International Review of Management and Marketing

ISSN: 2146-4405

available at http: www.econjournals.com

International Review of Management and Marketing, 2016, 6(4), 843-850.

Using Measures of Agreement to Evaluate a Chronic Problem of Congestion in the Empty Container Depots

Rosmaizura Mohd Zain[1]*, Mohd Nizam Ab. Rahman[2], Ainon Ramli[3], Zulkifli Mohd Nopiah[4], Nizaroyani Saibani[5]

[1]Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Pengkalan Chepa, Malaysia, [2]Department of Mechanical and Materials Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia, [3]Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Pengkalan Chepa, Malaysia, [4]Unit of Fundamental Engineering Studies, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia, [5]Department of Mechanical and Materials Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia. *Email: rosmaizura.mz@umk.edu.my

ABSTRACT

Container-based transportation through a combination of truck, rail and ocean shipping services has become an important part of supporting local and international cargo flows. However, congestion or bottlenecks of empty container-trucks may hinder the supply chain and logistics performance of the country in hinterland transport. The factors causing the bottlenecks of empty container-trucks at the depots around Port Klang, Malaysia, were first derived from interviews conducted with 30 participants working in different logistics companies. This paper intends to describe the application of Fleiss kappa to evaluate or validate the congestion factors and involves six experienced evaluators from the field of empty container logistics. Therefore, key steps and procedures are presented by extracting the qualitative data from a previous study that addressed this phenomenon. A resulting Fleiss generalised kappa value of 0.8203, which can be assumed as satisfactory and reflecting the strength of the congestion factors, as well as it would convince the real situation that occurs in bottlenecks at the empty depot. Truly, Fleiss kappa has proved to be an efficient and useful approach to measure the agreements for many purposes, including validating the chronic congestion factors.

Keywords: Interrater Agreement, Congestion, Empty Container-truck
JEL Classifications: L90, L91, L92

1. INTRODUCTION

In this world of global competition, logistics has become an important area to support the incoming and outgoing materials along the supply chain network. Logistics consists of the distribution of materials or goods from upstream to downstream channels; it begins with the suppliers, processers, intermediaries or wholesalers and finally ends with the consumer (Gunasekaran and Ngai, 2003). The accomplishment of local or international trade highly depends on logistic activities. Therefore, it is crucial to ensure the smooth flow of activities at a container terminal. At this point, this situation may reduce some interruptions when delivering goods through trucks, trains or ships. Finally, the effect of interruptions would reduce the vehicle turnaround time. Moreover, the transportation sector plays a crucial role in accomplishing the shipment of cargo to the end customer. For this reason, reliable transportation is required to meet the customer's demand. Moreover, container terminals become a medium to transport cargo by offering latest material handling, information technology and other facilities. According to Gudelj et al. (2010), the efficiency of transportation by container terminals depends both on supply planning and control strategies.

As per UNCTAD (2014), sea transport carried around 80% of cargo around the world. It can be observed that the container handling for "developing companies" in 2014 increased to 491.2 million TEUs (5.34%) when compared to the previous year, which accounted for 466.3 million TEUs (5.1%). Moreover, container handling for the "whole countries" with a total of 684.4 million TEUs in 2014 indicates that the growth is more than that of the previous year (651.2 million TEUs). Consequently, the enormous increase in the growth of containers would make sea transport the pillar of international and global trade. Container throughput in Malaysia handled 22.7 million TEUs when compared to Hong Kong (China) handling 22.3 million TEUs and Indonesia handling 11.9 million TEUs in 2014 (UNCTAD, 2014). In line with the economic growth, this situation will enormously support the demand of empty containers. However, the container growth would result in the expansion of the transportation industry that may cause several uncertainties in container chains, such as bottlenecks and late delivery thereby increasing the cost of operations. For example, in the United States, significant growth of container volume has slowed down the efficiency of container handling because of encountering more difficulties in truck access in or out of the terminal thereby resulting in congestion and induced delays. It can be observed that truck congestion is one of the hindrances that lead to inefficiency in operations at the container terminal (Namboothiri and Erera, 2004).

A bottleneck can be defined as a subset of congestion in a system that causes the entire process in each stage to slow down (Leporis and Kralova, 2010). A bottleneck in the transportation system is defined as a shortage of infrastructural resources of the intermodal and multimodal transportation chain (Möller, 2014). There are several reasons why the transportation system is experiencing a bottleneck. For instance, problems arise when a truck is stuck in traffic outside the terminal gates; handling equipment is unreliable or when any other inappropriate activities or practices occur (Rankine, 2003).

Therefore, the inefficiencies caused by the bottleneck lead to a long queue in trucks and a long overall cycle time is the consequence. According to Haralambides and Gujar (2011), congestion is a long-lasting problem in many ports as it increased the cost of shipment and depleted the development of trade thereby leading manufacturers, wholesalers and retailers to store huge stocks. As a result, this will increase the cost of transport and production; so, the prices of final goods shoot up; consequently, the development in international trade decreases.

The risk of terminal congestion may occur as a result of any hindrance at any given country; for example, a case in Malaysia wherein the truck drivers criticised the authorities over the charges imposed by the container depot firm while retaining the poorly managed services as is. Consequently, the drivers had to wait for 3-4 h to pick up and move in the containers; finally, this created longer trip and consumption times to deliver the containers to customers' premises. In addition, such delays cause congestion and increase the cost in the container chain (including repositioning empty containers, fuel consumption and maintenance) (Thestar, 2012). Moreover, Adam (2009) mentions that this congestion factor is caused as the land area allocated for the port is limited; so, this encourages frequent bottlenecks and congestion in the port area. His study attempts to examine the main constraints of bottlenecks that exist at Male' Commercial Harbour that affects the logistics performance.

The development of international trade and the increasing demand of container transportation can cause several hindrances, such as traffic congestion, to emerge. This is due to the container terminal being located at the port or hinterland thereby leading to a limitation in resources (e.g. space, machine, manpower) (Vacca et al. 2007). It is possible that the uncertainties related to congestion may impact several players or stakeholders in a container chain, such as shipping agents (problems in vessel delays, extra costs and missed feeders); port terminals (extra manpower, yard congestion and re-handling); truck companies (waiting time and loss of business) and importers (longer lead times) and container depot (delay or waiting time, cost of maintenance).

Previous researchers ignored to study the bottleneck factors or causes at empty container depot facilities. Only a few studies concentrate on identifying the factors of bottlenecks; most studies concentrate on the operational decision problems, such as berth allocation, stowage planning and port disruptions (Vacca et al. 2007; Ji and Zhou, 2010; Jones et al. 2011). Our previous studies (see Zain et al., 2014a; Zain et al., 2014b; Zain et al., 2015) have identified related factors or categories including the areas of bottlenecks at off-dock storage of empty container among the logistics providers or third party logistics, manufacturers and other container logistics chains that tend to interrupt or delay the delivery and return of the empty container on truck. At present, a high operational level efficiency seems to be a very crucial factor to provide a high service of quality to fulfil customer's specification and satisfaction; this view is also supported by Kaleappan (2006).

This paper focuses on measuring the rate agreement for nominal data by the proficient panel with skill in container logistics area through the application of Fleiss kappa. Particularly, it aims to evaluate and classify sub-categories into main categories (factors). Therefore, by measuring the rate of agreement through Fleiss kappa approach, it is hoped that this study, though in a small way, would help evaluators to strike a stronger agreement and further establish validity and consistency through the triangulation method.

2. LITERATURE REVIEW

In statistics, the terms of inter-rater agreement (IRA) and inter-rater reliability (IRR) are referred to the level of agreement among evaluators or observers/raters/judges. Landis and Kosh (1977) suggested that the level of agreement can be measured by Cohen kappa scale. However, if three or more evaluators evaluate the IRA, Fleiss (1971) proposes the extension of kappa. According to Shweta et al. (2015), Graham et al. (2012) and Hruschka et al. (2004), evaluation of IRA or IRR, either as a primary or secondary method of study, is widely used in various disciplines (such as in social research, education, medicine, psychology and marketing).

The two terms IRA and IRR are often used interchangeably by researchers or practitioners. Hallgren (2012) proposed that, IRR, which is also known as IRA, is the data collected through the ratings provided by evaluators. Thus, different evaluators assign the same score for each item that is being observed. According to Shweta et al. (2015) and Graham et al. (2012), there are dissimilarities in the terms of agreement and reliability; while IRA measures the frequency of two or more evaluators that allocate the same score, IRR quantifies the relative similarity between two or more sets of ratings. In this study, IRA has received more attention because the approach becomes an indicator to render some clarity in terms of the classification of sub-categories and categories of bottlenecks.

Qualitative analysis is another research field wherein IRA and IRR can be used in numerous applications. For example, in Mahamod and Ishak (2003), Jasmi (2012), Jasmi and Tamuri (2011); Yusri et al. (2012); Yusri et al. (2011); measuring the agreement of theme is done through Cohen kappa's index analysis. Recent studies conducted by Tamam et al. (2010); Timbang et al. (2011); Mahamod and Embi (2005) have also been published in this field. Validity in terms of expertise is very important so as to ensure that protocol interview can be trusted (Yin, 1994). Therefore, studies conducted by some local researchers have revealed their concerns relating to the evaluation of the reliability of qualitative data so as to increase the quality of research.

In addition, Schaer and Mayr (2010); Burla et al. (2008); Grant (2006); Zenk et al. (2007) consider to adopt kappa statistics to quantify the agreement between two or multiple evaluators in their studies. According to Timbang et al. (2011), reliability of interview data can be proved strongly through the evaluation of reliability by Cohen kappa indexes. However, the literature contains relatively little guidance on how to evaluate IRA by Fleiss kappa for multiple evaluators when compared to Cohen's kappa (related to the two evaluators in certain events). There are several studies focusing on Fleiss kappa or denoted as generalised kappa statistic (see Antonakos and Colling, 2001; Schaer and Mayr 2010; Gwet, 2012). Basically, the agreement for two evaluators can be measured by either two levels of performance (e.g. agree or disagree) or three and more standard levels of rating scales can also be measured by kappa (Graham et al. 2012).

Therefore, the development of measures with adequate IRA is essential for this study to measure similarity or agreement. This study aims to evaluate or measure the level of agreement for every item into the category. Particularly, the use of IRA will be justified in this study by placing each item (sub-category) into a specific category, and the agreement of each item will be identified (see Mahamod and Embi, 2005; Antonakos and Colling, 2001). Thus, the emphasis will be more on the application of the formula aspects rather than on the mathematical derivation of the indices.

3. METHODOLOGY

The special focus on this study is related to nominal variables and kappa statistics worked by programme designed in Microsoft excel to calculate rater agreement as denoted by Gwet (2012) and Fleiss (1971) guidelines. According to Antonakos and Colling (2001); Graham et al. (2012), there are many types of statistical software that can be used to calculate kappa, such as SPPSX, SAS, Stata, SPSS; or programs designed, such as AgreeStat, AGREE and ReCal.

The qualitative data collected based on in-depth interviews have been addressed in a previous study used to estimate a situation in which the selected evaluators independently classified the sub-category from interview approach into appropriate category. After deriving the category and sub-category (item), the researcher continues to review the item through Indexes Fleiss kappa coefficient. The values of proportion of agreement for each item were calculated based on the formula highlighted by Fleiss (1971).

Six panels with container logistics practices and academic experience in logistics independently rated each item in the category. Most panel members resided in Klang Valley, Selangor, Malaysia. Nominal scaled or categorical code data were tabulated into an appropriate format; thus, each evaluator needs to assign each unit as (0=disagree, 1=agree) into different categories, (j1=attitude, j2=information, j3=operations handling, j4= monitoring and j5= facilities/others) (Table 1). In this paper, "unit" (sub-category) would be used as a generic term for subject that is rated in this study. Similarly, Gwet (2012) also applied this term for subject or other researchers used people, things or events. Panel members required to finish with their expert judgment within 2 weeks. The following section reviews Fleiss' equations and the procedure that shows the computation of the agreement for multiple evaluators.

Table 1: The description categories of bottlenecks factors in empty container depot

Category	Description
Attitudes	Reflections on employees' competency levels (e.g., knowledge, characteristics or traits relevant to high-performances and skills in workplace). Attitudes hugely impact operational performance; for example, the staff in the depot, or other players in empty container logistics who always respond quickly to any enquiries or problems demonstrate professional attitude and ability to show good knowledge in a given job or situation (Vathanophas and Thai-ngam, 2007; Yeo et al. 2015)
Information	Concerned with the consistency and sufficient information flows among the depot players to facilitate container management or the level of technology applications in order to communicate the distribution of information, organise inventories or booking among firms (Yeo et al. 2015; Kia et al. 2000; Soares et al. 2013)
Operations handling	Activities or processes related to container handling works and other related activities (pick up, drop off, repair or wash and grading of container) and operational constraints (truck waiting time, scarcity of resources) (Kim and Kim, 1999; Lai et al. 1995)
Monitoring	Related to monitor the activities of empty container management by paying more attention to container operations, maintenance, claims response, monitor for inventory status and storage of container (Salekan, 1997; Umar, 2004)
Facilities/others	Condition of depot based on the infrastructure, and facilities among parties (e.g. road, gate access), or "others" may refer to relationship among players or how they efficiently link among the key players so as to solve a problem (Nasir, 2014)

Shweta et al. (2015) believed that, there are no specific rules relating to the level of agreement to ensure that the assessment process is reliable. Kappa scores can range from 0 (indicating no agreement) to 1.0 (indicating full agreement). However, there is less consensus on the sufficient multiple evaluators score among researchers. According to Fleiss (1981), the benchmark scale for the kappa values as Poor (<0.40), intermediate to good (0.40-0.75) and excellent (>0.75).

McCray (2013), highlighted that Landis and Koch (1977) benchmark kappa values' are as follows: Moderate (0.41 and 0.60); Substantial agreement (>0.60) and Perfect agreement (>0.80). Meanwhile, Altman's (1991) benchmark scale for the kappa values are as follows: Poor agreement (<0.2); fair agreement (0.21-0.40) and very good (0.81-1.00). It can be concluded that, the value of 0.80 is considered as the popular benchmark and this indicates a high level of agreement (Altman, 1991; Landis and Koch, 1977).

4. RESULTS AND DISCUSSION OF INTERRATER AGREEMENT MEASURES

According to Abramović et al. (2012), the container terminals are often congested due to factors such as, delays, long service time, poor infrastructure or any other unexpected event; thus, all these will reduce the quality of services thereby reducing the performance level of the logistics chain. Therefore, the categories of bottleneck that have been found in this study are also associated with quality. Miremadi et al. (2011), another study dealing with quality, identified five common dimensions, relevant to port industry such as reliability, empathy, assurance, responsiveness and tangibility, in shipping services. Nasir (2014) showed that the service standards in the intermodal transport chain are transit time, security, reliability and capacity. Morever, Kaleappan (2006) measured the quality in container terminal operations which related to personal contact ordering procedures, order accuracy, order condition, order quality, order discrepancy handling, timeliness, quality, order release quality and information quality. In other words, all of these factors liaised as root causes to increase the logistics chains performance. In this study, the bottleneck factors focused more on small container depot firms that manage the empty container, connected to manpower attitudes, information, operations handling, monitoring and facilities/others. The description categories of bottlenecks factors in empty container depot are given in Table 1.

Table 2 shows sub-categories (unit) of bottlenecks. The evaluators are required to assign each item into the category which they believe to be the most appropriate. Further, the evaluators were asked if the researcher needed to improve the unit and category. The equations (Fleiss, 1971) and procedures to quantify the proportion of agreement were highlighted as follows:

$$K=\frac{\bar{P}-\bar{P}_e}{1-P_e} \quad (1)$$

$\bar{P}$ is the per cent agreement (the proportion of agreements on judgments)

P_e is the per cent chance agreement (the proportion of agreements one would expect by chance)

The overall proportion of agreement is the average of the P_i's.

Where,

$$P_i=\frac{1}{n(n-1)}\sum_{j=1}^{k} n_{ij}(n_{ij}-1) = \frac{1}{n(n-1)}\left(\sum_{j=1}^{k} n_{ij^2}-n\right) \quad (2)$$

$$\bar{P}=\frac{1}{Nn(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{k} n_{ij}(n_{ij}-1) \quad (3)$$

$$p_j=\frac{1}{Nn}\sum_{i=1}^{N} n_{ij} \quad (4)$$

$$\bar{P}_e=\sum_{i=1}^{k} p_j^2 \quad (5)$$

Where,
N = Total number of items
n = Number of rating per items
k = The number of categories
i = Subscript i, where i=1,… N is the item
j = Subscript j, where j=1,… k is the category
n_{ij} = Number of evaluators who assigned item i into category j
p_i = The proportion of agreement among n evaluators for the i^{th} item
p_j = The proportion of all assignments for the j^{th} category.

Based on the previous study, the codes/categories/themes emerged to show the bottlenecks problems. For each major theme, all the text quotes were extracted from the transcripts (approximately 80 pages of transcripts) and grouped according to the subthemes from thematic analysis. Transcripts from interviews were independently reviewed by evaluators to identify major themes in terms of strengths and suitability. They proposed 10 items be dropped because some items posed repeated words or were not having a clear meaning. This step was implemented before the Fleiss kappa was calculated. Therefore, out of the 30 units, 20 were selected for IRA assessment.

Table 2 shows the list of units (i) which represent from i1 to i20 and the example of problem (perceptions by respondents interviewed from a previous study) which linked to five main categories (k=5). These units were used in the following table to calculate the coefficient of kappa (Fleiss). Six panels (n=6) were involved to classify 20 units (n=20) into each category.

The rating scale is assumed nominal because the five categories cannot be ranked. The basic objective is to quantify the extent to which the evaluators agree on the classification of units (sub-categories) into main categories. Table 3 shows the distribution of evaluators by participating unit and response category. "Total (a)"column indicates the number of evaluators who determine the appropriate category (j1-j5) based on unit (i) given in the table. Besides that, the "total (b)" represents the total number of categories selected by the evaluators. In this study, researchers are

Table 2: Sub-categories (unit) of bottlenecks

Unit (i) and example of problems (remarked by previous interviewees)
1. Alerting and communicating of any problem "Rarely meeting or communicating between parties. Particularly, in the case of container-truck movement which caused congestion at an off-dock depot"
2. Collaborate with channel partners "Choose the depot which offers lower-cost of services"
3. Handling operations "Unable to monitor the operation at depots because of being unaware of the quantity of available containers"
4. Consistency in the distribution of accurate information "Inaccurate information (for example, information on container serial number, depots location)"
5. Provide fast response "Late repair or delay in preparing the container after a damage"
6. Online container tracking facility "Information of container movement is not updated in the system; unable to trace the location of the container or lack of system's ability to trace the container"
7. Consistency in service standard "Inconsistency in terms of container delivery"
8. Condition of container fit to the customer's needs "Condition of containers does not achieve customer specifications"
9. The placement of containers "Lack of monitoring by the authority for the placement of container near depots"
10. Suitable operating hours "Container cannot move out on specific time or drivers need to wait long hours to pick up the container"
11. Issuing information on time "System (e.g., inventory) is not updated by the depots; therefore, information is either not received or is received late by the container operator"
12. Competence of manpower "Drivers carrying incomplete documents and preferring to arrive at depot area during peak hours"
13. Containers are segregated properly "Container arrangement is not classified according to grades, operator and size"
14. All parties involved are electronically linked to container depots "The communication chain was very slow"
15. Cost of operations "Liner delay to approve costs container repair (eg. maintenance, rent, claims)"
16. Timely settlement of problems "Problem solving with slightly less effort"
17. Conduct effective meetings with logistics parties "Discussions have been made, but no action has been taken"
18. Online container booking system facility "Container depot encounters difficulties in processing the manual documentation (which is a very slow process); thus, the documentation in the number of containers booked become inaccurate"
19. Availability of resources "Insufficient resources (employee, machine and other equipment)"
20. Container in the inventory "Poor quality of container and shortages of container"

able to minimise the number of missing values while conducting this test (see Tables 3 and 4). Therefore, this method does not deal with the missing rating as all evaluators are assigned each unit for each category.

According to Table 3, it can be seen that the categories are assigned to each unit whereas the frequency for category j1 = 22 is slightly higher than j2, which accounted for 21. Category j3 and j5 represent the highest rate compared to other categories, both evaluators accounted for 27. Category j2 becomes the lowest value compared to category j4 (23). Therefore, the information in Table 3 will be used for computing the per cent agreement ($\bar{P}$), per cent chance agreement (P_e) and Fleiss' generalised kappa (Fleiss 1971).

Table 4 contains the distribution of 20 units by evaluator and response category, whereby "total" column represents the total number of units assessed or scored by the evaluators. According to Table 5, for example, item i1 shows the proportion of agreement almost 100% because all evaluators placed these items in the same category titled "facilities/others," similarly with items i2 to i5 show high agreement rate liaised with respective categories. Evaluators believed that the problem of i2 was because of many container operators collaborating with depots that offer low cost for storage and maintenance, which is ideally suited under attitude category.

Rate agreement of i6 is accounted for 66.7% titled "information." According to the evaluators, information indicates that the players in container logistics chain require real time information to check inventory and trace the location of the container. According to the problem description, the information of container movement is not updated in the system. Therefore, all the panels were clearly identified in the category. As indicated in i10 (suitable operating hours), the evaluators agree to classify this item into j3 because the panel can relate certainly to the problem expressed in Tables 1 and 2 into that category. Thus, with a total score of 66.7%, it indicates an intermediate to good (Fleiss, 1981) or moderate agreement (Landis and Koch, 1977). In case of item i18, all evaluators agreed on the placement of 46.7%, whereby (40% for j2 and 0.067% for j5), reflecting moderate (Landis and Koch, 1977) and intermediate to good (Fleiss, 1981) among evaluators. Items i7, i11, i13, i14, i19 also accounted for 100% agreement by evaluators. They believed that the problem of insufficient resources (e.g. employees, machines, and other equipment) was mostly related to the operational problem. Therefore, the evaluators clearly classified i19 into j3. To calculate $\bar{P}$, both equations 2 and 3 were involved, and finally the value of $\bar{P}$ based on Table 5 is 0.8567.

Overall, the value of proportion agreement (from each item) seems more than 0.40, whereas Altman's (1991) benchmark these values as "Fair Agreement" and "Moderate" by Landis and Koch (1977); therefore, no items need further improvement or review in this study. Furthermore, the value of multiple-rater per cent chance agreement (P_e) was calculated in equations 4 and 5 as shown in Table 6.

Table 3: Distribution of evaluators by the participating unit and response category

Units (i)	Categories (k)					Total (a)
	j1	j2	j3	j4	j5	
1					6	6
2	6					6
3				6		6
4		6				6
5	6					6
6		5	1			6
7					6	6
8			5		1	6
9				6		6
10			6			6
11		6				6
12	5		1			6
13			6			6
14					6	6
15			1	5		6
16	5				1	6
17				1	5	6
18		4			2	6
19			6			6
20			1	5		6
Total(b)	22	21	27	23	27	

Table 4: Distribution of 20 units by evaluators and response category

Evaluators (n)	Categories (k)					Total
	j1	j2	j3	j4	j5	
1	2	4	5	4	5	20
2	4	3	6	4	3	20
3	4	3	5	3	5	20
4	4	3	4	4	5	20
5	4	4	4	4	4	20
6	4	4	3	4	5	20
Average	3.67	3.5	4.5	3.83	4.5	

The regular per cent agreement of equations 2 and 3 is given by $\bar{P}$ = 0.8567.

Fleiss' per cent chance agreement is Pe = 0.2022.

Therefore, Fleiss' generalised Kappa is calculated from equation 1 and is denoted by:

$$K = (0.8567-0.2022)/(1-0.2022) = 0.8203.$$

Therefore, Fleiss' Multiple-Rater Kappa is 0.8203.

Table 5: Value of per cent agreement ($\bar{p}$)

Units (i)	Categories					Total
	j1	j2	j3	j4	j5	
1					1	1
2	1					1
3				1		1
4		1				1
5	1					1
6		0.667				0.667
7					1	1
8		0.667				0.667
9				1		0.667
10			1			0.667
11		1				1
12	0.667		1			0.667
13			1			1
14					1	1
15				0.667		0.667
16	0.667					0.667
17					0.667	0.667
18		0.4			0.067	0.467
19			1			1
20				0.667		0.667
$\bar{P}$						0.8567

Table 6: Calculation of per cent chance agreement (Pe)

Units	Categories					Total
	j1	j2	j3	j4	j5	
1					1	1
2	1					1
3				1		1
4		1				1
5	1					1
6		0.83	0.17			0.667
7					1	1
8		0.83			0.17	0.667
9				1		0.667
10			1			0.667
11		1				1
12	0.83		0.17			0.667
13			1			1
14					1	1
15			0.17	0.83		0.667
16	0.83				0.17	0.667
17				0.17	0.83	0.667
18		0.67			0.33	0.467
19			1			1
20			0.17	0.83		0.667
pj	0.18	0.18	0.23	0.19	0.23	
pj^2	0.03	0.03	0.05	0.04	0.05	0.2022 (Pe)

Table 7: Provides the most complete information on the outcome of this inter-rater agreement

Category	Unit, i (Sub-category)
Attitude (j1)	Collaborate with channel partners (i2) Provide fast response (i5) Competence of manpower (i12) Timely settlement of problems (i16)
Information (j2)	Consistency in the distribution of accurate information (i4) Online container tracking facility (i6) Issuing information on time (i11) Online container booking system facility (i18)
Operations handling (j3)	Condition of container fit to the customer's needs (i8) Suitable operating hours (i10) Containers are segregated properly; for example, in terms of size, operator and grade (i13) Resources (eg., machine, manpower) availability (i19)
Monitoring (j4)	Handling operations (i3) The placement of containers (i9) Cost of operations, e.g., maintenance, rent, claims (i15) Container in the inventory (i20)
Facilities/others (j5)	Alerting and communicating any problem (i1) Consistency in service standards (i7) All parties involved are electronically linked to container depots (i14) Conduct effective meetings with logistics parties (i17)

To summarise (Table 7), in the case of attitude, the evaluators agree to include i2, i5, i12 and i16 as suitable items into this category. However, as agreed by evaluators, items for i4, i6, i11 and i18 are classified under the category of information. Accordingly, the evaluators agree to include items i8, i10, i13 and i19 under the category of operations handling. For monitoring, evaluators provide high rate of agreement for i3, i9, i15 and i20 when compared to the other items. Finally, items i1, i7, i14 and i17 were selected into the category of facilities/others. Finally, Fleiss' Multiple-Rater Kappa is 0.8203 for all the 20 codes, which may conclude that the extent of agreement among the evaluators is very good.

5. CONCLUSION

In this study, a useful procedure for evaluating the agreement was accomplished in order to gain the level of consistency among evaluators (by allocating units into category). Fleiss kappa value of 0.8203 has proven as a right approach for demonstrating the high validity of qualitative research results (based on a previous study). Thus, this may provide as a stronger agreement among the evaluators. Based on the qualitative analysis and kappa statistics, the strength of these factors would be a guideline for researchers to solve the problem across the logistics chain players including port management, depot firm, shipping lines, manufacturers and third party logistics link to the bottlenecks. In a future study, the research will focus on studying the operational problems that link to the efficiency of the container-truck movements and processes around the depot.

REFERENCES

Abramović, B., Lovrić, I., Stupalo, V. (2012), Analysis of intermodal terminals service quality in the republic of Croatia. PROMET-Traffic and Transportation, 24(3), 253-260.

Adam, S. (2009), Simulation and Analysis of Port Bottlenecks: The Case of Male. Master Thesis. New Zealand: Lincoln University.

Altman, D.G. (1991), Practical Statistics for Medical Research (Reprint 1999). Boca Raton, Florida: CRC Press.

Antonakos, C.L., Colling, K.B. (2001), Focus on research methods: Using measures of agreement to develop a taxonomy of passivity in dementia. Research in Nursing and Health, 24(4), 336-343.

Burla, L., Knierim, B., Barth, J., Liewald, K., Duetz, M., Abel, T. (2008), From text to codings: Intercoder reliability assessment in qualitative content analysis. Nursing Research, 57(2), 113-117.

Fleiss, J.L. (1971), Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378-382.

Fleiss, J.L. (1981), Statistical Methods for Rates and Proportions. 2nd ed. New York: John Wiley.

Graham, M., Milanowski, A., Miller, J. (2012), Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. Washington, DC: Center for Educator Compensation and Reform.

Grant, L.M. (2006), Second-Generation Chinese American Women and The Midlife Transition: A Qualitative Analysis. Doctoral Thesis. Alliant International University, California School of Profession Psychology Fresno Campus.

Gudelj, A., Krčum, M., Twrdy, E. (2010), Models and methods for operations in port container terminals. Promet-Traffic and Transportation, 22(1), 43-51.

Gunasekaran, A., Ngai, E.W.T. (2003), The successful management of a small logistics company. International Journal of Physical Distribution and Logistics Management, 33(9), 825-842.

Gwet, K.L. (2012), Handbook of Interrater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters. 4th ed. United States of America: Advanced Analytics, LLC.

Hallgren, K.A. (2012), Computing inter-rater reliability for observational data: An overview and tutorial. Tutorials in Quantitative Methods for Psychology, 8(1), 23-34.

Haralambides, H., Gujar, G. (2011), The Indian dry ports sector, pricing policies and opportunities for public-private partnerships. Research in Transportation Economics, 33, 51-58.

Hruschka, D.J., Schwartz, D., St. John, D.C., Picone-Decaro, E., Jenkins, R.A., Carey, J.W. (2004), Reliability in coding open-ended data: Lessons learned from HIV behavioral research. Field Methods, 16(3), 307-331.

Jasmi, K.A. (2012), Laporan Dapatan Kajian in Kursus Penyelidikan Kualitatif Siri, Mersing, Johor, Malaysia. p1-21.

Jasmi, K.A., Tamuri, A.H. (2011), Amalan Penggunaan Bahan Bantu Mengajar dalam Kalangan Guru Cemerlang Pendidikan Islam Sekolah Menengah di Malaysia. Journal of Islamic and Arabic Education, 3(1), 59-74.

Ji, S., Zhou, R. (2010), Simulation and bottleneck analysis of container port handling resources based on bounding theory. International Conference on Logistics Systems and Intelligent Management, Harbin China. p705-708.

Jones, D.A., Farkas, J.L., Bernstein, O., Davis, C.E., Turk, A., Turnquist, M.A., Nozick, L.K., Levine, B., Rawls, C.G., Ostrowski, S.D., Sawaya, W. (2011), U.S. import/export container flow modeling and disruption analysis. Research in Transportation Economics, 32(1), 3-14.

Kaleappan, C. (2006), A Study of Logistic Service Quality in A Container Terminal Operations. Master Thesis. Malaysia: Universiti Sains Malaysia.

Kia, M., Shayan, E., Ghotb, F. (2000), The importance of information technology in port terminal operations. International Journal of Physical Distribution and Logistics Management, 30(3/4), 331-334.

Kim, H., Kim, B. (1999), Segregating space allocation models for container inventories in port container terminals. International Journal of Production Economics, 59(1), 415-423.

Lai, K.K., Lam, K., Chan, W.K. (1995), Shipping container logistics and allocation. The Journal of the Operational Research Society, 46(6), 687-697.

Landis, J., Kosh, G.G. (1977), The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.

Leporis, M., Kralova, Z. (2010), A Simulation Approach to Production Line Bottleneck Analysis. International Conference Cybernetics and Informatics, Vyšná Boca, Slovak Republic. p1-10.

Mahamod, Z., Embi, M.A. (2005), Penggunaan strategi pembelajaran bahasa untuk menguasai kemahiran membaca. Jurnal Teknologi, 42, 1-18.

Mahamod, Z., Ishak, N.H.M. (2003), Analisis Cohen Kappa Dalam Penyelidikan Bahasa: Satu Pengalaman. Seminar Penyelidikan Pendidikan Guru Peringkat Kebangsaan. p1-7. Available from: http://www.ipbl.edu.my/portal/penyelidikan/seminarpapers/2003/zamriUKMkk1.pdf.

McCray, G. (2013), Assessing inter-rater agreement for nominal judgement variables. Paper Presented at the Language Testing Forum. Nottingham: University of Lancaster. Available from: http://www.norbertschmitt.co.uk/uploads/27_528d02015a6da191320524.pdf.

Miremadi, A., Ghalamkari, S., Sadeh, F. (2011), Customer Satisfaction in Port Idustry (A Case Study of Iranian Shipping). Singapore: International Conference on Sociality and Economics Development IPEDR © IACSIT Press. p58-62.

Möller, D.P. (2014), Introduction to Transportation Analysis, Modeling and Simulation: Computational Foundation and Multimodal Applications. London: Springer-Verlag.

Namboothiri, R., Erera, A.L. (2004), Set partitioning heuristic for local drayage routing under time-dependent port delay. IEEE International Conference on Systems, Man and Cybernetics, 1(7), 3921-3926.

Nasir, S. (2014), Intermodal Container Transport Logistics to and From Malaysian Ports: Evaluation of Customer Requirements and Environmental Effects. Doctoral Thesis. Stockholm, Sweden: Royal Institute of Technology.

Rankine, G. (2003), Benchmarking Container Terminal Performance, Container Port Conference, Rotterdam.

Salekan, I.K. (1997), A Study on Problems Affecting Import Movements Via Full Container Load (FCL), Case Study: Green Peninsula Agencies. Bachelor Thesis. Shah Alam, Malaysia: Mara Institute of Technology.

Schaer, P., Mayr, P. (2010), Implications of inter-rater agreement on a student information retrieval evaluation. ArXiv. p1-7. Available from: http://www.arxiv.org/abs/1010.1824.

Shweta, R.C., Bajpai, R.C., Chaturvedi, H.K. (2015), Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. Journal of the Indian Academy of Applied Psychology, 41(3), 20-27.

Soares, W.L., Pozo, H., Souza, C.P., Akabane, G.K., Lopes, C.P. (2013), Strategies in networks of collaborative transportation systems with RFID for oil rigs: A case study of technological innovations for a depot of empty containers. 24th Annual POM Conference Denver, Colorado, USA. p1-10.

Tamam, T., Nik Mohd Zamri, M., Rahimi, N.Y., Jamaludin, B. (2010), Masalah Membaca Murid-murid Sekolah Rendah Kerajaan di Brunei Darussalam : Satu Kajian Kes. Jurnal Pendidikan Malaysia, 35(2), 77-85.

Thestar. (2012), Available from: http://www.thestar.com.my/news/nation/2012/05/09/truckers-strike-in-port-klang-cost-acccim-members-rm50mil/. [Last accessed on 2012 May 09].

Timbang, T., Mahamod, Z., Hamat, A. (2011), Faktor dan kesan masalah membaca dalam kalangan murid Sekolah Rendah Kerajaan di Brunei Darussalam. Jurnal Pendidikan Bahasa Melayu, 1(1), 89-107.

Umar, S. (2004), The Development of Third Party Logistics in Malaysia: An Overview. Master Thesis. Malaysia: Malaysia University of Science and Technology.

UNCTAD. (2014), Maritime Transport Review, United Nations Conference on Trade and Development, United Nations Publication. Available from: http://www.unctad.org/en/PublicationsLibrary/rmt2014_en.pdf.

Vacca, I., Bierlaire, M., Salani, M. (2007), Optimization at container terminals: Status, trends and perspectives. In: Proceedings of the 7th Swiss Transport Research Conference. Monte-Verita, Ascona, Switzerland. p1-21.

Vathanophas, V., Thai-ngam, J. (2007), Competency Requirements for effective job performance in the Thai public sector. Contemporary Management Research, 3(1), 45-70.

Yeo, G.T., Thai, V.V., Roh, S.Y. (2015), An analysis of port service quality and customer satisfaction: The case of Korean container ports. Asian Journal of Shipping and Logistics, 31(4), 437-447.

Yin, R.K. (1994), Case Study Research: Design and Methods. 2nd ed. Thousand Oaks, CA: Sage Publications.

Yusri, G., Rahimi, N.M., Shah, P.M., Wah, W.H., Hassan, A.T. (2011), Kawalan Kepercayaan Pembelajaran Dalam Kalangan Pelajar Kursus Bahasa Arab Sebagai Bahasa Ketiga. ASEAN Journal of Teaching and Learning in Higher Education, 3(2), 41-47.

Yusri, G., Rahimi, N.M., Shah, P.M., Wah, W.H., Hassan, A.T. (2012), The use of peer learning strategy among Arabic language course students at Universiti Teknologi Mara. Asia Pacific Journal of Educators and Education, 27(2012), 37-50.

Zain, R.M., Ab. Rahman, M.N., Nopiah, Z.M., Saibani, N. (2014a), Understanding of empty container movement: A study on a bottleneck at an off-dock depot. AIP Conference Proceedings, SORIC, Sarawak, Malaysia.

Zain, R.M., Ab. Rahman, M.N., Nopiah, Z.M. (2014b), Quality logistics management: Managing empty container movement at the off-dock depot. Conference Proceedings, 18-ICIT, Sarawak, Malaysia.

Zain, R.M., Ab. Rahman, M.N., Nopiah, Z.M., Saibani, N., Ramli, A., Jusoh, N. (2015), Exploring the delays at empty container off-dock depots: Useful perceptions by stakeholders. Management, 5(5), 148-159.

Zenk, S.N., Schulz, A.J., Mentz, G., House, J.S., Gravlee, C.C., Miranda, P.Y., Kannan, S. (2007), Inter-rater and test-retest reliability: Methods and results for the neighborhood observational checklist. Health and Place, 13(2), 452-465.