Simulation Study of Item Validity Testing and Item Discrimination Index

Adi Setiawan
Department of Mathematics, Faculty of Science and Mathematics
Satya Wacana Christian University, Jl. Diponegoro 52-60 Salatiga 50711 Indonesia
E-mail : adi_setia_03@yahoo.com

Abstract— In psychology or educational research, we often want to test the validity of our instrument. Usually we use the Pearson coefficient of correlation in validity testing. However, the distribution of the data are not always normal distribution so it is also possible to use the Spearman and Kendall coefficient of correlation to test the validity of items. Valid items tend to have a positive and significant coefficient of correlation. It means the item is likely to be answered correctly by people who have a high total score and tend to be answered incorrectly by people who have a low total score. Item discrimination index is an index that measures the ability of item to be able to discriminate students who have high learning outcomes and students who have low learning outcomes. Simulation studies provide a clearer description and show that the item discrimination index has a significant relationship to validity testing that is commonly used in research.

Keywords — *validity testing, item discrimination index, Pearson coefficient of correlation, Spearman coefficient of correlation, Kendall coefficient of correlation*.

I. INTRODUCTION

We often use the Pearson correlation coefficient to test the validity of instrument tool in psychology and education research. However, the distribution of the data are not always normal distribution, it is also possible to use the Spearman and Kendall coefficient of correlation to test the validity of items. Valid items tend to have a positive and significant coefficient of correlation. It means the item is likely to be answered correctly by people who have a high total score and tend to be answered incorrectly by people who have a low total score. Item discrimination index is an index that measures the ability of item to be able to discriminate students who have high learning outcomes and students who have low learning outcomes [1]. Recent papers related to this paper are presented in [2] and [3].

In this paper it will be presented simulation studies on the relationship between the validity testing and item discrimination index to describing the relationship between the validity and item discrimination index that is widely used in the item analysis. The data used in the simulation is only types 0 and 1 data.

II. LITERATURE REVIEW

In the literature review it will be explained about the various kinds of correlation coefficient i.e. Pearson coefficient of correlation, Spearman coefficient of correlation and Kendall coefficient of correlation. Furthermore, it is also explained item discrimination index.

TABLE 1. RESPONSES OF 10 EXAMINEES TO 5 ITEMS, DICHOTOMOUSLY SCORED.

Examinees	1	2	3	4	5	Total Score
1	0	0	1	1	0	**2**
2	0	0	0	1	0	**1**
3	1	1	0	0	0	**2**
4	1	0	0	1	0	**2**
5	0	1	1	1	1	**4**
6	0	1	0	0	0	**1**
7	1	1	1	1	1	**5**
8	1	1	0	1	0	**3**
9	1	1	1	1	0	**4**
10	0	0	0	1	1	**2**
Total	**5**	**6**	**4**	**8**	**3**	

Suppose (X_1,Y_1), (X_2,Y_2), …., (X_n,Y_n) are bivariate random sample size n that is taken from a certain population. Pearson coefficient of correlation is defined by

$$\rho_{X,Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

where $\mu_X = E[X]$, $\mu_Y = E[Y]$, $\sigma_X = \sqrt{V(X)}$ and $\sigma_Y = \sqrt{V(Y)}$. Estimation of Pearson coefficient of correlation based on the sample can be found by :

$$r = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}}$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

To give an idea how to use coefficient of correlation in validity testing, the data is given in Table 1 which consists of 5 items of question answered by 10 examinees of the test [4]. A value of 1 means that the examinee responded correctly to an item and a value of 0 means the examinees responded incorrectly to the item. Each item scores has a Pearson correlation coefficient as follows 0.4685, 0.5419, 0.7332, 0.4295, 0.5453, respectively. The correlation coefficient is significant if more than 0.632 so that the only valid item is 3. However, it is only used to provide an overview in using of the Pearson correlation coefficient.

Spearman coefficient of correlation can be regarded as a measure of relationship between two ordinal variables or measures of the degree of relationship between the data that have been prepared according to the ranking. Let (X_1,Y_1), (X_2,Y_2), …., (X_n,Y_n) is a bivariate sample of size n. To calculate the Spearman correlation coefficient rankings are prepared in advance of the entire sample pairs X_i and Y_i then Spearman coefficient of correlation were calculated using the following formula:

$$r_S = 1 - \frac{6\sum_{i=1}^{n}[R(X_i)-R(Y_i)]^2}{n(n^2-1)}$$

where r_S is Spearman coefficient of correlation, $R(X_i)$ is rank of X_i and $R(Y_i)$ is rank of Y_i. This formula is used when there is no ties in the data.

Spearman correlation coefficients were used in the R package program based on the formula [5] :

$$r_S = \frac{A+B-\sum_{i=1}^{n}[R(X_i)-R(Y_i)]^2}{2\sqrt{AB}}$$

where $A = \frac{n^3-n}{12} - \Sigma T_X$, $B = \frac{n^3-n}{12} - \Sigma T_Y$ and ΣT_X is the number of tied observations in each group of ties on the X variable and $T = (t^3\text{-}t)/12$. Suppose the data is given as follows { (1,2), (0,1), (0,2), (0,2), (1,4) }. It means X is { 1, 0, 0, 1} so that $R(X_i)$ is {4.5, 2, 2, 2, 4.5 } and

$$\Sigma T_X = \frac{(4.5)^3-4.5}{12} + \frac{2^3-2}{12} = 2.5$$

(there exist 2 numbers with ties i.e. rank 4.5 and rank 2 with 3 ties). Furthermore, data Y is { 2, 1, 2, 2, 4} so that $R(Y_i)$ is {3, 1, 3, 3, 5 } and $\Sigma T_Y = \frac{3^3-3}{12} = 2.5$ (only one datum that has ties rank i.e. 3). Finally, we have

$$\sum_{i=1}^{5}[R(X_i)-R(Y_i)]^2 = 5.5,$$

A = 7.5 and B = 8 and we have Spearman coefficient of correlation 0.6455.

The command that can be used to obtain Spearman coefficient of correlation in package program R is *cor(x, y, method = "spearman")*. Based on Table 1, the Spearman coefficient of correlation between the scores of items and total score are 0.5061, 0.5166, 0.7011, 0.4519, 0.5128 respectively. An example of paper that use Spearman coefficient of correlation is [6].

Kendall Tau Correlation Coefficient

Kendall rank correlation coefficients without ties can be calculated by using the following formula [5] :

$$\tau = \frac{2K}{n(n-1)}$$

where

$$K = \sum\sum_{i<j} \mathrm{sgn}(x_i - x_j)\,\mathrm{sgn}(y_i - y_j)$$

$\mathrm{sgn}(x) = 1$ if $x > 0$, $\mathrm{sgn}(x)$ if $x = 0$ and $\mathrm{sign}(x) = -1$ if $x < 0$. Kendall rank coefficient of correlation with ties can be calculated as follows:

$$\tau = \frac{K}{\sqrt{\frac{1}{2}n(n-1)-T_X}\sqrt{\frac{1}{2}n(n-1)-T_Y}} \qquad (1)$$

where

$$T_X = \sum_{k=1}^{n1} t_X(t_X - 1)$$

$$T_Y = \sum_{k=1}^{n2} t_Y(t_Y - 1)$$

t_X = the number of objects of the same rank more than one variable X_i,
t_Y = the number of objects of the same rank to more than one variable Y_i,
i = 1, 2, ..., n,
m = the number of items,
n = sample size.

Using the same data as in the above example is {(1,2), (0,1), (0,2), (0,2), (1,4)}. It means the data X is {1, 0, 0, 0, 1} and the data Y is {2, 1, 2, 2, 4} so that $\mathrm{sgn}(x_i\text{-}x_j)\,\mathrm{sgn}(y_i\text{-}y_j)$ for i = 1, 2, ..., 5; j = 1, 2, ..., 5 and can be expressed in matrix form :

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

We have $K = \sum_{i<j}\sum \operatorname{sgn}(x_i - x_j)\operatorname{sgn}(y_i - y_j) = 4$ which is the sum of all elements in the matrix below the main diagonal. Finally, by using equation (1) Kendall coefficient of correlation is 0.6172.

Kendall coefficient of correlation can be obtained by using command *cor(x, y, method = "kendall")*. Based on Table 1, the Spearman correlation coefficient between the scores of items and total score are 0.4603, 0.4698, 0.6376, 0.4110, 0.4664 respectively.

Item Discrimination Index

Item discrimination index is the ability of an item to be able to discriminate between high-ability examinees with the low-ability examinees. It means that the high-ability examinees is examinees who have a high ability to answer more items correctly, meanwhile the low-ability examinees is examinees who have a low-ability to answer items correctly.

Item discrimination index can be expressed as

$$D = p_H - p_L$$

where p_H is the proportion of higher group who answered the item correctly and p_L is the proportion of lower group who answered the item correctly. In this case,

$$p_H = \frac{B_H}{N_H}$$

where B_H is the number of higher group who answered the item correctly and N_H is the number of lower group who answered the item correctly;

$$p_L = \frac{B_L}{N_L}$$

where B_L is the number of lower group who answered the item correctly and N_L is the number of lower group who answered the item correctly. Item dicrimination index can also be defined as follows :

$$\phi = \frac{p_H - p_L}{2\sqrt{pq}}$$

where ϕ is the item discrimination index numbers, p denote the proportion of examinees who answered correctly to the item (so that q = 1- p denote the proportion of examinees who answered incorrectly to the item) [1]. However, the index must lie between -1 and 1.

TABLE 2. THE RESULT OF SORTED SCORE TOTAL BASED ON TABLE 1.

Examinees	1	2	3	4	5	Total Score
7	1	1	1	1	1	5
5	0	1	1	1	1	4
9	1	1	1	1	0	4
8	1	1	0	1	0	3
1	0	0	1	1	0	2
3	1	1	0	0	0	2
4	1	0	0	1	0	2
10	0	0	0	1	1	2
2	0	0	0	1	0	1
6	0	1	0	0	0	1
Total	**5**	**6**	**4**	**8**	**3**	

When we want to find coefficient of correlation between the values of the correlation coefficient (Pearson, Spearman and Kendall) and item discrimination index of each item on the data in Table 1, then we obtain 0.7726, 0.8619, 0.8619, respectively. The critical point of significant value of correlation coefficient is 0.755. We see that the correlation coefficient is significant so that the item discrimination index is very closely related to the three types of the correlation coefficient.

III. RESEARCH METHODS

Simulation studies are carried out with the following steps :

1. Data is in matrix form with order $m \times n$ and it is generated by a Bernoulli distribution with parameter p where p denote the probability of the examinees answered correctly for each item. There are m examinees and n items in the test. It is considered that the ability of examinees to answer each question independently. In this simulation study m = 30, 50,

100 and 1000, while the number of test items n = 10, 20, 30, 40, 50 and 100.

2. Total score of examinees has a normal distribution such that if $X_1, X_1, \ldots, X_n$ are the total score then

$$Z_i = \frac{X_i - \overline{X}}{s}$$

has a standard normal distribution where $\overline{X}$ and s are mean and standard deviation, respectively. Thus, the probability of examinee to answer every item correctly is $\Phi(Z_i)$ for $i = 1, 2, \ldots, m$ where m is the number of examinees. In this simulation study we use m = 30, 50, 100 and 1000, while the number of test items n = 10, 20, 30, 40, 50 and 100.

Simulation that is used in this paper is Monte Carlo simulation. Recent related paper is [7].

IV. RESULTS AND DISCUSSION

The simulated data is generated by using first procedure as described in section methods. If it is used m = 10, n = 5 and p = 0.5, the simulation results is shown in Table 3. Based on Table 3, item discrimination index are 0.6667, 0.0000, 1.0000, 0.3333, 0.6667 respectively; Pearson's correlation coefficient is 0.7534, 0.0184, 0.7167, 0.2063, 0.8251, Kendall correlation coefficients are 0.7008, 0.0000, 0.6639, 0.20702, 0.7591 and Spearman coefficient of correlation 0.7570, 0.0000, 0.7171, 0.2236, 0.8199. In addition, the average proportion of correct answers for each item is 0.54 which is close to the value of the used parameter p = 0.5.

If the results obtained from different power attributed to each value of the correlation coefficient will be obtained correlation coefficients (Pearson) of 0.8883, 0.8940, 0.8940. We see that coefficient of correlation between item discrimination and Kendall and Spearman correlation coefficient always have the same value. If the generated data is done for m = 30 and n = 20, then calculated the correlation coefficient between item discrimination index and the Pearson coefficient of correlation and Kendall coefficient of correlation of each of item and the procedure is repeated in large number of replication B times, it will get the results of the values of correlation coefficients (Pearson) as presented in Fig. 1 for p = 0.1, 0.3, 0.5 and 0.7. In this case we use B = 1000. It is seen that the zero point is not contained in the histogram - histogram so that there is a significant relationship between them.

If the second procedure is used to generate samples with m = 30, n = 20 and B = 2000 replication it will be obtained the histogram of (Pearson) correlation coefficients values between item discrimination index and the Pearson correlation coefficient and Kendall correlation coefficient as presented in Fig. 2.

TABLE 3. THE RESULT OF SIMULATION BY USING m = 10 , n = 5 ITEMS AND p = 0.5.

Examinees	1	2	3	4	5	Total Score
1	1	0	1	1	1	**4**
2	1	1	0	0	1	**3**
3	1	1	1	0	1	**4**
4	0	1	0	0	0	**1**
5	1	0	0	1	0	**2**
6	0	1	0	0	0	**1**
7	0	1	0	1	0	**2**
8	1	1	1	0	1	**4**
9	1	1	0	1	1	**4**
10	1	0	0	0	1	**2**
Total	**7**	**7**	**3**	**4**	**6**	

Fig. 1. Histogram of coefficient of correlation between item discrimination index and Pearson and Kendall coefficient of correlation by using first procedure with $m = 30$, $n = 20$ and $B = 2000$.

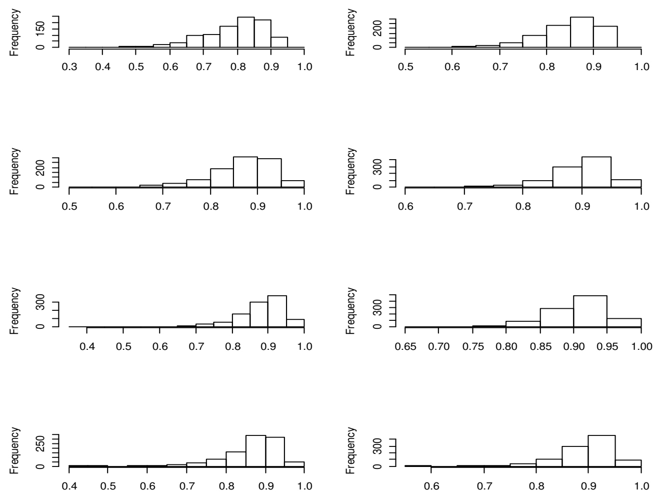

Fig. 2. Histogram of coefficient of correlation between item discrimination index and Pearson and Kendall coefficient of correlation by using second procedure with $m = 30$, $n = 20$ and $B = 2000$.

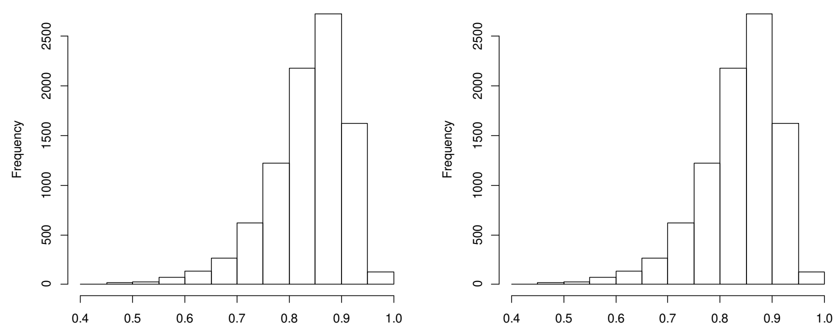

Table 4 presents the result of relation (in Pearson coefficient of correlation) between Pearson or Kendall coefficient of correlation and item discrimination index for several m, n and p. We see that there is significant relation between coefficient of correlation and item discrimination index. The last two columns of Table 4 presents the results of relation between different power index with a coefficient of Pearson and Kendall correlation coefficients for various m and n by using second procedure. The close relationship is also supported by the results obtained using both procedure/methods.

V. CONCLUSION

In this paper it has been described simulation studies to present the relationship between item discrimination index and validity testing of item by using Pearson (Kendall or Spearman) coefficient of correlation. The (simulation) study can be done also to the other data such as data with a Likert scale in the next following research.

REFERENCES

[1] Sudijono, A., 2007, *Pengantar Evaluasi Pendidikan* Jakarta: PT. King Grafindo Persada.

[2] Setiawan, A., 2014, (Monte Carlo) Resampling Technique in Validity Testing and Reliability Testing, *International Journal of Computer Application*, Vol. 91, No. 5, 6-11.

[3] Berkowitz, D. & Caner, M. & Fang, Y. , 2012. "The validity of instruments revisited," Journal of Econometrics, Elsevier, vol. 166 (2), pages 255-266.

[4] Croker, L & J. Algina, 2008, *Introduction to Classical and Modern Test Theory*, Cengage Learning, Mason Ohio.

[5] Siegel, 1956, *Nonparametric Statistics for the Behavioral Science*, McGraw-Hill Book Company Inc, New York.

[6] Yap, C. K, 2013, Can the Number of Correlation Coefficient (R Value) as a Statistical Indicator of Metal-Polluted Ecosystem? A Comparison

between Polluted and Unpolluted Ecosystems, *International Journal of Advances in Applied Sciences* Vol 2 No 4 pages 209-216.

[7] Elabd, Ali A., Abdel-Aziz T. Shalaby, El-Sayed M. El-Rabaie, 2012, Monte Carlo Simulation of Single Electronics Based on Orthodox Theory, *International Journal of Nano Devices, Sensor and Systems (IJ-Nano)* Vol 1. No. 2, pp 65-76.

TABLE 4. RESULT OF RELATIONSHIP (IN PEARSON COEFFICIENT CORRELATION) BETWEEN PEARSON AND KENDALL COEFFICIENT OF CORRELATION AND ITEM DISCRIMINATION INDEX FOR SEVERAL *m*, *n* AND *p*.

		Pearson	**Kendall**	**Pearson**	**Kendall**	**Pearson**	**Kendall**	**Pearson**	**Kendall**	**Pearson**	**Kendall**	**Pearson**	**Kendall**
n	***m***	***p* = 0.1**	***p* = 0.1**	***p* = 0.3**	***p* = 0.3**	***p* = 0.5**	***p* = 0.5**	***p* = 0.7**	***p* = 0.7**	***p* = 0.9**	***p* = 0.9**		
20	30	0.79	0.85	0.87	0.90	0.86	0.88	0.87	0.90	0.80	0.87	0.82	0.84
	40	0.79	0.85	0.87	0.90	0.88	0.91	0.87	0.90	0.80	0.85	0.84	0.85
	50	0.80	0.86	0.87	0.90	0.88	0.91	0.87	0.90	0.80	0.86	0.85	0.86
	100	0.80	0.87	0.88	0.90	0.88	0.91	0.88	0.90	0.80	0.87	0.84	0.84
	1000	0.80	0.87	0.87	0.90	0.89	0.92	0.88	0.90	0.80	0.87	0.84	0.85
30	30	0.82	0.87	0.88	0.91	0.88	0.91	0.88	0.91	0.82	0.86	0.83	0.85
	40	0.82	0.87	0.88	0.91	0.89	0.92	0.88	0.91	0.82	0.87	0.85	0.86
	50	0.82	0.87	0.88	0.91	0.89	0.92	0.88	0.91	0.82	0.88	0.86	0.87
	100	0.82	0.87	0.88	0.91	0.89	0.92	0.88	0.91	0.82	0.88	0.85	0.86
	1000	0.82	0.87	0.89	0.92	0.89	0.92	0.88	0.91	0.82	0.87	0.86	0.86
50	30	0.84	0.88	0.88	0.91	0.89	0.92	0.88	0.91	0.84	0.88	0.84	0.86
	40	0.84	0.89	0.88	0.92	0.89	0.92	0.89	0.92	0.84	0.89	0.86	0.87
	50	0.85	0.89	0.89	0.92	0.89	0.92	0.89	0.92	0.85	0.89	0.86	0.87
	100	0.85	0.90	0.89	0.92	0.89	0.92	0.89	0.92	0.85	0.90	0.86	0.86
	1000	0.85	0.90	0.89	0.92	0.89	0.92	0.89	0.92	0.86	0.90	0.86	0.86
100	30	0.86	0.89	0.89	0.92	0.89	0.92	0.89	0.92	0.87	0.90	0.85	0.86
	40	0.86	0.90	0.89	0.92	0.89	0.93	0.89	0.92	0.86	0.90	0.86	0.87
	50	0.86	0.90	0.89	0.92	0.90	0.93	0.89	0.93	0.86	0.90	0.87	0.88
	100	0.87	0.91	0.89	0.92	0.90	0.93	0.89	0.92	0.87	0.91	0.86	0.87
	1000	0.87	0.92	0.89	0.92	0.89	0.92	0.89	0.92	0.88	0.90	0.87	0.91