raters. The value of this risk rating is also stable and reliable over the time when test-retest reliability is conducted within the stipulated time frame by the same rater.

Keywords
Covid-19 Risk Management System, Cohen's Kappa Statistic, Percentage agreement, Test-retest Method

1. Introduction
Universities should be a safe place from Covid-19 virus transmission due to the availability of research experts and its conducive environment, however, with the presence of a large number of staff and students, the result could turn the other way round. According to Nivette et al. (2021), Oosterhoff & Palmer (2020) and Tomczyk et al. (2020), Covid-19 SOP non-compliance practice mostly contributed from the student (young adult) which lead to the spreading of Covid-19 infection. In relation to this, hygiene (handwashing and coughing behavior) is one of the behavioral perspectives that has driven a higher percentage in non-compliance practice among the students across 10 countries in Europe (Wismans et al., 2020).

Working out the feasibility of the unparalleled worldwide measures to limit the coronavirus spread is now one of the scientists' most pressing questions. Among several techniques to avoid spreading Covid-19 discussed globally, there are four successful approaches. Firstly, the social, economic, government and educational institutions sector's restriction of movement (Han et al., 2020). Secondly, through systematic risk assessment which emphasizes all Covid-19 potential infections and the determination of appropriate risk control (Randolph & Barreiro, 2020). Thirdly an efficient vaccine that was estimated to be ready may take at least 12 – 18 months for clinical testing and approval (Pagliusi et al., 2020). The fourth one is through herd immunity (Randolph & Barreiro, 2020). Effective Immunization programs and herd immunity play a very important role in reducing the severity of cases due to Covid-19. The National Covid-19 Immunization Program (PICK) is on track due to factors such as the increase in the number of vaccinations and the effectiveness of the vaccination which has seen serious positive cases record a significant decline (Jayaraj et al., 2021). It is a good sign that is recognized not only among health experts or non -governmental organizations (NGOs) but also industry players and economists who also described the effectiveness of the immunization program as a hope for the economic sector to recover. Covid-19 death cases show a declining trend after week 36 epidemiology, in line with an increase in vaccination in Malaysia (See Figure 1).



Source: Ministry of Health (2021)

Figure 1: Cumulative Distribution of Covid-19 Cases in Malaysia

Countries such as Sweden and South Korea do not enforce movement restriction because they depend on the netizen's high degree of self-regulation and self-discipline. The people were committed and inspired to do their part in preventing the spread of the Coronavirus. Meanwhile, both movement restriction and self-regulation methods are implemented by Malaysia and Singapore. In Malaysia, the restriction of movement to contain the outbreak is imposed in the first few weeks. Once the infection curve of Covid-19 has been flattened, self-regulation approaches are taken to restart and restore the social and economic sector (Jamaludin et al., 2020).

One technique to avoid spreading Covid-19 is through systematic risk assessment which emphasizes all Covid-19 potential infections and the determination of appropriate risk control (Randolph & Barreiro, 2020). The OSH risk management system used in this study is known as Covid-19 Hazard Identification and Assessment of Risk and Opportunities (C19-HIRAO). The components in this C19-HIRAO were extracted from clauses 6.1.2.1, 6.1.2.2, 6.1.2.3, and 8.1.2 in International Standard of 45001: 2018. This system is able to identify hazards, recognize appropriate control measures as well as opportunities for improvement of control measures based on the value of risk rating (before control measure) and residue risk rating (after control measure). To date, no such study has been developed to address the transmission of covid-19 in the workplace, especially at the university. Before this system is available for use, it is important to ensure that the C19-HIRAO template used undergone an appropriate validity assessment. Two types of validation processes were used in this study and they are face validity and content validity. The result showed that face validity was found quite impressive with no fundamental remark while content validity indicated all risk management contents were retained due to their representative, relevance, and clarity.

5 main activities in UMP were assessed through developed C19-HIRAO. Each activity consists of several related sub-activities for example program facilitation activity represented by "lecture entering the classroom, a student entering the laboratory, chemical and apparatus handling and sitting for examination". The risk matrix adapted from DOSH Guideline Hazard Identification, Risk Assessment and Risk Control (HIRARC, 2008) comprises the severity and

likelihood column. The guidance from the Centre for Disease Control and Prevention (CDC) and National Security Council (MKN) were used to develop a severity impact column table while data from a recent study (personnel factor and self-movement restriction) was used to formulate a likelihood column table. The developed C19-HIRAO then was tested for its reliability.

1.1 Objective

To test in terms of reliability for the Covid-19 Risk Management System in a Malaysia Higher Learning Institution.

2. Literature Review

Reliability concerns the extent to which a measurement of a phenomenon provides a stable and consistent result (Carmines and Zeller, 1979). Reliability is also concerned with repeatability (Kennedy et al., 2019). For example, a test will be reliable if measurement is repeated and conducted under constant conditions, hence will give the same result (Moser and Kalton, 1989). Reliability is mainly divided into two part namely stability and internal consistency reliability (Mohajan & Mohajan, 2017). Two method commonly used to test the stability are test-retest reliability (test-retest method) and parallel-form reliability while internal consistency reliability is represented in two formats namely the inter-rater consistency and split-half reliability. Split-half reliability also known as internal consistency by some researcher. The description and how each type of reliability is measured described as per Table 1.

Table 1: Four types of reliability

Type of reliability	Description	How it is measured
Test-retest	Measure the stability of scores between 2 points of time within the same participant	The correlation between response of time 1 and time 2
Internal consistency	Measure the degree to which the items measure a single construct	Cronbach's alpha
Parallel forms	Measuring the degree to which changing the form of the questionnaire changes the response	Correlation of scores using form 1 and form 2
Inter-Rater	Measure the agreement between 2 raters or observers who make measurement	The percentage agreement between the 2 raters, or the correlation of their scores or can use Cohen's Kappa Statistic/Correlation

Source: (Mohajan & Mohajan, 2017)

The Cohen Kappa is used as a reliability statistic and it's a measure of rater agreement but the difference between this measure of rater agreement (using SPSS) and the conventional method (using excel) is that Cohen's Kappa corrects for a rater agreement due to chance (Gwet, 2008).Anything due to the chance is taken out from the analysis or out of the equation by using Cohen's Kappa analysis (McHugh, 2012). This analysis over the conventional agreement although the conventional rater agreement analysis can still be useful under certain circumstances. There will be a slight difference value between conventional percent agreement and Cohen's Kappa. The reason is that Cohen's Kappa has corrected for agreement due to chance. Cohen Kappa However provides the researcher with a p-value (approximate significance) which if the p-value is smaller than 0.05 there is a significant similarity between the two raters. Perhaps the best advice for researchers is to calculate both percent agreement and Kappa. If there is likely to be much guessing among the raters, it may make sense to use Kappa statistic, but if the raters are well trained and little guessing is likely to exist, the researcher may safely rely on percent agreement to determine interrater reliability (Schober & Schwarte, 2018). Meanwhile, the concept of interrater percentage agreement is fairly straightforward. Interrater reliability was measured as a percentage of agreement among data collectors for many years (Schober & Schwarte, 2018). Good test re-test reliability indicates a test's internal validity and ensures that measurements obtained in a single sitting are both representative and stable over time. Test re-test reliability analyses are frequently conducted over 2-time points (T1, T2) in a relatively short period to mitigate against conclusions being due to age-related changes in performance rather than poor test stability.

3. Methods

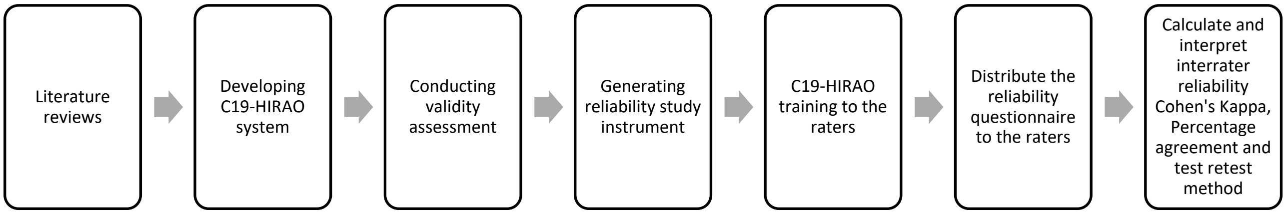


Figure 2: Methodology flow

C19-HIRAO development phase

The process of developing C19-HIRAO was involved in three phases (Pre-development, during development and post-development). The first phase (pre-development) is to attend training on HIRAO Training. The researcher was sent to the training on "understanding element in HIRAO". The purpose of this training is to emphasize the researcher with the element in HIRAO. The researcher also attended another training course regarding on risk of Covid-19 and its control measures. The second phase (during development) of C19-HIRAO consists of two-part. The first part is the development of the C19-HIRAO template and the second part consist of the activities that related to the establishment of the C19-HIRAO system. The C19-HIRAO was developed in Microsoft Excel. Besides free licenses, Microsoft excel provides the researcher with the ability to calculate, organize, and evaluate quantitative data, allowing the researcher to freely explore its functionality that best suited the type of semi-quantitative data in this study. The third phase (post-development) of the C19-HIRAO is conducted after the system is fully developed. The draft of the C19-HIRAO system has been validated through the face validity test and content validity test. After the validation process was completed, the reliability of the developed C19-HIRAO was tested subsequently. (Figure 2)

Interrater Cohen's Kappa formula

Calculation of Cohen's kappa was performed according to the following formula:

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr[e]}$$

Where $Pr(a)$ is the observed agreement among raters and Pr[e] is the hypothetical chance of the raters arriving the same answer. It can be calculated easily using SPSS.

Interrater percentage agreement formula

Percent agreement was obtained by calculating the total same risk rating for specific sub-activities in the university divided by numbers of raters (in percentage) using Microsoft excel. The number of raters who joined this study was 7. The risk rating varies from 1 to 25 which 1 to 4 justified as Low, 5 to 12 Medium and 15 to 25 High. The same highest amount in the risk rating is calculated as an agreement, for example, if 4 out of 7 raters indicate a risk rating score of 10 for the first sub-activity then the agreement percentage is 4/7 x 100% = 57%. Table 2 shows the score for Kappa statistic correlation and the interpretation. (Table 2)

Table 2: The Kappa Statistic level of agreement score

Value of Kappa	Level of Agreement	% of Data that are reliable	Interpretation
0.00 – 0.20	Less than chance agreement	0 - 4%	None
0.21 – 0.39	Slight agreement	4 - 15%	Minimal
0.40 – 0.59	Fair agreement	15 - 35%	Weak
0.60 – 0.79	Moderate agreement	35 – 63%	Moderate
0.80 – 0.90	Substantial agreement	64 – 81%	Strong
Above 0.90	Almost perfect agreement	82 – 100%	Almost perfect

Source: (McHugh, 2012)

Test and retest method formula

For a correlation between rater 1 and rater 2, the formula for calculating the sample Pearson's correlation coefficient (Pearson's r) is given by:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

Where X is test score, Y is retest score and N is the total number of pairs of test and retest scores. The calculation is perfomed in Microsoft excell.

Correlation coefficients are used to determine the strength and direction of the linear relationships between two variables (Schober & Schwarte, 2018). Reliability value is interpreted based on Guilford's reliability rule of thumb in Table 3.

Table 3: Strength of relationship score of test-retest method

Correlation coefficient (r)	Strength of relationship
< 0.2	Negligible relationship
0.2 – 0.4	Low relationship
0.4 – 0.7	Moderate relationship
0.7 – 0.9	High relationship
> 0.9	Very high relationship

Source: (Schober & Schwarte, 2018)

Conducting reliability test

Prior to conducting a reliability study, selected participants were invited by email. Eight staff from several departments in UMP who had a basic concept of workplace risk assessment were joined as raters. To ensure less guessing arise during the reliability test all participants have been trained the fundamental of risk management based on the latest version as stipulated in ISO 45001:2018.

The training has emphasized basic understanding and definition of the 3 pillars of risk management which typically involved hazard identification, risk assessment, and opportunities. The important part of this session is on how to calculate risk rating based on C19-HIRAO risk matrix. The training was considered to meet its objective when all participants were clearly understood how to do risk management on their own. From 8 participants who had joined the training 7 of them volunteered to proceed for the reliability test. The minimum number of the raters required for this purpose is two (Gwet, 2008). They were administered a set of interrater reliability test questionnaires which were conducted right after the training and 2 of them has joined the test-retest method 2 days after the training. The risk matrix is calculated by the given formula: Likelihood x severity.

The Questionnaire consisted of two parts. The first part is the respondent's information sheet which contains the objective of the reliability test, some instructions, the benefits of the study, and confidential agreement. The second part was a questionnaire form which has 4 sections namely section A, B, C, and D. Section A was a sociodemographic background and occupational information, section B was an interrater reliability test, section C was test-retest method study and section D is usability scale study (SUS). For newly developed system, 2 different raters required to establish test-retest reliability with minimum time gap 2 days (Gunda, 2009). Figure 3 was a screenshot of some examples of interrater reliability test questionnaires. As for sections B and C, the raters need to identify severity and likelihood rating based on the Covid-19 risk matrix as shown in Table 4,5 and 6. The system then will automatically calculate the risk rating based on severity and likelihood rating selection. The risk level will indicate whether LOW, MEDIUM, or HIGH based on a risk rating that varies from 1-25. There are 5 sub-activities from program facilitation activity and 7 sub-activities from office operation activity needed to be evaluated by the raters. The risk ratings that have the same score between raters will be considered to achieve the same agreement, otherwise, different values will lower the agreement percentage. The bigger the difference the lower the percentage agreement will be.

Hazard Identification And Assessment Of Risk & Opportunities (Interrater Reliability Test)

Hazards Identification (6.1.2.1)								Risks Assessment (6.1.2.2) & (8.1.2)												Opportunities
No.	Activity	Sub Activity	Condition of assessment	Context Of Organization (4.1 & 4.2)	Hazard	Hazard Category	OHS Risk	Severity	Likelihood	Risk Rating	Risk Level	Accident Record		Existing Control Measure		Hierarchy of control 1 Elimination (0.1) 2 Subtitution/ Isolation (0.3) 3 Engineering Control 4 Administrative (0.6 5 PPE (0.6)	ECM Factor	Residu Risk Rating	Risk Level	Consideration to implement include: 1. Historical occurences; 2. Identification of additional controls; 3. Effectiveness of control(s); 4. Severity;
4	**Office Operation (In General)**	Using meeting/ Discussion room	R	Internal Issue/Workers	Covid19 virus infection	Biological	can cause death and respiratory illnesses when transferred from other infected people through aerosol sneeze& coughing and hand shake					-	i	Arrangement of Seating chair with 1 metre distancing	4	Administrative	0.6	0		No further Opportunity Required
													ii	Sanitize the meeting space daily	4	Administrative				No further Opportunity Required
		Using Toilet	R	Internal Issue/Workers	Covid19 virus infection	Biological	can cause death and respiratory illnesses when transferred from other infected people through aerosol sneeze& coughing and hand shake					-	i	Only allowed 1 person to use toilet one person per time.	4	Administrative	0.6	0		No further Opportunity Required

Figure 3: Interrater reliability sample of the questionnaire.

Table 4: Covid-19 Risk Matrix

Likelihood / **Severity**	**1**	**2**	**3**	**4**	**5**
1	1 (Low)	2 (Low)	3 (Low)	4 (Low)	5 (Medium)
2	2 (Low)	4 (Low)	6 (Medium)	8 (Medium)	10 (Medium)
3	3 (Low)	6 (Medium)	9 (Medium)	12 (Medium)	15 (High)
4	4 (Low)	8 (Medium)	12 (Medium)	16 (High)	20 (High)
5	5 (Medium)	10 (Medium)	15 (High)	20 (High)	25 (High)

Table 5: Likelihood rating and its description

Likelihood	**Description**	**Rating**
Inconceivable	<u>Self-movement restriction</u> Only spend the whole time at home. Activities involve in very minimal contact with other people Source: (Murphy et al., 2020)	1
Remote	<u>Self-movement restriction</u> Enter the Campus for learning/research activities. Keep social distancing at all times. Return home after the job is done. (Murphy et al., 2020)	2

Conceivable	Self-movement restriction Travelling to red zone within the country (Murphy et al., 2020)	3
Possible	Self-movement restriction Travelling to other country and has higher chances to get infected throughout touching items. (Murphy et al., 2020)	4
Most likely	Self-movement restriction Travelling to other red zone country and has higher chances to get infected throughout meeting many peoples. (Murphy et al., 2020)	5

Table 6: Severity rating and its description

Severity	**Description**	**Rating**
Insignificant/good	Impact to people (Slight ill health) asymptomatic or no associated symptom but might pass on the virus to others. Source: (CDC, 2020) Impact to university reputation (Slight or no impact) Public awareness may not exist and no public concern. Green zone: 1 positive cases Source: (MKN, 2020)	1
Minor/acceptable	Impact to people (Minor ill health) mild symptoms e.g. sore throat, headache, muscle aches Source: (CDC,2020) Impact to university reputation (Minor impact) Public awareness may exist but there is no public concern. Yellow Zone: 2-10 cases cumulative positive cases Source: (MKN, 2020)	2
Moderate/anomaly	Impact to people (Moderate ill health) moderate symptom e.g. a fever, fatigue, a cough. Source: (CDC, 2020) Impact to university reputation (Considerable impact) Some public concern; Some local/state media and/or local attention. Orange Zone: 11-20 cumulative positive cases Source: (MKN, 2020)	3
Significant/serious	Impact to people (Serious ill health) high fever above 39.4C, difficulty breathing, blue lips or face Source: (CDC, 2020) Impact to university reputation (National impact) Regulatory improvement notice; National media coverage. Red zone: 21-40 cumulative positive cases Source: (MKN, 2020)	4
Major/Defective	Impact to people (Major ill health) Death Source: (CDC, 2020) Impact to university reputation (International impact) Regulatory prohibition notice; Extensive adverse attention (National) Black zone: 41 and above cumulative positive cases Source: (MKN, 2020)	5

4. Results and Discussion

4.1 Percentage agreement reliability (Conventional method)

Before proceeding with the data, it is important to note that the risk rating score is 10 for each respective sub-activity determined by the researcher. Other than this value was considered as a difference value and such discussion should be made for further research recommendation. Table 7 shows the percentage agreement reliability result which demonstrates that all sub-activities have the same agreement (Risk level=10) except for sub-activity "student entering lab/classroom", "Apparatus/chemical Handling during Lab activities by undergraduate students (Final year students)", and "Using Pantry" which score 0.71, 0.71, and 0.86 respectively. The average percentage agreement score is 0.92 which can be interpreted as an "almost perfect" agreement based on Table 2. The higher percentage agreement between rater indicated that the developed C19-HIRAO system is practically well-functioning.

Table 7: Percent agreement (conventional method)

Main Activity	No. of sub-activity	Risk management sub-activity	Risk Level (Likelihood x Severity)								
			Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7	differences	% agreement
Program Facilitation	1	Lecturer / Technical Staff Entering Lab / Classroom	10	10	10	10	10	10	10	0	1.00
	2	Student Entering Lab / Classroom	10	10	10	10	8	8	10	2	0.71
	3	Lecturing session / Lab demonstration by Lecturer & Technical Staff	10	10	10	10	10	10	10	0	1.00
	4	Apparatus / chemical Handling during Lab activities by undergraduate students (Final year students)	10	10	10	10	8	8	10	2	0.71
	5	Mid-term examination / Test	10	10	10	10	10	10	10	0	1.00
Office Operation (In General)	1	Using meeting/ Discussion room	10	10	10	10	10	10	10	0	1.00
	2	Using Toilet	10	10	10	10	10	10	10	0	1.00
	3	Using Pantry	10	10	10	10	10	8	10	1	0.86
	4	Using Prayer room	10	10	10	10	10	10	10	0	1.00
	5	Staff Work station	10	10	10	10	10	10	10	0	1.00
	6	Open and close any door in office (meeting room, pantry, main office)	10	10	10	10	10	10	10	0	1.00
	7	Handling sick student (illness e.g fever, internal body paint)	10	10	10	10	10	10	10	0	1.00
All raters Average score											0.92
Raters 5 and raters 6 average score											0.95

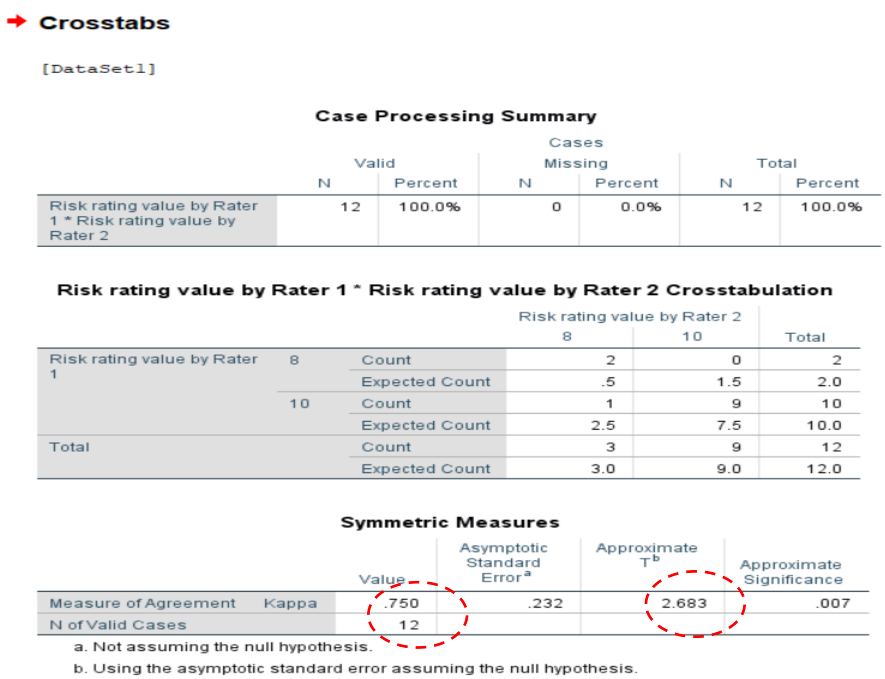
4.2 Cohen's Kappa reliability

In this part, the two worst data were selected for Cohen's Kappa analysis (rater 5 and rater 6). In a simple word, out of 5 raters who gave a score of 10 (risk rating) for each sub-activity were excluded in Cohen's Kappa assessment. This aims to obtain the lowest data to strengthen the % agreement by assuming that this system is still new and needs continuous training if implemented in the workplace. To understand this result, it is suggested to read the line in conjunction with the understanding of the percentage agreement result that was elaborated in sub-chapter 4.1. Out of 12 sub-activities evaluated by the two raters 9 of them has scored the same risk rating (10) except for sub-activity 2 and 4 from program facilitation activity and sub-activity 3 from office operation activity. Cohen's Kappa does not represent the result based on evaluated activity but merely on the risk rating score. For example, the risk rating that has score 8 contributed by 2 sub-activities evaluated by rater 5 and 3 sub-activities evaluated by rater 6 (refer to Table

8), while the risk rating that has score 10 contributed by 10 sub-activities evaluated by rater 5 and 9 sub-activity by rater 6. In short, the result has demonstrated a moderate level of agreement when the Cohen's Kappa score is 0.75 (refer to Figure 4). Level agreement for this value can be considered as a moderate agreement between the raters (Carmines and Zeller, 1979). This value is lower when compared to the percent agreement value of 0.95. The p-value (approximate significance) is 0.007 for these two raters (refer to figure 4). According to McHugh, if the p-value is smaller than 0.05 there is a significant similarity between the two raters. Assuming that there is a similarity of understanding between these two raters on the severity rating of 4 (Covid-19 stage 4 symptoms), thus it can be considered that a value of 0.75 is good because Cohen's Kappa has corrected for agreement due to chance. In conclusion, Cohen's Kappa analysis takes into account the factors that raters only guess in choosing the severity and likelihood rating score when this study was conducted.

Table 8: Cohen Kappa statistic

		Risk Rating value by Rater-6					
		Risk Level 8		Risk Level 10		Total	
		N	%	N	%	N	%
Risk Rating value by Rater-5	Risk Level 8	2	66.7	0	0.0	2	16.7
	Risk Level 10	1	33.3	9	100	10	83.3
Total		3	100	9	100	12	100



Crosstabs

[DataSet1]

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Risk rating value by Rater 1 * Risk rating value by Rater 2	12	100.0%	0	0.0%	12	100.0%

Risk rating value by Rater 1 * Risk rating value by Rater 2 Crosstabulation

			Risk rating value by Rater 2		
			8	10	Total
Risk rating value by Rater 1	8	Count	2	0	2
		Expected Count	.5	1.5	2.0
	10	Count	1	9	10
		Expected Count	2.5	7.5	10.0
Total		Count	3	9	12
		Expected Count	3.0	9.0	12.0

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.750	.232	2.683	.007
N of Valid Cases		12			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 4: Cohen's Kappa calculation using SPSS

Test-retest method result

Data from interrater reliability for rater 5 was recalled as test 1 (conducted on 15.11.2021) in this study while test 2 was conducted 2 days after that. The rater judgments remain consistent over this period with a correlation coefficient value of 1.00. Based on the Guildford rule of thumb for the test-retest method the value above 0.9 can be interpreted as a very high relationship. The result indicated an impressive correlation between test 1 and test 2 due to the two-gap days implied in this assessment (Refer to Table 9). However, it is important to note that on some occasions, the agreement and disagreement may change over time. A common practical issue faced by the application of test-retest reliability is the influence of both the dynamic nature of the construct being measure over time and the duration of the time interval (Haynes et al., 2019). Most of the other factors that drive the stability of this result are due to psychological phenomena such as mood, poor memory, and the failure of the observer to understand the concepts to be conveyed in the relevant exercises (Geisinger, 2013). The test-retest reliability result may as well be contributed from both memory effect and practice effect. One example of memory effect is an individual may simply recall their responses from the first test occasion when attempting the re-test occasion while practice effect is defined as the familiarization of similar test which merely on a cognitive-type test. Thus, it is important to ensure that training can be delivered in such an interesting way to ensure the participant can fully understand the gist of the training concept. Another factor to be considered is the observer selection criteria such as their knowledge, expertise, and their time commitment. On top of that, the time gap between test 1 and test 2 should be ensured not taken too long and match the topics presented in the exercises conducted.

Table 9: Test-retest method (rater 5)

Main Activity	No. of sub-activity	Risk management sub-activity	Risk Level		Correlation coefficient
			Date		
			15.11.2021 Test 1	17.11.2021 Test 2	
Program Facilitation	1	Lecturer / Technical Staff Entering Lab / Classroom	10	10	1.00
	2	Student Entering Lab / Classroom	8	8	
	3	Lecturing session / Lab demonstration by Lecturer & Technical Staff	10	10	
	4	Apparatus / chemical Handling during Lab activities by undergraduate students (Final year students)	8	8	
	5	Mid-term examination / Test	10	10	
Office Operation (In General)	1	Using meeting/ Discussion room	10	10	
	2	Using Toilet	10	10	
	3	Using Pantry	10	8	
	4	Using Prayer room	10	10	
	5	Staff Work station	10	10	
	6	Open and close any door in office (meeting room, pantry, main office)	10	10	
	7	Handling sick student (illness e.g., fever, internal body paint)	10	10	

5. Conclusion

In short, the developed C19-HIRAO meets all respective risk assessment criteria that emphasized hazard and risk recognition, a full assessment of risk, and dynamic opportunities plan to mitigate the risk to the acceptance level which may less harm to the staff and student in the university. The objective of this paper which is to present the result of a reliability analysis in the development of the C19-HIRAO in Malaysia Higher Learning Institution meets its purpose throughout the result discussed earlier on. A better result can be achieved by ensuring less guessing among the raters when the reliability study is conducted, it is important to ensure that each participant is given comprehensive training through a clear conceptual understanding of the system developed and their understanding of the training attended should be well assessed. Finally, it can be concluded that both inter-rater consistency (Cohen's Kappa analysis and percentage agreement) and test-retest reliability capable to measure the agreement between rater either the stability of scores between 2 points of time within the same raters. In addition to this conclusion, another method that can be suggested to measure reliability is known as Fleiss Kappa Analysis (FPA). FPA can be applied to determine the level of agreement between two or more raters as long as it meets a few basic requirements and assumptions of Fleiss' kappa. Cohen's kappa only limited the researcher with the measurement between two raters which can be replaced by FPA for future research development in the same context of this study. This paper is very useful as previous studies related to the development of a risk management system at the workplace had overlook the reliability assessment. This is often happening due to less exposure among the researcher to the reliability assessment concept emphasized on semi-quantitative analysis. Therefore, this paper can be used as a reference to explore the reliability analysis concept that can be applied in the related-risk management system at the workplace.

References

Carmines, E. G., & Zeller, R. A. Reliability and validity assessment. Sage publications, 1979.

Chatterjee, R., Bajwa, S., Dwivedi, D., Kanji, R., Ahammed, M., & Shaw, R. COVID-19 Risk Assessment Tool: Dual application of risk communication and risk governance. Progress in Disaster Science, 7, 100109, 2020.

Geisinger, K. F. Reliability. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 21–42). American Psychological Association, 2013.

Gunda AK. The inter-rater reliability of lifting and carrying of OccuPro's functional capacity evaluation [unpublished powerpoint]. Concordia University Wisconsin; 2009.

Gwet, K. L. Intrarater Reliability. In Wiley Encyclopedia of Clinical Trials. John Wiley & Sons, Inc., 2008.

Han, E., Tan, M. M. J., Turk, E., Sridhar, D., Leung, G. M., Shibuya, K., Asgari, N., Oh, J., García-Basteiro, A. L., Hanefeld, J., Cook, A. R., Hsu, L. Y., Teo, Y. Y., Heymann, D., Clark, H., McKee, M., & Legido-Quigley, H. Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. The Lancet, pp. 1525–1534, 2020.

Haynes, T., Bishop, C., Antrobus, M., & Brazier, J. The validity and reliability of the My Jump 2 app for measuring the reactive strength index and drop jump performance. Journal of Sports Medicine and Physical Fitness, pp. 253–258, 2019.

Jamaludin, S., Azmir, N. A., Mohamad Ayob, A. F., & Zainal, N. COVID-19 exit strategy: Transitioning towards a new normal. Annals of Medicine and Surgery, 59(October), pp. 165–170, 2020.

Jayaraj, V. J., Rampal, S., Ng, C. W., & Chong, D. W. Q. The Epidemiology of COVID-19 in Malaysia. The Lancet Regional Health - Western Pacific, 17, 2021.

Kennedy, L. G., Kichler, E. J., Seabrook, J. A., Matthews, J. I., & Dworatzek, P. D. N. Validity and Reliability of a Food Skills Questionnaire. Journal of Nutrition Education and Behavior, 51(7), pp. 857–864, 2019.

Mohajan, H., & Mohajan, H. K. Munich Personal RePEc Archive Two Criteria for Good Measurements in Research: Validity and Reliability Two Criteria for Good Measurements in Research: Validity and Reliability, 2017.

MOSER, C. A. & KALTON, G. Survey methods in social investigation, Aldershot, Gower, 1989.

McHugh M. L. Interrater reliability: the kappa statistic. Biochemia medica, 22(3), pp. 276–282, 2012.

Nivette, A., Ribeaud, D., Murray, A., Steinhoff, A., Bechtiger, L., Hepp, U., Shanahan, L., & Eisner, M. (2021). Non-compliance with COVID-19-related public health measures among young adults in Switzerland: Insights from a longitudinal cohort study. Social Science and Medicine, 268(September 2020), 2021.

Oosterhoff, B., & Palmer, C. A. Attitudes and Psychological Factors Associated with News Monitoring, Social Distancing, Disinfecting, and Hoarding Behaviors among US Adolescents during the Coronavirus Disease 2019 Pandemic. JAMA Pediatrics, 174(12), pp. 1184–1190, 2020.

Pagliusi, S., Jarrett, S., Hayman, B., Kreysa, U., Prasad, S. D., Reers, M., Hong Thai, P., Wu, K., Zhang, Y. T., Baek, Y. O., Kumar, A., Evtushenko, A., Jadhav, S., Meng, W., Dat, D. T., Huang, W., & Desai, S. Emerging manufacturers engagements in the COVID −19 vaccine research, development and supply. Vaccine, 38(34), pp. 5418–5423, 2020.

Randolph, H. E., & Barreiro, L. B. Primer Herd Immunity: Understanding COVID-19. Pp. 737–741, 2020.

Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. Anesthesia and Analgesia, 126(5), pp. 1763–1768, 2018.

Wismans, A., Letina, S., Thurik, R., Wennberg, K., Franken, I., Baptista, R., Barrientos Marín, J., Block, J., Burke, A., Dejardin, M., Janssen, F., Mukerjee, J., Santarelli, E., Millán, J. M., & Torrès, O. Hygiene and social distancing as distinct public health related behaviours among university students during the COVID-19 pandemic. Social Psychological Bulletin, 15(4), 2020.

Zitek, E. M., & Schlund, R. J. Psychological entitlement predicts noncompliance with the health guidelines of the COVID-19 pandemic. Personality and Individual Differences, 171, 2021.

Biography

Khairil Anuar Abdul Hamid has completed his Degree in Chemical and Process Engineering from University Kebangsaan Malaysia (UKM) and master's in science (Occupational Safety & Health) from University Malaysia Pahang (UMP) respectively. Currently he is working as a Senior Vocational Training Officer at the Centre for Design and Innovation Technology (PRInT) in University Malaysia Pahang (UMP). Have been served for UMP since 2006 and 2 years in Fertilizer industry. He was instructor for several Chemical Engineering Technology Laboratory Subjects including Process Control Instrumentation, Chemical Reaction, Fluid Technology, Gas Turbine Operation and Maintenance Supervision, Basic Lifting and Rigging, HSE in Oil and Gas, Ethic in Profession and many more.

Ezrin Hani Sukadarin is an Associate Professor at the Occupational Safety and Health Program, Faculty of Industrial Sciences and Technology, University Malaysia Pahang (UMP). She has a bachelor's in computer science (majoring in Software Engineering) University Teknologi Malaysia (UTM), master's in occupational safety and Health Management, University Utara Malaysia (UUM) and PhD in Mechanical and Material Engineering (Specialized in Occupational Ergonomics), University Kebangsaan Malaysia (UKM). She has been recognized as an Ergonomic Trained Person (ETP) for Initial and Advanced ERA by DOSH (Malaysia). She has published research articles in high-impact (ISI/SCOPUS) journals, chapters of books, and proceedings of the national and international conference in the discipline of Human Factors and Ergonomics. She is actively involved as a principal investigator and a team member of national and international research projects. Her main area of research interest is Human Factors and Ergonomics. She has provided consultations to Lembaga Perindusrian Nanas Malaysia (MPIB), Lembaga Kemajuan Perusahaan Pertanian Negeri Pahang (LKPP Negeri Pahang). As an active member of Human Factors and Ergonomics Malaysia (HFEM), she manages activities related to the academic conference and national student chapters under the prestige association.

Mirta Widia is currently a Senior Lecturer in the Occupational Safety and Health Program, Faculty of Industrial Sciences & Technology, University Malaysia Pahang (UMP). She has been recognized as one of the Ergonomic Trained Person (ETP) for Initial ERA by DOSH (Malaysia). She has a Degree in Industrial Engineering from Universitas Andalas (UNAND), an M.Eng.Sc, and a Ph.D. in Ergonomics from the University of Malaya (UM). She has published research articles in high impact (ISI/SCOPUS) journals, a chapter of book and the proceeding of an international conference in the discipline of Ergonomics and Manufacturing System. She is actively involved as a principal investigator, as well as a team member of local and international research projects. Recently, she is the principal investigator of the "Rear Impact Crash Characteristics in ASEAN countries" international grant awarded by ASEAN NCAP. She has cooperated in a variety of ergonomic projects for 13 years as a researcher and consultant member with Aerospace Malaysia Innovation Centre (AMIC), Proton Berhad, DRB HICOM, MIROS, and Malaysian Palm Oil Berhad (MPOB). She participated in several program committees of national, international conferences and workshops. Her research activities primarily focus on issues of design and workplace ergonomics as these affect the health especially musculoskeletal disorder, comfort and productivity of workers. Her research interest includes ergonomics (human factor), occupational safety, health and industrial engineering.