

The Development of Alternative Test Innovation for Learning Assessment

Aungkana Koolnapadol¹, Nuengruthai Mekwathat²
and Narin Kooknapadol³

Received October 17, 2020; Revised October 27, 2020; Accepted December 9, 2020

Abstract

This research has developed a concept map as a tool for learning evaluation intended to analyze the size of the variance in each element, compare the coefficient of reference summary, compare the exactness when the pattern and number of examiners are different while developing training courses to apply concept mapping techniques for learning assessment and evaluation of training results. The sample in this study is divided into 2 groups: 4 examiners, and 48 students from Rajabhat Rajanagarindra University. The research tool used is an activity plan for the use of concept maps for learning outcome assessment, concept mapping evaluation form, quizzes, training assessment. The research results indicate that;

When the examination pattern is different even though the number of inspectors is the same, the Generalizability Coefficient then is statistically different with a significant level of 0.05. When the examination pattern is the same but the number of assessor is different, then the Generalizability Coefficient is statistically different at a significant level of 0.05. The score in all conditions has a high Concurrent Validity and statistically significant difference at 0.05. Training courses in applying the concept mapping techniques for learning measurement and evaluation are the most appropriate for all items. Evaluation results of training courses in applying concept mapping for measuring and evaluating learning outcomes are at a high level for all items except for the knowledge on the topic prior to training which is at a moderate level.

Keywords: concept mapping, generalizability coefficient, evaluation of a learning achievement

Introduction

The measurement and evaluation of the study is a psychological assessment where items to be measured are quite abstract, and there are extraneous variables involved. The operator therefore may not have full control over those variables. As a result, there is a potential error that may occur at any time. Therefore, for each test, the obtained score will be the sum of two points: the true score and the error score. The tolerance score can be either positive or negative. And the discrepancy that occurs may be due to many reasons, for example, poor quality measurement tool, operator lacking expertise, the variation of the examinees, discrepancy in sampling the content, behaviors, etc.

Due to the aforementioned reasons, it can be

seen that the tools used in the measurement are important parameters that the assessor must pay attention to and care for. Ruiz-Primo, Schultz, and Shavelson (1996) sought to find out how to measure and evaluate results in science. They wanted to investigate what the students know and what they can do. In this study, they used the concept mapping to measure and evaluate the results. West, Pomeroy, Park, Gerstenberger, and Sandoval (2000) studied the application of concept mapping to assess the critical thinking of medical students. And Srinivasan, McElvany, Shay, Shavelson, and West (2008) studied the implementation of conceptual mapping in measuring and evaluating medical students. This can be seen that in foreign countries, there are people interested in applying conceptual maps to

¹ Assistant Professor, Faculty of Education, Rajabhat Rajanagarindra University

² Lecturer, Faculty of Education, Rajabhat Rajanagarindra University

³ Assistant Professor, Industrial technology, Rajabhat Rajanagarindra University

measure and evaluate learners' learning outcomes or even to assess critical thinking, which is an advanced thinking. For Thailand, the use of concept mapping to measure and evaluate still receives very little attention from people interested.

Concept mapping is not a newly emerging technique or method. It was developed and has been used in education for over 30 years (Buldu & Buldu, 2010). Novak & Gowin (1972) pioneered and developed this concept based on Ausubel's Theory of Meaningful Learning (1968). Ausubel believes that learning is a result of assimilation, knowledge, and stories that have been newly acquired combined with the existing concept the learners already possess and it is the expansion of knowledge by the learners themselves. Ausubel also pointed out that teaching methods to achieve meaningful learning depend on three conditions as follows:

- 1) New knowledge must contain logical meaning as resulted from the learners' prior knowledge.

- 2) The structure of the prior knowledge must be related to the new knowledge.

- 3) The learner must understand and have the determination to learn meaningfully, otherwise, the first two conditions will result in memorization learning. Novak visualized Ausubel's ideas into concrete charts and continuously pioneered the research since 1972.

Measurement and evaluation using concept mapping is an important and challenging process since there are several sources of variability. This is consistent with the concept of McClure, Sonak, and Suen (1999) who studied using concept mapping as a tool and found that the quality of the concept mapping may have multiple sources of variation, such as scoring methods, discrepancies from students' knowledge, and discrepancies from the assessor, etc. Such of these discrepancies affect the efficiency of the concept mapping. Ruiz-Primo and Shavelson (1996) stated that analyzing the quality of concept mapping by using traditional test theories such as retesting is not suitable for determining the quality of mind mapping. Instead, the importance of the assessor should be considered. This is consistent with Chaiyapruet Serirak et. al (2540) who stated that in the analysis of the validity of continuous measurement

for student's performance, the traditional validity theory like splitting the Cronbach's alpha coefficient test into two halves would not be suitable. Because these theories have an agreement that the test must measure the same thing in parallel and must perform the exam once. But the assessment using concept mapping is of different characteristics. Thus, the theory of validity approximation or the so-called Generalizability Theory is, therefore, a suitable theory to analyze the validity of the assessment using concept mapping.

From the above document and research, the application of concept mapping for measurement and evaluation in Thailand receives a little attention. But in contrast, in foreign countries, they are widely studied. In 2013, in the educational research course, the researcher developed the concept map scoring criteria, analyzed the size of the variance in each of the scoring components, compared the Generalizability Coefficient of the scoring criteria when the scoring criteria and the number of assessors are different. In the study, the researcher found the problem that during the examination, students periodically raised their hands in question indicating a lack of understanding in the implementation of concept mapping for measurement and evaluation. Therefore, in this study, the researchers develop a concept mapping for the course of the Principles of Education Measurement and Evaluation in the topic of the Instrument for measuring learning behavior and analyzing the size of variance in each component of the concept mapping, compare the Generalizability Coefficient of the concept mapping in case the pattern of examination and number of assessor are different and when the Concurrent Validity of the concept mappingscores are already compared. And the researcher developed a training course for the application of mind mapping techniques for measurement and evaluation and evaluated the training course. The researchers wanted to disseminate the knowledge gained from this research to teachers/assessors in the educational area of Chachoengsao Province in order to learn the alternative assessment pattern for use in the actual scenario and to bring the results of the assessment to effectively improve and develop the learners.

Research Objectives

1. To compare the Generalizability Coefficient of the concept mapping when the pattern of examination and the number of examiners are different
2. To compare the validity of the concept mapping scores when the pattern of examination and the number of examiners are different
3. To develop a training course on applying concept mapping techniques for learning measurement and evaluation.
4. To evaluate the training course on applying concept mapping techniques to measure and evaluate learning achievement.

Scope of the Study

For responding to the objectives 1 and 2

1. Target groups include 4 teachers in the teaching professional group and teacher students enrolled for a course of the Principles of Education Measurement and Evaluation in the first semester of the academic year 2018 of 48 students from Rajabhat Rajanagarindra University

2. Research variables

2.1 There are 2 independent variables, namely;

2.1.1 Three patterns of scoring including:

2.1.1.1 The assessor inspects every concept map of every student.

2.1.1.2 The assessor inspects every concept map of some students.

2.1.1.3 The assessor inspects some concept maps of all students.

2.1.2 The number of inspectors is classified into 3 levels, namely;

2.1.2.1 2 persons

2.1.2.2 3 persons

2.1.2.1 4 persons

2.2 There are 2 dependent variables, namely

2.2.1 Generalizability Coefficient

2.2.2 Concurrent Validity

For responding to objectives 3 and 4

1. Target groups include 48 teacher students who enrolled for a course of the Principles of Education Measurement and Evaluation in the first semester of the academic year 2018 from Rajabhat Rajanagarindra University

2. Research Variables

2.1 There are 2 independent variables, namely;

2.1.1 Training courses on applying the concept mapping techniques to measure and evaluate learning achievement.

2.1.2 Students

2.2 There are 2 dependent variables, namely

2.2.1 Course Quality

2.2.2 Satisfaction of the training participants

3. The period of study is the first semester of the academic year 2018.

4. Content used in the development of this concept mapping is the content of the course of the Principles of Education Measurement and Evaluation.

Research hypothesis

1. The Generalizability Coefficient of concept mapping with the examination pattern where the assessors inspect every concept map of all students will provide higher value than in other patterns of all assessors.

2. The Generalizability Coefficient of concept mapping with a higher number of assessors will provide a higher value of the Generalizability Coefficient than those with a smaller number of assessors.

3. The Concurrent Validity of concept mapping with the examination pattern that the assessors inspect every concept mapping of all students will provide higher value than those of other patterns in every number of assessor.

4. The Concurrent Validity of concept mapping with the higher number of assessors will be higher than those with less number of assessors in all patterns of examination.

Research Methodology

The research tools consisted of activity plans in the implementation of concept mapping for learning evaluation. The concept map evaluation form for assessing learning achievement in 12 topics. The researcher collected data by themselves in the Semester 1/2018 in the course Principles of Education Measurement and Evaluation from 48 persons of the sample group. Researchers organized teaching and learning activities each time and allow students to complete the concept map assessment form in 12 topics. After that, all concept map forms are photocopied into 4 copies, given to 4 assessors with the researchers explained the examination methods and scoring criteria created by the researchers and have the assessors to inspect every student's concept map along with taking notes and then bringing the results of the examination to prepare and analyze the information as follows

1. Compare the Generalizability Coefficient of concept mapping when the pattern and number of assessors are different using Woodruff and Feldt's UX1 formula (1986, p. 393-413) (for responding to purpose 1).

2. Compare the Concurrent Validity of the concept mapping score when the pattern and number of assessors are different using the Pearson Product Moment Correlation coefficient analysis for scores obtained from different test patterns and different numbers of assessor with the final exam scores on the course of the Principles of Measurement and Evaluation of education. And compare the Concurrent Validity of the scores when the pattern and number of assessors are different. The correlation coefficient will be converted into Fisher's standard Z score and then tested for the parallel difference of the correlation coefficient by using Chi-Square statistics (for responding to objectives 2)

3. Consider the suitability of the training plan by using mean and standard deviation values (for responding to the objective 3)

4. Evaluate the results of training on the course of the application of concept mapping for measuring and evaluating learning results by calculating the mean and standard deviation values (for answering to objec-

tives 4)

Research Results

Comparing the Generalizability Coefficient of mind map scores when the pattern and number of assessor are different and testing for the difference of the Generalizability Coefficient when the examination pattern is different but the number of assessors is the same, the results reveal that all conditions are statistically significantly different at a level of 0.05 indicating that at least 1 pair of Generalizability Coefficient for each condition is different. When testing the parallel difference of the Generalizability Coefficient where the test pattern is different but the numbers of assessors are the same, it showed a statistically significant difference at 0.05 level. And when considering the assumptions set out in Clause 1 that "the Generalizability Coefficient of the mind map score with the examination pattern where the assessors inspect of every mind map of all students would have a higher value than other scoring patterns in all numbers of assessor". It can be seen that this is consistent with all the assumptions defined. When the pattern is the same but the number of assessors is different, it is found that all conditions have a statistically significant difference at a level of 0.05, indicating that at least 1 pair of the Generalizability Coefficient for each condition was different. When testing the parallel difference of the Generalizability Coefficient values when the test pattern is the same but the number of assessors is different, it is found a statistically significant difference at the 0.05 level for all pairs. And when considering the assumptions set out in Clause 2 that "the Generalizability Coefficient of the concept mapping score with a higher number of assessors, will provide the higher value than the Generalizability Coefficient with a lesser number of assessors". This can be seen that all hypotheses are true.

Comparing the Generalizability Coefficient of the concept map score when the pattern and number of assessor are different, the results in comparing the Generalizability Coefficient of the concept map scores when the pattern is different but the number of assessors is the same, it shows statistically significant differ-

ence at the 0.05 level following the hypothesis 3 stating that “the Generalizability Coefficient in the concept map scores in the course of the Academic Research with the examination pattern where the assessors inspect every concept map of all students will have a higher value than in other patterns in every number of assessors”. This indicates that the scores obtained from the different examination patterns but the same numbers of assessors will have different values of Generalizability Coefficient. The results in comparing the Generalizability Coefficient with the same the test patterns but the different numbers of the assessor, it shows statistically significant difference at the 0.05 level following the hypothesis 4 which states that “Generalizability Coefficient of concept map score in the course of the Academic Research in education with a higher number of assessors will provide higher value than those with lesser number of assessors in all patterns of examination”. This indicates that the scores obtained from the same examination pattern but different numbers of assessors will have different values of Generalizability Coefficient.

The results in developing the training course in applying the concept map techniques for measuring and evaluating learning outcomes (responding to Objective 3), all 3 experts considered the suitability of the curriculum, objectives, procedures for organizing activities and documents related to organizing the activities, the application of concept mapping for assessing the learning outcomes on the basic knowledge of mind map and the utilization of mind map for learning evaluation, and found that they are appropriate in the highest level indicating that the curriculum developed is appropriate.

Results in evaluating the training course in applying concept mapping for measuring and evaluating learning outcomes (responding to objective 4), the result found that among 48 respondents, 39 are female students, accounting for 81.25 percent, and 9 males, accounting for 18.75 percent. All respondents agree that the time to attend the training is reasonable, accounted for 100.00 percent. Most of the respondents have a high level of overall opinion about the suitability of the course with a mean value of 4.11. When sorting the mean value in descending, it is found that the content

is useful at a high level with the mean value of 4.41, followed by the opinion that the knowledge gained can be used in practice at a high level with the mean of 4.26. And the lowest mean belongs to the opinion that the availability of audiovisual equipment is at a high level with the mean of 3.58. Most respondents have opinions about training materials with a total mean of 3.90. When sorted the mean values in descending, it is found that the knowledge gained can be utilized for practice to a large extent with the mean of 4.39. The suitability of the content of the training is at a high level with the mean of 4.38 and the lowest mean fell to the opinion that the knowledge of this topic before attending the training was at the medium level with the mean of 2.61. Respondents are satisfied with the training course with a total mean of 4.21. When sorted the mean in descending, it is found that participants are very satisfied with the registration process at a high level with a mean of 4.31, followed by the satisfaction with snack food with the mean of 4.30 and the least mean falls to the location/training room with the mean of 4.05.

Discussion

1. The results in comparing the Generalizability Coefficient of concept map scores when the test pattern and the number of assessors are different, it is found that all conditions have a statistically significant difference at a level of 0.05, which is consistent with the assumption 1 stating that “the Generalizability Coefficient of the concept map score with the pattern that assessors inspect all the concept map of all students have a higher value than those of other scoring patterns. This may be due to the score obtained from the examination pattern that assessors inspect every concept map of all students is obtained by bringing the examination results from 2, 3, and 4 assessors to be averaged as the scores of each student which is the ideal form of scoring (Sudweeks et al., 2005).

When the examination patterns are the same but the numbers of assessor are different, it is found that all conditions have a statistically significant difference at the 0.05 level, which is consistent with the hypothesis set out in Article 2 stating that “Generalizabil-

ity Coefficient of the mind map score with the higher number of assessors will be higher the value than those the Generalizability Coefficient with a lesser number of assessors". This is consistent with Sirichai Kanjanawasi (2007) stating that whatever pattern of examination, multiple assessors should be used in scoring each test maker and then the average or total scores should be provided instead of the test maker's ability. This will help reduce the variance due to the assessors and the remaining discrepancies which will help increase the validity.

2. Comparing the concurrent validity of concept map score when the pattern and number of examiners are different, it is found that each condition, examination pattern, and number of assessors that differ will have a statistically significant difference at a level of 0.05, indicating that scores obtained from all conditions of the number of assessor and examination patterns have different values of concurrent validity. This is consistent with the assumption that "the Concurrent Validity of the concept mapping score with the examination pattern where the assessors inspect every concept map of all students have a higher value than those of other patterns in every number of assessors" and "the Concurrent Validity of mind map score with a higher number of assessors will have a higher value than those of lesser number of assessors in all patterns of examination". This may be due to the researchers explained the method of checking the mind map in detail to the assessors, thus allowing the assessors to understand and create the concept mapping through careful consideration of experts in every step. Thus, this enables students to fully understand and express their knowledge and ideas, resulting in the Concurrent Validity.

3. Experts commented on the arrangement of activities to apply the concept mapping for learning evaluation, the introduction, and the application of concept mapping for learning evaluation saying that they were the most appropriate. This indicated that the developed activity plans are appropriate for the implementation of the concept mapping for learning evaluation. This is because the researchers conducted a

detailed synthesis of course content, therefore, making the results of the expert's assessment the most appropriate.

4. Most respondents have a high overall opinion about the suitability of the course at a high level averagely and the lowest mean falls to the availability of audio-visual equipment. This is because the researcher may not be ready in the preparation of the conference room for training. And the results of this evaluation will be considered for the next training.

Most respondents have overall opinions about training materials with the mean value at a high level, and the lowest average mean value fell to knowledge on this topic before training, which is consistent with the concept of Novak (1983).

The respondents are satisfied with the overall training at a high average level. When sorting the mean in descending, it is found the lower average level in the matter of the location/training room. This is because the researchers may not be ready in the location/training room, and the results of this evaluation will be considered for the next training.

Recommendations

1. Recommendations in applying the research results

For scholars and educators interested in implementing the concept map, the pattern of examination where the assessors inspect every concept map of all students should be used. This is due to there are higher levels of Generalizability Coefficient and Concurrent Validity than other patterns of scoring. And the greater number of assessors should be used since the Generalizability Coefficient and Concurrent Validity are higher than those with lesser number of assessors.

2. Suggestions for the next research

2.1 Study in the feasibility of scoring the concept mapping using a computer or other scoring tools should be conducted.

2.2 Study in the implementation of the concept mapping should be conducted to assess students' learning outcomes using research methodology and development and to apply to other courses.

2.3 Study in the application of concept mapping for learning evaluation by using other techniques such as filling words in the map, creating the mind from words, creating words in lines so as to expand the knowledge in a broader range.

References

- Thatmanee Chukwan. (2005). A comparison in the quality of abstract notion in mathematic using mind map concepts evaluated using mind map with different scoring. Master of Education Thesis, Educational Measurement and Evaluation Program, Graduate School, Chulalongkorn University
- Wiyada Rawangsuk. (2002). Assessment of the abstract notion using mind mapping. Thesis. Master of Education, Department of Educational Measurement and Evaluation, Graduate School, Chulalongkorn University.
- Sirichai Kanjanawasee. (2007). New Test Theory. Bangkok: Chulalongkorn University Press.
- Angkhana Kunnapadol. (2012). Development of mind map scoring for assessing student learning outcomes. Dissertation. Doctor of Philosophy, Department of Research and Statistics in Cognitive Science, College of Research Methodology and Cognitive Science, Burapha University.
- Ausubel, D.P. (1968). Educational Psychology. New York.
- Buldu, M. & Buldu, N. (2010). Concept mapping as a formative assessment in college classroom: Measurement usefulness and student satisfaction. *Procedia Social and Behavioral Sciences*, 2(2010), 2099-2104.
- McClure, J.R., Sonak, B., & Suen, H.K. (1999). Concept map assessment of classroom learning: reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475-492.
- Novak, J.D. & Gowin, D.B. (1972). *Learning How to Learn*. New York: Cambridge University Press.
- Sudweeks, R.R., Reeve, S., & Bradshaw, W.S. (2005). A Comparison of Generalizability Theory and Man-Facet Rasch Measurement in an Analysis of College Sophomore Writing. *Assessment Writing*. 9(3):239-261.