Development and validation of analytic and holistic rubric guides in assessing concept cartoons

Randel D. Estacio[a]*, Department of Natural Science, Faculty of Science, Polytechnic University, 0860, Quezon City, Philippines.

Abstract

This study focuses on the development and validation of a teacher made analytic and holistic rubric guide in assessing Concept Cartoons in teaching science lessons. This study is a descriptive-developmental type of research. In validating and testing the reliability of the developed rubric guides, science teachers and experts, and students from Quezon City Polytechnic University served as respondents. Three analytic and three holistic rubric guides to assess three different forms of Concept Cartoons were developed, used and subjected for validity and reliability test. Results showed that the developed analytic and holistic rubric guides were valid and reliable. In conclusion, valid and reliable analytic and holistic rubric guides are important in giving a conclusive analysis of student's learning. With the use of valid and reliable rating scales strengthen the claims that concept cartoons help teachers in identifying and correcting students' misconceptions.

Keywords: Concept cartoon; holistic rubric; analytic rubric; assessment of learning; validity; reliability.

* ADDRESS FOR CORRESPONDENCE: **Randel D. Estacio**, Department of Natural Science, Faculty of Science, Polytechnic University, 0860, Quezon City, Philippines.
E-mail address: mr.rdestacio@yahoo.com / Tel.: +63998-2917507

1. Introduction

Assessment of learning is an integral part of the teaching and learning process. It is a process of collecting evidences of students' performance over a period of time to determine learning and mastery of skills (Navarro & Santos, 2012). Generally assessment aims to improve student's learning and provide them, their parents, and teachers with a reliable information regarding their progress and extent of attainment of the expected learning outcomes (Navarro & Santos, 2012). Teachers can assess their students' understanding and knowledge in various ways but it is important to infer from certain indicators of understanding through written descriptions (Navarro & Santos, 2012).

There are variety of assessment instruments or tools that can be use when assessing student learning outcomes. One of which is an assessment rubric. A rubric is an authentic assessment tool which measures student's work. It is a scoring guide that seeks to evaluate a student's performance based on a full range of criteria rather than single numerical score (Navarro & Santos, 2012; Malini-Reddy, 2007; Wolf & Steven, 2007). Authentic assessment tool like rubric allows students to perform real-world tasks which are either replicas or simulations of the kind of situation encountered by adult citizen, consumers, or professionals. Rubrics are used to assess non-objective test performance like psychomotor tests and written reports (Malini-Reddy, 2007; Wolf & Steven, 2007).

Recently, concept cartoons are used as teaching strategy to draw out student's ideas and eliminate misconceptions by means of presenting thought provoking questions and situations that can stimulate thinking and reasoning through the use of graphical representation of scientific facts with minimal text in a dialogue form (Naylor & Keogh, 2013). It presents different characters arguing about a given situation and a scientifically accepted and an alternative ideas are presented in each situation with equal status (Naylor & Keogh, 2013; Keogh, Naylor & Wilson, 1998).

Surveyed literature and studies shows that concept cartoon has the capability to increase the level of student's achievement in science (Cetin, Pehlivan, & Hacieminoglu, 2013; Estacio, 2015; Inel & Balim, 2012; Kaptan & Izgi, 2013; Letsaolo, 2011; Sahin & Cepni, 2011) and in mathematics (Sengul & Uner, 2010; Sexton, 2010). In addition, concept cartoons are used as an assessment of learning tool, studies revealed that when it is used as an assessment of learning tool it can help the teachers to easily identify student's misconceptions about the subject matter (Naylor & Keogh, 2013; Kabapinar, 2005). By means of identifying students' misconceptions the teacher may have an opportunity to improve his or her teaching strategies and styles to promote conceptual change (Keogh & Naylor, 1999) and at the same time correcting it (Ekici, Ekici & Aydin, 2007; Kabapinar, 2005).

According to Khalid, Meerah and Halim (2010) majority of the teachers believed that concept cartoons can be a great tool in assessing student's cognitive skills and can be used to identify and eliminate student's misconception in teaching physics subject, similar results were also found in the study conducted by Birisci, Metin and Karakas (2010). And this claim was supported by the results of the study conducted by Estacio (2015), Cetin, Pehlivan and Hacieminoglu (2013), Naylor and Keogh (2013), Akamca, Ellez, and Hamurcu (2009), Chin and Teou (2009), and Ekici, Ekici and Aydin (2007).

Literatures and studies revealed the effectiveness of concept cartoon as a tool for identifying and evaluating students' misconceptions (Estacio, 2015; Cetin, Pehlivan & Hacieminoglu, 2013; Naylor & Keogh, 2013; Akamca, Ellez, & Hamurcu, 2009; Chin & Teou, 2009; Ekici, Ekici & Aydin, 2007) without intimidating the student during the discussion, but the question now is how do the teachers assess student's response in a concept cartoon? How do they grade or give scores on student's answers when they use it as a formative or summative assessment tool? It is important to give ratings that are free from bias and personal judgment to this kind of assessment tool so as the teacher will arrived at a sound and valid interpretations on what the students really know towards the lesson. Therefore the primary goal of this study is to develop a valid and reliable rubric guides which can be used in assessing students' response in a concept cartoons.

2. Methodology

The study was conducted in Quezon City Polytechnic University, Quezon City, Philippines during the Academic Year 2016-2017. A total of six concept cartoons were developed based on the Table of Specification (TOS) of the topic "Earth's External Processes" and classified into three different types (see figures 1 - 3). Concept cartoons were then used as a formative assessment in a group of 30 Senior High School students after receiving a lesson about the topic "Earth's External Processes".
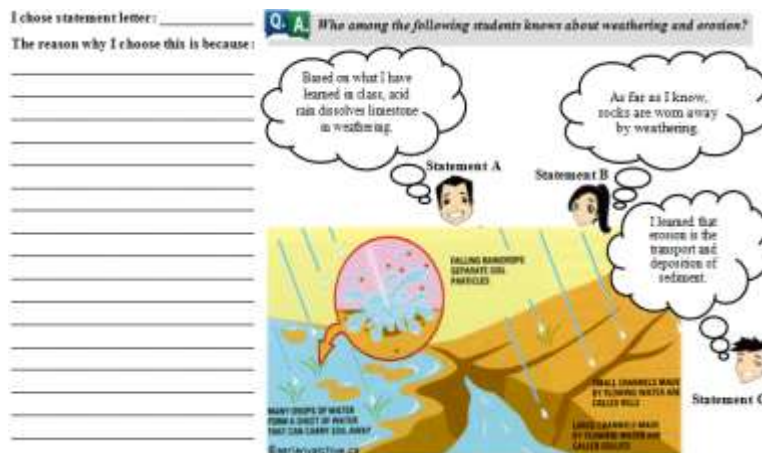

Figure 1. Concept Cartoon with Three Options

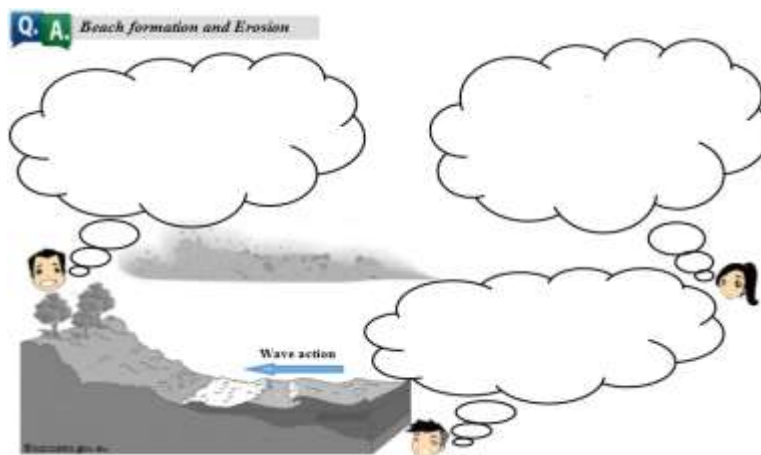**Figure 2. Concept Cartoon with One Empty Speech Bubble**

Figure 3. Concept Cartoon with Three Empty Speech Bubbles

Analytic and holistic rubric guides were then developed (see Appendix A) based on the objectives given in the Table of Specifications (TOS) and developed concept cartoons. During the initial development of the analytic and holistic rubric guides, six (6) science teachers in the collegiate and senior high school level and six (6) students (three from collegiate level and three from the senior high school level) served as the primary evaluator of the developed rubrics. Recommendations of the respondents were considered and reflected to improve the rubric guides (version 2). The second version of the rubric guides were evaluated by another group of teachers who are experts in the field of science education, a total of six experts (majority of them were masters' degree holders and teaching science subjects for more than ten years) validated the rubric guides. They were asked to establish the content and face validity of the of the final analytic and holistic rubric guides through the use of a standard checklist (see Appendix B). The checklist for content and face validity developed by Morales (2012) was used and slightly modified for this study. It consists of 20 items and a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) was used to measure the respondents' evaluation. Collected data were then analyzed using Microsoft Office Excel 2007 Statistical Data Analysis Tool. Suggestions of the experts were reflected in the final rubric guides (version 3).

In testing the reliability of the proposed rubric guides, Cohen's kappa, Fleiss' kappa, and Cronbach's alpha test were employed. For Cohen's kappa test, two science teachers evaluated the final rubric guides and the results of their ratings were analyze and interpreted by the researcher. For Fleiss' kappa test a total of twelve raters (six science teachers and six senior high school students) were asked to rate the final rubric guides and the results were subjected for analysis and interpretation. And finally, for the Chronbach's alpha test, thirty students were asked to answer the concept cartoons as part of their formative assessment, the students' responses were evaluated and scored using the validated analytic and holistic rubric guides. Collected data were then analyzed using Microsoft Office Excel 2007 Statistical Data Analysis Tool and interpreted by the researcher.

3. Results and Discussion

Generally, the study aims to develop a validated and reliable analytic and holistic rubric guides that can be use to assess students' response in a concept cartoons.

3.1. Validity of the analytic and holistic rubric guides

Generally validity has something to do with whether an instrument such as rubric guides is measuring what is intended to measure. Establishing validity involves examining the logical relationships that should exist between assessment measures. Results of an assessment are used to

predict future achievement and current knowledge, thus establishing the validity of the analytic and holistic guides in assessing concept cartoons is important.

Developed analytic and holistic rubric guides were subjected to descriptive and analytic methods of content validation by the experts. Descriptive validation was conducted for face validation, it emphasized the use of phrases or words to describe the assessment of the items. These are presented as comments, remarks, or suggestions of the experts. On the other hand, quantitative content validation made use of the 20-item validation checklist. The results of the validation are presented in Tables 1 and 2 with comments and suggestions of the experts.

The second version of the analytic and holistic rubric guides were validated by six experts in the fields. A validated and reliable 5-point Likert-scale evaluation checklist developed by Morales (2012) was used for validation purposes. Table 2 shows that the over-all mean is 3.67, suggesting that the experts evaluated the rubric guides within the highest continuum of the Likert scale. The value also suggests a good quality of rubric guides in construction and valid content wise. Comments and suggestions were used as bases for the revision of the rubric guides.

Table 1. Content validity of the second version of the analytic and holistic rubric guides

	Expert 1	Expert 2	Expert 3	Experts 4	Expert 5	Expert 6
Mean	3.73	3.66	3.64	3.60	3.70	3.68
Comments	Language used and instructions are somewhat vague	There are very few typographical errors. Descriptions are vague. Use language and descriptors that are suitable to the learners.	Typo errors are observed. Make a clear directions. Use descriptors that can easily be understood by the learners as well as the users.	Rephrase some descriptors under "Organization of Scientific Ideas". Modify the general direction.	There are some typographical errors.	Some descriptors are ambiguous, rephrase and make it clear and direct to the point.
Over-all Mean	3.67					

After the revision of the rubric guides based on the first validation cycle, the revised rubric guides (third version) were subjected to a second round of content and face validation. The rating improved with an over-all mean of 4.59 out of 5.00 by the same group of evaluators.

The new rating was an improvement of the rubric guides from the first cycle of validation. Each of the experts evaluated the rubric guides as close to 5.00 (see Table 2).

Table 2. Content validity of the third version of the analytic and holistic rubric guides

	Expert 1	Expert 2	Expert 3	Experts 4	Expert 5	Expert 6
Mean	4.68	4.57	4.55	4.53	4.62	4.60
Comments	Easy to understand. No typographical errors being observed.	No typographical errors found. Clear and easy to comprehend.	Better than the previous version.	Good and easy to understand.	No comment.	Better.
Over-all Mean	4.59					

In addition to mean values of experts, content validity coefficient was determined per checklist item to ensure that the rubric guides were actually rated as a content valid rubrics. Based on the results, the average content validity coefficient is 0.88, and according to Aiken (1985) the closer the coefficient value to 1.00, the higher content validity an item has. The experts who rated the items found the content of the rubric guides valid as shown in the values of content validity coefficients.

Table 3. Content validity coefficient of the third version of the analytic and holistic rubric guides

Checklist Item	Aiken's V (Content Validity Coefficient)
1	0.88
2	0.89
3	0.88
4	0.89
5	0.89
6	0.87
7	0.87
8	0.87
9	0.89
10	0.89
11	0.88
12	0.87
13	0.88
14	0.88
15	0.88
16	0.88
17	0.89
18	0.89
19	0.89
20	0.89
Average	0.88

3.2. Reliability of the analytic and holistic rubric guides

Reliability is an indicator of an assessment tool's consistency. The reliability of an instrument can be measured through inter-rater and internal consistency method.

Inter-rater reliability establishes the equivalence of ratings obtained with an instrument when used by different observers (Hafner & Hafner, 2004; Newell, Dahm, & Newell, 2002). If a measurement process involves judgments or ratings by observers, a reliable measurement will require consistency between different raters. Inter-rater reliability requires completely independent ratings of the same event by more than one rater. No discussion or collaboration can occur when reliability is being tested. Reliability is determined by the correlation of the scores from two or more independent raters or the coefficient of agreement of the judgments of the raters. Cohen's kappa is commonly used to determine the coefficient of agreement between two raters while Fleiss' kappa is used to determine the coefficient of agreement of more than two raters. Kappa is used when raters classify observations into categories based on rating criteria. Instead of a simple percent agreement, kappa takes into account the agreement that could be expected only by chance. Table 4 shows the results of the Cohen's kappa and Fleiss' kappa test of the rubric guides.

Table 4. Inter-rater reliability test

Rubric Guide	Cohen's kappa	Fleiss' kappa
	0.650	0.621
Analytic Rubric Guide for Concept Cartoon with three Options	0.654	0.656
Analytic Rubric Guide for Concept Cartoon with One Empty Speech Bubble	0.650	0.632
Analytic Rubric Guide for Concept Cartoon with Three Empty Speech Bubble		
Holistic Rubric Guide for Concept Cartoon with three Options	0.659	0.632
Holistic Rubric Guide for Concept Cartoon with One Empty Speech Bubble	0.673	0.651
Holistic Rubric Guide for Concept Cartoon with Three Empty Speech Bubble	0.656	0.650

Trends on the effectiveness of rubrics focuses on inter-rater reliability. The study conducted by Penny, Johnson, and Gordon (2000) revealed a very high inter-rater reliability scores for their rubrics while others have reported a low or moderate reliability.

The Cohen's kappa of the analytic rubric guides are 0.650, 0.654, and 0.650, while Fleiss' kappa value are 0.621, 0.656, and 0.632 which denotes that the developed analytic rubric guides are moderately reliable, since the kappa values fall within the acceptable range. On the other hand, holistic rubric guides have a value of 0.659, 0.673, and 0.656 for Cohen's kappa test and 0.632, 0.651, and 0.650 for Fleiss' kappa test, similar to analytic rubric guides, the developed holistic rubric guides are also considered moderately reliable due to the fact that the kappa values fall within the acceptable value.

Internal consistency gives an estimate of the equivalence of sets of items from the same test. The coefficient of internal consistency provides an estimate of the reliability of measurement and is based on the assumption that items measuring the same construct should correlate. Perhaps the most widely used method for estimating internal consistency reliability is Cronbach's alpha. Cronbach's alpha is a function of the average inter-correlations of items and the number of items in the scale. Table 5 shows the result of the internal consistency reliability test of the rubric guides.

Table 5. Internal consistency reliability test

Rubric Guide	k	Σvar	Var	Cronbach's alpha
Analytic Rubric Guides	6	1.528	3.512	0.678
Holistic Rubric Guides	6	1.599	3.806	0.696

Both analytic and holistic rubric guides are reliable based on the result of the internal consistency reliability test. The Cronbach's alpha of the developed analytic rubric guides is 0.678 while the holistic rubric guides is 0.696 which falls on the acceptable value. Even though the results were not greater than 0.70 the developed rubric guides were generally considered a reliable instruments.

Scoring rubrics' reliability and validity have been studied from several viewpoints. The study of Spandel (2006), and Wolfe (1997) focus on the objectivity of rubrics while Kohn (2006) and Mabry (1999) claimed that scoring rubrics are excessively reductive. While for some rubrics have not led to a more objective or more reliable grading. Ideally, the response given by using a scoring rubric is far better than by giving simple numerical or letter grade. However, untrained users of rubrics may simply use this guides to justify their subjective assessment. In addition, this educators believed that the irrelevant variables affecting analytic or holistic assessment of a performance or a written response may still affect rubric-based assessment. Decisions eventually turn on adjectives that are not clear and end up being left to the teacher's personal impressions (Kohn, 2006; Lumley, 2002). In contrast, the researcher observed that the raters, even though they are not trained users of rubrics graded the

Estacio, D. R. (2017). Development and validation of analytic and holistic rubric guides in assessing concept cartoons. *New Trends and Issues Proceedings on Humanities and Social Sciences.* [Online]. 4(1), pp 27-35. Available from: www.prosoc.eu

concept cartoons solely based on the content of the rubric guides and not on their personal impressions.

4. Conclusion and Recommendations

Rubric guides developed in this study after going through a validation process can be utilized as an assessment tool in scoring student's response in a concept cartoon. Valid and reliable rubric guides support the teachers in rating, evaluating and interpreting student's responses in a concept cartoon. By using analytic and holistic rubric guides teachers even though they are not trained to use rubrics allows to score student's answers in a concept cartoons objectively and free from bias. The results of this study should be examined very carefully. One should note that, in real life, the quality of answer and student's conceptual knowledge are associated to some degree. As a conclusion, valid and reliable analytic and holistic rubric guides are important in giving a conclusive interpretation of student's learning. With the use of valid and reliable rating scales, like the analytic and holistic rubric guides developed in this study will strengthen the claims that concept cartoons help teachers in identifying and correcting students' misconceptions.

References

Aiken, L. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45*(1), 131-142.

Akamca, G., Ellez, A. & Hamurcu, H. (2009). Effects of computer aided Concept Cartoons on learning outcomes. *Procedia-Social and Behavioral Sciences, 1*(1), 296-301.

Birisci, P., Metin, M. & Karakas, M. (2010). Pre-service elementary teachers' views on concept cartoons: A sample from Turkey. *Middle East Journal of Scientific Research, 5*(2), 91-97.

Cetin, E., Pehlivan, M. & Hacieminoglu, E. (2013). The effect of the Science and Technology course integrated with cartoons on students' achievement and attitudes. *Procedia - Social and Behavioral Sciences, 116,* 973 – 978.

Chin, C. & Teou, L. (2009). Using concept cartoons in formative assessment: Scaffolding students' argumentation. *International Journal of Science Education, 31(*10), 1307-1332.

Ekici, F., Ekici, E. & Aydin, F. (2007). Utility of Concept Cartoons in diagnosing and overcoming misconceptions related to photosynthesis. *International Journal of Environmental & Science Education, 2*(4), 111-124.

Estacio, R. D. (2015). The Effect of Concept Cartoons as an Instructional Material and Formative Assessment Tool in Teaching Evolution and Diversity on the Achievement of Freshmen College Students. In *Biology Education and Research in a Changing Planet* (pp. 71-79). Singapore: Springer.

Hafner, J. & Hafner, P. (2004). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education,* 25(12), 1509-28.

Inel, D. & Balim, A. (2012). Concept cartoons assisted problem based learning method in science and technology teaching and students' views. *Procedia - Social and Behavioral Sciences, 93,* 376–380.

Kabapinar, F. (2005). Effectiveness of teaching via concept cartoons from the point of view of constructivist approach. *Educational Sciences: Theory and Practice, 5*(1), 135-146.

Kaptan, F. & Izgi, U. (2013) The effect of use concept cartoons attitudes of first grade elementary students towards science and technology course. *Procedia - Social and Behavioral Sciences, 116*, 2307 – 2311.

Keogh, B. & Naylor, S. (1999). Concept cartoons, teaching and learning in science: An evaluation. *International Journal of Science Education, 21*(4), 431-446.

Keogh, B., Naylor, S. & Wilson, C. (1998). Concept cartoons: A new perspective on physics education. *Physics Education, 33*(4), 219-224.

Khalid, H., Meerah, S. & Halim, L. (2010). Teachers' perception towards usage of cartoon in teaching and learning physics. *Procedia Social and Behavioral Sciences 7,* 538–545.

Kohn, A. (2006). The trouble with rubrics. *English Journal*, *95*(4), 12–15.

Letsaolo, R. (2011). *The use of Concept Cartoons and Prompt Sheets in supporting learners in the planning of scientific investigations* (Unpublished doctorate thesis). South Africa: University of Johannesburg.

Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, *19*(3), 246–276.

Mabry, L. (1999). Writing to the rubric: Lingering effects of traditional standardized testing on direct writing assessment. *Phi Delta Kappan*, *80*(9), 673–679.

Malini-Reddy, Y. (2007). Effect of rubrics on enhancement of student learning. *Educate Journal, 7*(1), 3-17.

Morales, M. (2012). Development and validation of a concept test in introductory physics for biology students. *The Manila Journal of Science, 7*(2), 26-44.

Estacio, D. R. (2017). Development and validation of analytic and holistic rubric guides in assessing concept cartoons. *New Trends and Issues Proceedings on Humanities and Social Sciences.* [Online]. 4(1), pp 27-35. Available from: www.prosoc.eu

Navarro, R. & Santos, R. (2012). *Assessment of learning outcomes (assessment 1).* Quezon City, Philippines: Lorimar Publishing, Inc.

Naylor, S., & Keogh, B. (2013). Concept cartoons: what have we learnt?. *Journal of Turkish Science Education, 10*(1), 3-11.

Newell, J., Dahm, K. & Newell, H. (2002). Rubric development and inter-rater reliability: Issues in assessing learning outcomes. *Chemical Engineering Education*, *36*(3), 212–215.

Penny, J., Johnson, R. & Gordon, B. (2000). Using rating augmentation to expand the scale of an analytic rubric. *Journal of Experimental Education*, *68*(3), 269–287.

Sahin, C. & Cepni, S. (2010). Developing of the concept cartoon, animation and diagnostic branched tree supported conceptual change text:"gas pressure". *Eurasian Journal of Physics and Chemistry Education, 1*(1), 25-33.

Sengul, S. & Uner, I. (2010). What is the impact of the teaching "Algebraic Expressions and Equations" topic with concept cartoons on the students' logical thinking abilities?. *Procedia Social and Behavioral Sciences, 2,* 5441–5445.

Sexton, M. (2010). Using concept cartoons to access student beliefs about preferred approaches to Mathematics learning and teaching. *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia.* Fremantle: MERGA.

Spandel, V. (2006). In defense of rubrics. *English Journal*, *96*(1), 19–22.

Wolf, K. & Steven, E. (2007). The role of Rubrics in advancing and assessing student learning. *The Journal of Effective Teaching, 7*(1), 3-14.

Wolfe, E. W. (1997). The relationship between essay reading style and scoring proficiency in a psychometric scoring system. *Assessing Writing*, *4*(1), 83–106.