Critical thinking abilities assessment tools: reliability generalization

Natcha Mahapoonyanont*, Rewadee Krahomwong, Duenpen Kochakornjarupong, Worawanninee Rachasong

*Department of Evaluation and Research, Faculty of Education, Thaksin University, Songkla Campus, 90000, Thailand
Department of Mathematics, Faculty of Science, Thaksin University, Phatthalung Campus, 90000, Thailand
Department of Psychology, Faculty of Education, Thaksin University, Songkla Campus, 90000, Thailand

Received October 6, 2009; revised December 15, 2009; accepted January 4, 2010

Abstract

The purpose of this research was to synthesize the research related critical thinking abilities and to examine the potential impact of variables on the score reliability of critical thinking abilities measurement tools from articles that published on ThaiLIS database during 1988-2008. There were 108 studies that related keyword “critical thinking”. There were only 59 articles that critical thinking abilities were dependent variables and were reported the measure development processes. Reliability generalization (RG) was used to examine the potential impact of variables on the score reliability. Besides the most researchers used conceptual of Robert H. Ennis to develop CTA measuring instruments. In addition the research finding described as the potential impact of variables on those score reliability.

© 2010 Elsevier Ltd. All rights reserved.

Keywords: Critical thinking abilities; assessment tools; reliability generalization; researches synthesize; meta analysis

1. Introduction

The concern of teaching thinking skills is penetrating the education system everywhere in the world. All levels of society agree that thinking skills are crucial for one to remain relevant and proficient in this fast-paced and competitive world. In the era of massive information and technology explosion as such, there is an urgent need for teaching critical thinking to students. So thinking and thinking teaching is an important issue in an educational management which has been emphasized the characteristics of people to have the ability in circumspect, reasonable, and serious issue for develop country nowadays. It is the result of The Eighth National Economic and Social Development Plan that underscored people to be the center or a country development. Consequently, thinking or critical thinking or aware of thinking is a necessary purpose of an education because it will help learners be able to solve problems efficiently and it is a tool for continuous self-studying. Learners need to have the ability in careful thinking for separating to select useful information and use benefits from them efficiently. Current school’s duties; therefore, are not only teaching content in subjects but also teaching and supporting students to be able to think.
proficiently or discreetly. Moreover, they should have the habit in thinking properly in order to help them live in the society happily and can appropriately do a favor of country development.

So a development of critical thinking is significant for learners. This kind of abilities, accordingly, has more role in schools since the development of learners thinking’s ability could be done in limitation and it didn’t achieve the required aims in the previous decade. Thailand has had a movement about thinking development for several years; nevertheless, it hasn’t widely used and an advanced thinking is not as good as it should be. Thai students achievement, according to the Institute for the Promotion of teaching Science and Technology, indicates that their average score in quite low level.

Thai educational policy; furthermore, must be accentuated enhancing the process of Thai students’ thinking, seeing that a quality of their studying at the present is worrying, the average abilities of Thai student are decreasing including an analysis, a sensible synthesis, as well as the creativity.

From all above, it obviously shows that thinking is very vital in human development, especially in the education circle, the concept is a way to improve people to become aware of thinking is developing them to have a skill of critical thinking.

In conclusion, therefore, in an educational reform considers that developing a wisdom, improving people to become aware of thinking and having a critical thinking as well as having the ability to judge things properly are essential. As a consequence, every level of the education department must realize and find solutions with reasons above.

A following key point is that a method to identify the development above needs to have procedures and suitable measuring instruments. From a survey of researches regarding critical thinking abilities published in 1978-1997, from ThaiLIS database, there were 106 studies, which have been trying to study the variables involving this type of thinking. The researchers thought that the researches should be synthesized the measuring instruments of critical thinking abilities and considered if there were any variables and factors affected the errors of measurement of the instruments by using Reliability Generalization technique (RG) (Vacha-Hasse.1998). In this meta analysis will benefit in sponsorship a development of critical thinking abilities and it can bring information to use in the development correctly, efficiently, and effectively in an education’s advance of the country. Because the results of a RG study will provide researchers with a better understanding of the reliability of scores obtained for tests in their particular studies as well as test characteristics that contribute most to score reliability in future studies.

3. Instruments and Methods

The purpose of this study was to synthesize independent researches, theses, as well as dissertations which had been published in 1978-2008 and broadcasted on ThaiLIS database by using “critical thinking” as a keyword. It was found that there were 106 studies on the database, but only 88 studies that had information corresponding to the researchers’ requirement (critical thinking abilities was the one of variables). The collection of this study use a research characteristic recording form, quality research evaluation form, then brought the information to analyze and synthesize by descriptive statistic, content analysis as well as reliability generalization technique (RG) (Vacha-Hasse.1998) had been used in this meta analysis.

The variables that researcher interested in this study were reliability as a dependent variable, science subject, secondary level, sample size, N of try out, N of item, and multiple choice scale were the independent variables.

The coefficient was transformed in to it’s Z statistic using Fisher’s r-to-Z transformation and weight by sample size. This transformation from r to -Z is performed because correlations come from a skewed distribution. Sample size typically is important in the stability of a statistical estimate. For this reason, sample size is considered relevant in meta analysis and a weighting is applied to the individual estimate with more weight given to an effect size measure from a larger sample. A mean weighted Z value was calculated and 95% confidence intervals were constructed. The Z-to–r conversion was then performed and a mean weight r (or coefficient alpha) was reported as well as 95% confidence intervals. The Z-to–r conversion is computed to provide ease when interpreting the results of the meta analysis.
4. The Researches Result

The beginning result was found that most of the studies were published in 1997. It was mostly KhonKhaen University’s studies (22.50%); in addition, they were theses (86.80%), as well as the theses in Curriculum and Instruction course (21.10%). Most of the studies were done in Science content (21.00%), and 50.50% of the sampling are studied with secondary school students. Only 59 studies reported the procedure of instrument development and reported reliability scores only 75%, according to 31.80% of those had high confidence level (α between .60-.79). The number of sampling to use in examine the instruments had the variation between 7-860 people and the variation of the number of items were between 5-80 items.

Besides, there were only 25 studies that used Robert H. Ennis’s critical thinking concept to measure critical thinking abilities; as a result, researcher assumed that all of these researches were the same construct and use to measure as the same things so that these 25 documents were used by RG technique. But unfortunately there were only 14 of 25 studies that reported the reliability score and appropriate to analyze by using RG technique.

According to the study of RG outcome, we found that when those alpha scores were adjusted (T_i) by using Fisher’s Z Transformation, the result was shown in table 1, as well as confidence interval of reliability scores 95% (represented by the error bars) to estimate an average score of Alpha Coefficient that was weighed by an amount of samplings was shown in table 2. In table 3, likewise, shows correlation coefficient among variables we wanted to study: alpha scores and 5 independent variables, namely, secondary school students, science content, sample size, numbers of try out the instruments and numbers of items.

From the outcome of correlation analysis, it was found that sample size and the number of items variable had high relationship (r xy = .902**).

Table 1. confidence interval of reliability scores

| Estimate type         | 95% confidence interval |         |         |
|-----------------------|-------------------------|---------|---------|
|                       | Lower limit M Upper limit |         |         |
| Alpha unweighted      | .71 .83 .96             |         |         |
| Alpha adjusted        | .34 .53 .66             |         |         |
| T. = .494882          |                         |         |         |
| \( \rho_{\alpha} = [1 - T.]/T. \) = .8788 | |         |         |

From table 1 found that the average of alpha unweighted was .83, the average of alpha adjusted was .53, the average of alpha transformed was .495, and the average when transformed back to raw score was .879.

Table 2. Associates between variables (Pearson and Spearman Correlation Method)

| Estimate type        | 95% confidence interval |         |         |         |         |         |         |
|----------------------|-------------------------|---------|---------|---------|---------|---------|---------|
|                      | 1           2           3           4           5           6           7           |
| 1. Alpha coefficient | -           -3.56       -2.59       - .081       -.294       -.078       -.681       |
| 2. Science subject  | -           - .026       .262        .163        .026        .141        .055        |
| 3. Secondary level  | -           - .072       .152        .059        -.059       .141        .440        |
| 4. Sample size      | -           - .027       .902**      -.504       -.059       -.463       -           |
| 5. N of try out     | -           -           -           -           -           -           -           |
| 6. N of items       | -           -           -           -           -           -           -           |
| 7. Multiple choice scale |                  |         |         |         |         |         |         |

**Correlation is significant at the 0.01 level (2-tailed) *Correlation is significant at the 0.05 level (2-tailed).

According to table 2, it can be concluded that there is a weak positive association between alpha coefficient and independent variables. Namely, sample size (r xy = .081) and N of items (r xy = .078) and there is a negative association between alpha coefficient and secondary level (r xy = -.259), science subject (r xy = -.356), N of try out (r xy = -.294), and Multiple choice scale (r xy = -.681). The most positive association of independent variables and alpha coefficient was sample size (r xy = .081) and the most negative association of independent variables and alpha coefficient was Multiple choice scale (r xy = -.681). In addition, there was a strength positive and significant association among independent variables namely, sample size and N of items (r xy = .902**).
Figure 1. The confidence interval of reliability scores (95%) (Represented by the error bars)

Table 3. Weighted Least Squares Regression Results of Alpha adjusted

| Estimate type          | df  | Sum of Squares | Explained |
|------------------------|-----|----------------|-----------|
| Regression             | 6   | 1435.294       | 41.90%    |
| Residual               | 7   | 636.876        |           |
| Total                  | 13  | 2072.170       |           |

| Variables              | B   | SE  | Beta | t    | P   |
|------------------------|-----|-----|------|-----|-----|
| Constant               | 28.925 | 4.810 |  6.013 | .000 |
| Multiple choice scale  | -18.345 | 5.692 | -.681 | -3.223 | .007 |

In table 3 was shown formula hence;
Raw score;  
Alpha coefficient = 28.925-18.345 Multiple choice scale
Standardized score;  
Alpha coefficient = -.681 Multiple choice scale

5. Conclusion

All of the independent variables there was only 1 variable; Multiple choice scale that subjected to regression analysis explain 41.90% ($R^2_{adj} = .419$) of the total variance in the reliability score.

6. Limitation of this study

The limitation of this study was had rather little sampling, that is to say, there were only 14 alpha scores that used in this study. From the literature review of RG studies, the traditional of Reliability Generalization technique was often used appropriately with a same form of assessment but it had been brought to measure or study for many times; apart from, there was a limitation in the clearness to transform or weigh several score. It was found that this weighting just had a few effects to a variance of the alpha scores; therefore, it had a few researches using this technique to build up knowledge. In other words, there were only 12 studies had been published (Rodriguez, M. C. and Maeda, Y. 2006) since reliability generalization technique (RG) was officially presented by Vacha-Haase in 1998. Another limitation was many of studies didn’t report alpha coefficient from a development of instruments and several necessary statistics for a meta analysis were in moderately high level. Besides, many international journals determined a main condition that there would be no publishing of study which didn’t report of their measuring instruments statistics.

In order to expand a result and build knowledge about RG up, large numbers of researchers had tried to study for an improvement of mentioned knowledge from the contents they had; still, there were rather little sampling in the researches. For example, Henson Robin K. (2002) had only 10 studies in the study and Zangro, G. A. and Soeken, K. L. (2205) had 14 studies of sampling; as well as, some of them had piloted this studied so that it would have wide studied about it by studying RG during varied researches that had different measuring instruments but measured in
the same construct or the same thing. According to a surfing online on the data base, in contrast, it had no an official and obvious study in Thailand. We hope that this research will advantage and could be the beginning of building mentioned knowledge which can bring benefits to an educational development of the country in the future.

Acknowledgement

This study has been fund by TRF (Thailand Research Fund) and Office of the Higher Education Commission as TRF grant no. MRG5180351 (fiscal year 2009 - 2010).

References

Henson, R. K., & Thompson, B. (2001, April). “Characterizing measurement error in test scores across studies: A tutorial on conducting “reliability generalization” analyses,” Paper presented at the annual meeting of the American Educational Research Association. Seattle, WA.

Caruso, J. C. (2000). “Reliability generalization of the NEO personality scales,” Educational and Psychological Measurement. 60, 236–254.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. (1st ed.). Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (2000). “Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge,” International Journal of Selection and Assessment. 8, 275–292.

Hunter, J. E., & Schmidt, F. L. (2004). Methods of meta-analysis: Correcting error and bias in research findings. (2nd Ed.). Thousand Oaks, CA: Sage.

Michael C. Rodriguez and Yukiko Maeda. (2006). "Meta-Analysis of Coefficient Alpha," Psychological Methods. Vol. 11, No. 3, 306–322.

Thompson, B., & Vacha-Haase, T. (2000). “Psychometrics is data metrics: The test is not reliable,” Educational and Psychological Measurement. 60, 174–195.

Vacha-Haase, T. (1998). “Reliability generalization: Exploring variance in measurement error affecting score reliability across studies,” Educational and Psychological Measurement. 58, 6–20.

Zangaro, George A, Soeken, (2005). "Meta-Analysis of the Reliability and Validity of Part B of the Index of Work Satisfaction Across Studies," Journal of Nursing Measurement. HighBeam Research. (May 1, 2009). http://www.highbeam.com/doc/1P3-931222531.html