International Journal of
Science and Engineering Investigations vol. 2, issue 17, June 2013
ISSN: 2251-8843

Application of Meta Analysis in Examination of Validity Items

Syuul. T. Karamoy
Lecturer Staff of PGSD-FIP-Manado State University, Jl.Kaaten Matani I Tomohon
(lientjekaramoy@yahoo.com)

Abstract- Process validation and reliability of the scale enumeration of psychology Considered to be one of important step. This ascription can comprehend because compilation and examination of scale is step which is very Determine analysis and conclusion of research. This step is important even same step with him before all that is infusion of concepts into indicators or into scale. Mistake in operational concept will cause fatal for the entirety of conclusion of research. This matter will happened especially if scale of do not measure what ought to be a measured and not also give constancy of result can be pledged.

Keywords- *Meta-Analysis, Validity Items*

I. INTRODUCTION

In the 21st century, the development of science and technology was the faster. The rapid development cannot be separated from donations from the results of research, both basic and applied research. Psychology study area either directly or indirectly affected by these developments. In terms of the more quantitative studies conducted on the same topic would increase the possibility of variation in the results or conclusions of research. It is not even rare study on the same topic show conflicting results. This situation is especially problematic course of constructing a comprehensive theory or make it as a decision-making basis. In terms of qualitative data is a description of the answer is that the form of words and are arranged in a sentence. So instead of the numbers that are prevalent in quantitative approaches. Therefore, to enhance the research results, it is necessary union between quantitative and qualitative approaches.

Meta-analysis of studies performed to overcome problems in the fields of social sciences, including psychology. The findings initially seem contradictory and difficult to accumulate eventually become more integrative and systematic meta-analysis. Thus integrating the findings into a solid foundation for the development of theory and policy and decision making.

The main task is a scientific analysis to find out scientific truths that are objective, verifiable and communicated to fulfill its function, i.e.: create a description, explaining, development theory, make predictions and perform control. To be able to meet the scientific functions, the results of scientific research requires a systematic method for integrating findings or research results.

But the reality has been known that the decision made by the decision makers are usually based on the amount of research on a particular phenomenon. Because of the various limitations of single studies, is extremely rare decision that is strategic especially based on the results of a single study, whether derived from survey research, experimental field or laboratory experiments. The limitations on this single study can occur due to the weakness in the measuring instruments used, method of measurement, time measurement, research subjects and other factors are not specified.

Process validation and counting of psychology scale reliability is considered as one important step. This assumption can be understood as the preparation and testing of a scale is a crucial step analysis and conclusion of the study. This step is even as important as the previous step is pouring the concepts into indicators or to scale. Errors in operationalizing the concept would be fatal to the overall conclusion of the study. This will happen especially if the scale does not measure gauge what should be measured and did not provide reliable results constancy. Thus researchers can get stuck to reject the null hypothesis when should have to accept (Type I error). Alternatively, researchers may fail to reject the null hypothesis that should have rejected (Type II error).

Sometimes researchers feel the need to prepare yourself for the psychological scale research. It's only natural if the scale of psychology itself is compiled with empirical evidence regarding the validity and reliability. There are also researchers who utilize psychological scales developed by other researchers. Problems arise when researchers use a scale of adaptation or owned by other researchers. Such scales are usually equipped validity and reliability. If the scale has been tested other problems arise is whether we need to do a re-test reliability and validity?. Many of the arguments put forward to answer that question. One common argument leveled at sample differences. Most of the users use the scale and the number of samples with different characteristics possessed by the constituent scale samples. Perhaps for this reason as the basis of the "necessity" to do the retesting.

From the author's experience of many researchers conduct retesting psychological scales or scale adaptations of others who have been considered 'raw'. Retesting results are often different from the test results obtained by the scale authors. As an illustration we can look at Table I below. Scale used for

comparison is the anxiety scale compiled by Janet Taylor at Northwestern University in 1955 (I. Made, S, 2001). Testing scale popularly known TMAS (Taylor Manifest Anxiety Scale) did not show similarity between the results of the researchers and other researchers. Testing validity and reliability in the table using the data obtained by I Made. S (2001) Rusmin.M (2003) and Debby.K (2003). Though I.Made.S (2001) and Debby.K (2003) to obtain the number of correlation with different results, but they can use the full 50 item TMAS because all valid stated item minimal at significance level of 5 percent. On the other hand, Rusmin M (2003) using the basic item significantly correlated with the total score took the decision to abort 3 items and only use 47 items declared invalid. Decision of course is arbitrary though using the significance test as a basis for decision-making. Though perhaps true that the characteristics of the subjects are item trials are the variables that can be used to explain differences in these results. item but not abortion is a very arbitrary inequalities increase the likelihood of scale and interpretation of results.

TABLE I. SOME TEST RESULTS VALIDITY ITEM TMAS

Research	The number and extent of item Significance		
	p >0.05	p > 0.05	p > 0.05
I Made S (2001)	--	35	15
Rusmin M (2003)	3	17	30
Debby K (2003)	--	46	4

As far as the theoretical basis used in calculating the parameter is classical theory items, the sample is a problem that always accompanies the reliability and validity testing item (Azwar. 1994). How can we prove that the samples do not make us get the results fluctuate?. Can we obtain a value that can be used to summarize the differences that occur? This paper seeks to present an alternative to the two problems. So far the technique can be used to solve the problem is a meta-analysis. Therefore, in the next section will be presented a brief description of the meta-analysis. Will then be presented on the use of one technique procedures in meta-analysis to examine differences in the sample. In the final section will discuss the use of constraint meta-analysis techniques in the research.

II. A BRIEFLY OF META-ANALYSIS

Fast-growing social science allows for the assessment of a topic that is reviewed from various aspects or a variety of approaches. Many attempts to, by empirical verification or replication of a topic allow assessing the constancy or inconstancy research. More and more studies are conducted frequently would increase the variation in the results or conclusions of research. Attempts to summarize the research findings and concludes range have been carried out. Meta-analysis is an attempt to summarize the results of quantitative research. That is, the meta-analysis as a technique aimed to re-analyze the results of the study were processed statistically based on primary data collection.

According to Hedges and Olkin (1985). Meta-analysis is a statistical technique or quantitative methods are used to synthesize the results of research. The term meta-analysis was first introduced by Glass in his article published in 1976 in the journal Educational Researcher (Hedges and Olldn 1985; Hunter and Schmidt, 1990). Discussed in the article for the first time the concept of the primary analysis (analysis of primary data), secondary analysis (analysis of data collected other researchers), and meta-analysis (an analysis of the results of statistical analysis of other researchers) in the study. Hedges and Olkin (1985) meta-analysis techniques to categorize as a secondary analysis technique, because this technique focuses on the efforts the results of the primary analysis.

Bangert-Drowns (1986) tried to do a review of some basic methods of meta-analysis. He tried to do a comparison of the Glass method, method of Rosenthal, Hunter-Schmidt method, and the method of Cooper. The comparison is done based on writings that have been recorded by the three researchers that Glass, published book's in 1981, Rosenthal with his book published in 1984, and Cooper with his book published in 1984. Glass method is a method which seeks to examine the large-small effect sizes (sampling error). Rosenthal give suppression method to study differences in the value of p or $\propto$ from a variety of studies. Hunter-Schmidt method of assessing not only the influence of effect-sizes. As performed by Glass, but also includes the assessment of the variance of the effect-sizes as well as the influence of measurement error. While Cooper method gives more emphasis to the reorganization of research results and less attention to statistical techniques for this purpose.

Meta-analysis method and its application is growing rapidly in many areas until now. According to Johnson, Mullen, and Salas (1995), in the development of the final turns out there are three methods of meta-analysis is considered as a meta-analysis approach to principal. The first is the method in pointed out by Hedges and Olkin (meta-analysis techniques Hedges-Olkin). The second is the meta-analysis technique proposed by Rosentahal and Rubin (meta-analysis techniques Rosenthal-Rubin). Third methods are proposed by Hunter and Schmidt (meta-analysis techniques Hunter-Schmidt). Among the three, it turns out Rosenthal-Rubin technique has the longest history as these techniques have been developed 15 years before the term meta-analysis presented by Glass. Summary of the three main techniques of meta-analysis are as follows.

A. Technical Rosenthal-Rubin

This technique is actually a technique of most previous efforts that have been done to integrate the research results. Rosenthal seeks to develop techniques ever developed early among others by Fisher, Pearson, and Thorndike. The basic assumption used this technique is to change the results of the study into a standard score (z score) in connection with the conversion of the normal curve and the curve of the correlation coefficient to a standard score of Fisher. Conversion results are then combined to produce a weighted average of several studies that investigated.

B. Hunter-Schmidt technique

This technique was developed at the same time with the technique proposed by Glass and their colleagues. Technique proposed by Hunter and Schmidt based studies in the areas of industrial and organizational psychology is considered as the most sophisticated approach. In addition to offering techniques to examine the effect-size (sampling error), this technique can also be used to assess measurement error, the reliability of the dependent variable, and independent variable reliability. After sampling error can be assessed and determine its effect on various studies that investigated, this technique offers the following steps to perform other tests.

C. Engineering Hedges-Olkin

This technique is highly influenced by the approach proposed by Glass through the introduction of the coefficient g. Coefficient g is meant by the Glass is the estimation of effect sizes, the difference between the average standard control group and the experimental group. This approach seeks to convert the results of previous studies into standard deviation units called coefficient g is then used to correct the bias research. Value is referred to as a result of the transformation g is then combined, are reviewed for consistency and variability explained by using a model (continuum or categorical).

Of the three approaches, Hunter-Schmidt technique turned out a lot more techniques associated with the validation problem. Technique known as validity generalization turns out to have the basic approaches developed from Hunter-Schmidt approach this. Validity generalization is a procedure for estimating the average (mean) and variance of the validity of the pure (true validity) studies were examined by way of correction include the measurement error and sampling error (Schmidt and Hunter in Osburn and Callender. 1990).

According to Hunter and Schmidt (1990) there are 11 artifacts that can be used as criteria to understand why there are differences in the results of research on the same topic, namely (1) sampling error, (2) measurement error in the dependent variable, (3) measurement error in independent variables, (4) the dependent variable continuous dichotomization, (5) the independent variable continuous dichotomization, (6) variations in the distribution of the independent variables, (7) variations in the distribution of the dependent variable, (8) deviation from the right to construct the independent variables, (9) deviation from the proper construct the dependent variable, (10) reporting errors or transcriptional errors, and (11) the variance external factors. Among the artifacts is the eleventh most relevant to testing the validity item is the first artifact, that errors that arise due to differences in the sample. In the following section will describe the procedures for testing the sampling error using the Hunter-Schmidt technique.

III. SAMPLE TESTING DIFFERENCE

Item validity testing, better known by the name of testing the item validity (Anwar, 1994) using the correlation technique as a common testing methods. In fact, the correlation coefficient is additive and not systematic was influenced by the sampling error (Hunter and Schmidt, 1990). By using classic approach, Hunter and Schmidt (1990) says that if the actual correlation coefficient denoted by p and the correlation coefficient based on the sample denoted by r. Thus the sampling error is symbolized as ℓ can be added to the following equation: r = + ℓ

The small-large of sample error is determined by the small-large of size samples. The following are the equations proposed by Hunter and Schmidt (1990) to test whether differences in variance and correlation are derived from studies that have in common with each other or even show a significant difference. So far, said by the experts that the second coefficient of variance testing differences and correlations based on the results of the calculation have the same distribution with chi-square distribution (x^2)

1. Testing differences in correlation figures through test(x^2)

$$x^2\kappa - 1 = \sum i \frac{(N_i - 1)(r_i - r)^2}{(1 - r)} \quad (1)$$

Degrees of freedom equal to (K-1)

K is the number of comparisons (studies) were conducted.

2. To get the average r ($\bar{r}$) use the following formula

$$\bar{r} = \sum i \frac{(N_i \times r_i)}{\Sigma N_i} \quad (2)$$

3. Average sample ($\bar{N}$) is obtained by the formula

$$(\bar{N}) = \frac{\Sigma(N_i)}{K} \quad (3)$$

IV. DIFFERENCE MEASUREMENT SAMPLE ILLUSTRATION

As an illustration of the use of equations presented in the previous section presented a table that illustrates the steps to obtain a chi-square. The first step is to calculate the sample mean ($\bar{N}$) and the average correlation coefficient ($\bar{r}$). Table 2 are used to illustrate the calculation of ($\bar{N}$) and ($\bar{r}$). The first column is the correlation coefficient total item (corrected) were obtained from 10 samples of the study. The second column is the number of samples for each correlation coefficient. While the last column is the correlation coefficient multiplying each by the number of that sample.

TABLE II. CORRELATION ITEM-TOTAL OF 10 DIFFERENT SAMPLES

r	N	(N x r)
0,3810	65	24,7650
0,2493	82	20,4426
0,2499	71	17,4426
0,3880	37	14,3560
0,4240	41	17,3840
0,2807	70	19,6490
0,1102	36	3,9672
0,0846	42	3,5532
0,5310	43	22,8330
0,0895	41	3,6695
*	528	148,3624
Mean	52,8	0,280989

Source: Jeiner. R (2001, 2007)

Table 2 we obtain the total sample and the total product of the number of samples and the results addition of multiplication between the numbers of samples with correlation coefficients. We will obtain the sample average by dividing the total sample (528) with the observed number of studies, the 10 studies. Thus we will obtain a sample mean of 52.8. Average correlation coefficient can be calculated by dividing the total of column 2 column 3 with the total in this table, so that we obtain the average coefficient of 0.280989.

By using the average correlation coefficient and the average number of samples we can do further calculations to get the figures chi-square. Table 3 below gives illustration the chi-square calculation. In the table 3 we obtain by 12.118737. With 9 degrees of freedom or the number of studies (K) minus 1 we obtain evidence that the number of chi-square was not significant (less than the chi-square table for $\propto = 0{,}05$). In this case the null hypothesis which states that there are no differences in the study sample of 10 accepted. This means that the difference in correlation coefficients item-total of 10 studies are used as an illustration of this calculation was not the result of differences in sample. The correlation coefficient difference may come from other artifacts outside the artifact sample differences.

TABLE III. KAI CALCULATION SQUARES CORRELATION COEFFICIENT DIFFERENCES

r	N-1	$\left(r-\bar{r}\right)$	$\left(r-\bar{r}\right)^2$	$(N-1)\left(r-\bar{r}\right)^2$	$(N-1)\frac{(N-1)(r-\bar{r})}{(1-\bar{r})^2}$
0,3810	64	0,100011	0,010002	0,640136	0,890301
0,2493	81	-0,031690	0,001004	0,081342	0,113130
0,2499	70	-0,031090	0,000967	0,067659	0,094099
0,3880	36	0,107011	0,011451	0,412246	0,573351
0,4240	40	0,143011	0,020452	0,818081	1,137788
0,2807	69	-0,000290	8,37E-08	5,78E-06	0,000008
0,1102	35	-0,170790	0,029169	1,020916	1,419889
0,0846	41	-0,196390	0,038569	1,581321	2,199301
0,5310	42	0,250011	0,062505	2,625223	3,651160
0,0895	40	-0,191490	0,036668	1,466728	2,039925
				X^2	12,118731
				db	9
				p	> 0,05

So far the researchers have difficulty to make decisions about the validity of the test. Especially if researchers want to test the validity of a way to correlation between item with a total score of the scale, the result of adaptation or scale developed by other researchers. Decision-making is meant here is decided to use a compiler item validity testing or re-validate the scale. Not a few researchers who conduct retesting and item decision to use the test results. There seems to be a tendency to overlook the hard work making up the scale to test the validity item blueprint drafted by carefully designed. If users have an abortion item scale has been standardized, meaning that little or a lot of researchers have sacrificed the principle of equality. Inequalities in terms of scale certainly have a contribution to the research conclusions inequality. The absence of information on the procedures of decision making inferences about differences in outcomes they take a very brave way.

Testing measures the correlation coefficient differences offered in this paper is intended as an alternative to making decisions that have been made by researchers in arbitrary. Hunter-Schmidt technique for testing differences in this sample should be used to conduct an assessment of studies on a given topic in the social sciences. Very least the results of a study assessing the same topic make this technique less attention and less popular. Although the main purpose of the test sample error is between one study with other studies. Proving fault samples have high value practicality. Mainly is practicality in terms of testing different power item.

Measures offered relatively simple and easy to understand. Calculations can be easily done by using spreadsheet software. Utilization of sample testing techniques Hunter-Schmidt will be more extensive if the calculations can be integrated into existing software, example package SPSS. With the integration

testing of the sample differences into a program package will facilitate researchers to take a decision. Decision-making with regard to whether or not abortion items certain of the scale developed by other researchers. Despite the controversy regarding whether or not a person to conduct a validation test of the scale re-arranged another researcher, this technique at least provide an alternative to facilitate the decision making process.

REFERENCES

[1] Azwar, S.1987 Tes Prestasi, Yogyakarta ; Liberty

[2] Azwar, S. 1994. Seleksi Aitem dalam Penyusunan Skala Psikologi, vol 2, no.2.pp.27-34

[3] Bangert-Drowns, R.L 1986. Review of Developments in Meta-Analytic Methods Psychological Bulletin, vol.99, pp. 388-399.

[4] Debby, K. 2003 .Hubungan antara Tingkat Kecemasan dan Umur dengan Emesis Gravidarum pada kasus Primigravida di Poliklinik Pemeriksaan Kehamilan Ante Natal RS Bethesda Tomohon. Skripsi: Fakultas Psikologi Universitas Kristen Tomohon di Manado

[5] Hedges, L.V. and Olkin. I, 1985 Statistical Methods FOR Meta-Analysis London Academic Press, Inc.

[6] Hunter, J.E. and Schmidt, F.L. 1990. Methods of Meta-Analysis ; Correcting Error and Bias in Research Findings. Newbury Park; Sage Publications.

[7] I Made, S .2001, Studi Tentang perbedaan Kecemasan antara Anak yang Berbeda Status dalam Keluarga pada Siswa SMP Kristen Kaaten Tomohon dan SMP Kristen Pniel Bahu Manado. Skripsi : Fakultas Psikologi Universitas Kristen Tomohon di Manado.

[8] Jeiner. R. 2001. Hubungan antara Jumlah anak dalam keluarga. Pola Asuh Demokratik, dan Status pekerjaan Ibu dengan Kemandirian Pada siswa kelas I SMA Negeri di Kota Manado. Skripsi: Fakultas Psikologi Universitas Kristen Tomohon di Manado.

[9] Jeiner.R. 2007 Aplikasi Meta-Analisis untuk Analisis Aitem Skala Pola Asuh Demokratik, Laporan penelitian, Fakultas Psikologi Universitas Kristen Tomohon di Manado.

[10] Johnson, B.T. Mullen, B. And Salas,E 1995 Comparison Of Three Major Meta-Analytic Apprpoaches. Journal of Applied psychology vol 80 no.1. pp.94-106.

[11] Osburn, H.G and Callender, J.C 1990. Bias in validity Generalizatio Variances Estimates: A Reply to Hoben Thomas. Journal of Applied Psychology, vol 75. no.3. pp. 328-333.

[12] Rusmin. M. 2003. Studi Pendahuluan tentang Hubungan antara Kecemasan Dengan lama Persalinan di RS.Teling Manado Skripsi: Fakultas Psikologi Universitas Kristen Tomohon di Manado