The Misuses of Reliability Coefficient and Sampling Variance in Educational Research

Dali S. Naga

Abstract: Many educational researchers report the reliability of their data. They also report the sampling techniques used in their research. Some of the reports treat the reliability coefficients and sampling variances incorrectly. It is a misuse of the reliability coefficients and sampling variances in educational research.

Key words: Reliability coefficient, correlation coefficient, sampling variances.

Many educational research reports, either in research institutions or in graduate and postgraduate studies, consist of two major parts. The deductive part elaborates on the theoretical aspect of the research while the inductive part engages in experimenting and collecting data through research samples. The collection of data is usually done through custom-made instruments in which the validities and reliabilities are checked through several efforts of tryouts. The research samples are usually acquired through several kind of sampling techniques.

The data collected during the research is usually probabilistic in nature that the data analysis is computed through a series of statistical procedures. It is by this means that the testing of hypotheses are endorsed and the results of the research are reported. But it is also in this part

Dali S. Naga is a professor at the Faculty of Technology and Vocational Education, IKIP Jakarta.

of the research that some misuses of the statistical procedures appear in the research reports.

This article focuses on two such misuses. The first one is the reliability coefficient of the data obtained from the measuring instrument and the second one is the sampling variance due to the selection of the kind of sampling technique. There is a tendency that these misuses are spread wider and wider in the circle of educational research because later researchers ignorantly copy these misuses from the earlier researchers. It seems that it is now the time to check that rampant propagation of those misuses among the younger researchers.

COEFFICIENT OF RELIABILITY

There are many formulas for the coefficient of reliabilities of scores. Some of them bear the names of their inventors such as Spearman and Brown, Kuder and Richardson, Cronbach, Flanagan, Guttman, Mossier, Rulon, Feldt, Kristof, and more (Naga, 1992: 134—155). Kerlinger even suggested that the reliability of scores could also be analyzed through analysis of variance between items and respondents (Kerlinger, 1975: 447—451). But out of these famous names the most well known coefficients among educational researchers are Spearman-Brown coefficient of reliability on the one hand and Kuder- Richardson 20 and Cronbach Alpha on the other hand.

One common way of examining score reliability is through variances. In this case, it is usually considered that the sum of the item p-scale variances p (1 - p) as errors and is deducted from the total variances. The remaining variance is regarded as a measure of score reliability. In politomous model this p-scale variance is substituted by the internal variance of each item. Another common way of computing score reliability is through correlation coefficient between test score and its retest score or its equivalent test score.

Basically all of the coefficients of reliability can be traced back to correlation coefficients which is restricted within the same test score. It is a sort of internal coefficients of correlation which correlate a part of the test score with another equivalent part of the same test score. A high correlation coefficient between equivalent parts of the same test scores indicates that this conformity should lead to high reliability.

In some educational research reports, test score quality is reported beyond the values of the coefficients of reliability. Assuming that it is equivalent to the coefficient of correlation, the coefficient of reliability in the reports is further inferred to its population through statistical t-test with a certain degree of significance. Particularly this statistical testing is accomplished through sampling distribution of the coefficient of product moment correlation. The problem now is: can this statistical inference be justified?

For the sake of simplicity, some abbreviations are used in this article. Coefficient or reliability is noted as CR. Pearson product moment coefficient of correlation is further noted as CC. Test-retest is noted as TR and parallel tests as PT. Spearman-Brown is noted as SB, Kuder-Richardson as KR, and Cronbach Alpha as CA.

In varying frequencies, the mostly used CRs are TR CR, PR CR, Split-half SB CR, and KR-20 or CA CRs. Imitating hypothesis testing for CC in the research hypothesis, the researchers tested the CRs as if they are the CC in their studies. Let us see how justifiable is the use of the t-test statistical inference in these CRs.

We begin with SB CR. In computing SB CR, the test is split into two equivalent halves. The respondent's scores of those two halves are correlated and produce CC. This CC is then put into a formula to produce SB CR. If this SB CR is then tested statistically using Student-t distribution then it is assumed that, firstly, the SB CR is a sample and, secondly, the sampling distribution of SB CR is a Student-t distribution. Judging from SB CR formula, it is very unlikely that its sampling distribution is a Student-t distribution. Hence this assumption and this statistical inference of CR is incorrect.

In a sense, KR-20 CR and CA CR are extensions of SB CR formula by expanding the two halves into n parts and treating test items as parts. It is also very unlikely to take KR-20 CR and CA CR as samples and it is equally unlikely that their sampling distributions are Student-t distribution. Therefore this statistical inference of CR is also incorrect.

It is true that TR CR and PT CR are CC in nature. But in the field of reliability, we are not looking for the existence of CC. We do not concern about the problem of whether the CC is equal to zero or not. What we actually concern is how high the value of CC is. We are eager to have a CC of around 0.70 or higher. With the usual size of sample

there is no sense of statistically testing a CC (here it means CR) which already has a value of around 0.70 or higher. It is merely a wasting of the research time in the part of the researchers.

SAMPLING VARIANCE

Educational research is conducted through samples and, using statistical techniques, the population scores are decided through the use of those sample scores. Since there are various kinds of sampling techniques, the researchers can select which sampling technique is used in the research. In many cases, the nature of the populations require the right kind of sampling technique in the research.

There are two major kinds of populations, the finite population and the infinite population. Within them, the populations can be concentrated in one area or spread in a very wide area. The populations can be traced into several strata or are unable to be separated into strata. Each kind of population might have a proper kind of sampling technique of its own.

Several sampling techniques are well-known among educational researchers. They are simple random sampling, stratified random sampling, cluster random sampling, systematic sampling, quota sampling, purposive sampling, multistage sampling, et cetera. Due to the nature of the sampling, each sampling technique has its own formulas of variances and sampling variances (Parel, 1973). The simplest formula is that of the simple random sampling and the most complex one is that of multistage sampling.

Research reports usually refer various kind of sampling technique used in the research. Some of them mention stratified random sampling or cluster random sampling together with the reason of choosing such kind of sampling technique. Most of the selections of the sampling techniques are justifiable. But when we look into the statistical part of the report, irrespective of the kind of sampling technique used in the research, almost all of the them use sampling variance of that of the simple random sampling.

For those researches which use simple random sampling technique, this sampling variance is of course correct. But for those which use stratified or cluster random sampling, this sampling variance is a misuse.

This misuse of sampling variance in the statistical computation of the research reports appears widely in the educational research.

CONCLUSION

Many educational researchers copy the procedures of reporting the reliability coefficients and the sampling variances from the previous research reports. There are now cases of misuses of reliability coefficients and sampling variances in the previous research reports. These misuses breed further and wider misuses in the research reports in the later research.

These misuses have to be stopped. Later educational researchers have to pay more attention to their research reports by avoiding all misuses inherent in the previous research reports especially those in the fields of reliability coefficients and sampling variances.

REFERENCES

Kerlinger, F.N. 1975. *Foundations of Behavioral Research*. Second Edition. London: Holt, Rinehart and Winston.

Naga, D.S. 1992. *Pengantar Teori Sekor dalam Pengukuran Pendidikan*. Jakarta: Penerbit Gunadarma.

Parel, C.P., Caldito, G.C., Ferrer, P.L., De Guzman G.G., Sinsioco C.S., and Tan, R.H. 1973. *Sampling Design and Procedures*. Quezon City: PSSC Social Survey Series 1.

Shavelson, R.J. 1996. *Statistical Reasoning for the Behavioral Sciences*. Third Edition. Boston: Allyn and Bacon.