



Relationship between Point-Based Evaluation and Comment-Based Evaluation in Self-Evaluation

Kenichi Goto

Curriculum Research Center, National Institute for Educational Policy Research (Japan)

ken1510@nier.go.jp

Abstract

This research focuses on “enhancement of learning evaluation”. That is, the research plan was formulated with a thought that analyzing the relationship between comment-based evaluation and point-based evaluation under the evaluation standards for a mutual evaluation sheet based on criterion-referenced proactive design leads to curriculum management, and to improvement of teaching of teachers and learning for children. In this research, comparative reviews of the respective evaluations and the relationship between them were made. Specifically, the research was conducted through the following three comparative reviews: comparative review I: effects of rewriting; comparative review II: effects of repetition; and comparative review III: effects of retention. In the comparison, the relationship between point-based evaluation and comment-based evaluation was reviewed using a statistical approach. Consequently, the comparison has revealed that there is an interesting relationship between point-based evaluation and comment-based evaluation.

1. Introduction

Evaluation that contributes to the integration of teaching and evaluation based on criterion-referenced proactive design is required for “enhancement of learning evaluation.” In traditional learning evaluation, there were many methods under which a teacher measures achievements of learners according to the purpose. Although the conventional evaluation methods show the strengths in ranking and screening of children, they fall short of contributing to their improvement. Future evaluation requires improvement of the situation and an approach of evaluation in which not only teachers but also learners participate while keeping progress in mind. As such a view, Mr. Miyoshi and Ms. Tanaka (2001) presented the necessity and vision of “participatory evaluation.” Ms. Minamoto (2008) stated that the approach “arose from a spirit of attempting to make use of evaluation as a method to improve society.” From this view, it is thought that while evaluation was originally utilized for the purpose of “improving society,” positioning such an evaluation approach as a learning evaluation approach may lead directly to “improvement of learning.”

Comment-writing based evaluation is a participatory evaluation method. In education in Japan, teachers traditionally have written comments on deliverables submitted by children as learning outcomes and returned the deliverables to them on a regular basis. The comment-based evaluation in utilization of a mutual evaluation sheet (Mr. Goto, 2013) covered in this research has a similar approach. These evaluation approaches are deemed to contribute to improvement. Analyzing the relationships between point-based evaluation under evaluation standards and evaluation comments will lead to further improvement of teachers’ teaching and of learning for children.

2. Questions and purpose

In this research, point-based and comment-based self-evaluations were made. Also, comparative reviews of changes in points of self-evaluations that were made four times and the analysis of the relationship between these changes and self-evaluation comments were both conducted. Such analysis is expected to reveal learners’ tendencies regarding what and how they learn and how they comment to describe their learning in the experiment. The results may lead to improvement of teachers’ teaching and students’ learning. In this context, the following purpose is set:

To clarify the relationship of learning by analyzing changes in points of self-evaluations and words seen in self-evaluation comments.



3 Method of research

3.1 Overview of practice

Table 1 shows the overview of practice done in this research.

Table 1 Overview of practice in this research

Theme Practice period	(1) Experiment 1 Identification of oxygen by fundamental measurement May 2014, 150 people	(2) Experiment 2 Identification of urea by cryoscopic measurement June 2014, 147 people	(3) Experiment 3 Development of plan for identification of sugar and salt and identification by measurement January 2015, 118 people
Content	Volume and mass measurement and identification of gas	Cryoscopic measurement and identification	Experiment plan and measurement identification of the nature of material
Issues	Regarding two types of gases, oxygen and nitrogen, identify which is oxygen, using various information.	Regarding two types of aqueous solutions, urea and sodium chloride, identify which is urea, using data on cryoscopy.	Form an <u>experiment plan</u> to identify white powder, which is sugar or salt, perform experiment tests and identify sugar or salt.
Overview	Experiment to identify oxygen from two cylinders, each of which contains oxygen or nitrogen.	Perform experiment tests in advance based on the experiment plan to identify urea from solutions, each of which contains urea or sodium chloride at the same concentration.	Experiment to identify unknown samples (sugar and salt) Work out a “plan, experiment and conclusion” on an individual basis
Issues for consideration	Offer a conclusion on which is oxygen and give the reason	Offer a conclusion on which is a urea aqueous solution and give the reason	Offer a conclusion on which is salt and give the reason

3.2 Research method and analysis method

In the research, three comparative reviews for evaluating description of considerations are performed.

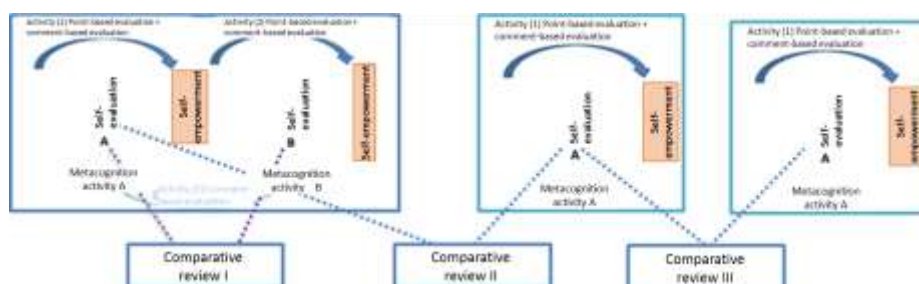


Chart 1 Comparative reviews in this research

In each review, a paired comparison was made on the percentage of students who could give a description according to each item of evaluation standards. Comparative review I examined effects of rewriting. That is, this review is aimed to find any change of students in preparation of a report by rewriting. Next, Comparative review II examined effects of repetition. This review is aimed to find any change of students in preparation of a report by making a comparison with evaluation of description of considerations submitted at first in Experiment 1 and Experiment 2. Furthermore, Comparative review III examined effects of retention. That is, an analysis was made using a statistical method in both point-based evaluation and comment-based evaluation, to know which kind of change was found in description in students' reports over time. In comment-based evaluation, words were extracted by reference to the research of Mr. Iida (2015). In this research, the point-based evaluation approach and comment-based evaluation approach in self-evaluation were reviewed in light of Mr. Iida's research. With the use of WordMiner, commercially available software, the analysis utilized the word divider function for self-evaluation comments and covered words ranked in the top 1,000 for the appearance frequency. In addition, statistical analysis was made in both comment-based evaluation and point-based evaluation, using SPSS.

4. Research results

Table 2 shows results of analysis obtained through McNemar's test for point-based evaluation in Comparative reviews I to III



Table 2 Results of point-based evaluation in Comparative reviews I to III

Evaluation items	Comparative review I Effects of rewriting	Comparative review II Effects of repetition	Comparative review III Effects of retention
(1) Did he or she describe a content that suits the purpose?	n, s	n, s	n, s
(2) Necessary Key words	++	n, s	--
(3) Did the description contain his or her own opinion (impression and feelings)?	++	++	n, s
(4) Did he or she give only explanation of experimental outcomes when making examination?	++	++	++
(5) Did he or she give a reason for the conclusion = "considerations"?	++	n, s	n, s
(6) Is the content of his or her argument right?	n, s	n, s	n, s
(7) Did he or she clearly describe what "way of thinking" he or she used?	++	n, s	++
(8) Was evidence presented clearly in consideration?	++	++	--
(9) Is there any error such as incorrect correspondence between subject and predicate, typographical error, omission, incorrect use of postpositional particle or connective (including conjunction and demonstrative pronoun)?	++	n, s	++
(10) Is one sentence too long, does one sentence contain too many information, or does the amount of document exceed the given limit?	++	n, s	n, s
(11) Is the text easy to read?	++	n, s	n, s

The meanings of signs are as follows; n.s: insignificant,

significant increase: ++ · · · p<0.01 + · · · p<0.05

significant decrease: -- · · · p<0.01 - · · · p<0.05

Likewise, Table 3 shows results of analysis obtained through McNemar's test for comment-based evaluation in Comparative reviews I to III.

Table 3 Results of comment-based evaluation in Comparative reviews I to III

	Comparative review I	Comparative review II	Comparative review III
Appearance of affirmative words	++	-	+
Appearance of negative words	--	-	n, s

The meanings of signs are as follows; n.s: insignificant,

significantly increase: ++ · · · p<0.01 + · · · p<0.05

significantly decrease: -- · · · p<0.01 - · · · p<0.05

Mr. Goto and others (2015) summarized results on point-based evaluation in Comparative reviews I to III in Table 2 as follows. Specifically, except for "(1) Did he or she describe a content that suits the purpose?" and "(6) Is the content of his or her argument right?" which relate to effects of rewriting, all items increased significantly in Comparative review I. In Comparative review II, no effects of repetition decreased significantly. In Comparative review III, there were items that increased significantly and those that decreased significantly in evaluation of retention (thinking individually without an experiment plan given).

5. Considerations

The research showed that the patterns of frequency in words that showed up were similar in comment-based evaluation of students. Reasons for the similarity are considered to be "1) attributed to the purposes" and "2) attributed to factors of the content of experiment, etc."

5.1 Attributed to the purposes

(1) Comparative review I: effects of rewriting

In Comparative review I, rewriting brought about effects as seen in the significant increase in point-based evaluation except for two elements. The relationship between this situation of point-based evaluation and comment-based evaluation is examined below.



When comments for oxygen (1) and those for oxygen (2) are compared in Tables 2 and 3, as common points, the percentage of appearance of the affirmative words increased significantly and the percentage of appearance of the negative words decreased significantly (both $N=126$, significance probability = 0.000, $p < 0.01$). This is because students' realization that they became "able" to write by rewriting appeared as increases in affirmative words and decreases in negative words in their evaluation comments. It also shows that students' pleasure that they could not write but then became able to write may have led to their confidence in learning.

(2) Comparative review II: effects of repetition

Although no items decreased significantly in point-based evaluation when students repeated, both affirmative words ($N=132$, significance probability = 0.049, $p < 0.05$) and negative words ($N=132$, significance probability = 0.012, $p < 0.05$) decreased significantly in comment-based evaluation. The reason for the significant decrease in affirmative words can be that students are seeking learning of higher quality. The reason why negative words also decreased significantly can be that those descriptions did not need negation.

(3) Comparative review III: effects of retention

Affirmative words increased significantly ($N=110$, significance probability = 0.013, $p < 0.05$) in comment-based evaluation in self-evaluation of students for retention. There was no significant difference in negative words. The reason why affirmative words increased significantly is thought to be that students made efforts for retention and a sort of confidence was being developed in them. Just as in Comparative review II, the reason why there was no significant difference in negative words is thought to be that the descriptions did not need negation.

5.2 Attributed to the content of experiment

The themes of experiments covered in this research, "(1) Experiment 1: Identification of oxygen by fundamental measurement, (2) Experiment 2: Identification of urea by cryoscopic measurement, (3) Experiment 3: Development of plan for identification of sugar and salt and identification by measurement" all relate to the concept of "identification." Issues for consideration: "offer a conclusion on which is oxygen and give the reason," "offer a conclusion on which is a urea aqueous solution and give the reason" and "offer a conclusion on which is salt and give the reason," also give similar challenges for consideration. In this regard, it is primarily assumed that results of comments obtained did not affect the experimental themes themselves. However, in terms of experimental methods, teachers presented an experimental plan in advance in "(1) Experiment 1: Identification of oxygen by fundamental measurement" and "(2) Experiment 2: Identification of urea by cryoscopic measurement," while students thought about an experimental plan in "(3) Experiment 3: Development of plan for identification of sugar and salt and identification by measurement." Since Experiment 3 requires students to think more, comment-based may have been affected in Experiment 3.

5.3 General overview

Common grounds of tendency in the graphs were seen in both affirmative words and negative words. Effects of rewriting were shown as frequency in students' evaluation comments. Teachers need to be sensitive to changes in these words and link those changes to teaching. In the experiment of "salt," both affirmative and negative words were on the increase. Potentially, students could learn deeper, and therefore recognize the affirmative part and the negative/improvement part from various aspects while monitoring, evaluating and performing metacognition of them, instead of merely describing affirmation or negation and improvement.

Given the frequency of appearance of affirmative words, the frequency of appearance of negation/improvement, and the tendency, it is thought that students are becoming able to monitor and perform metacognition from various aspects and directions. In addition, if other-evaluation (point-based evaluation and comment-based evaluation) activities are incorporated, as Mr. Iida (2015) presented, we can shape a learning model in which students get encouragement from others and thus gain self-confidence, resulting in further learning.

6. Conclusion

In the future, it is necessary to raise the precision of the method by which simply picking up expressions from children's words and monitoring their learning lead to improvement in teaching and their learning as well as that of the method by which actual situations of children's learning and teaching can be understood from those words. This should allow the improvement of learning and lessons.



References

- [1] The Central Council for Education (2015), the special committee for educational curriculum planning, summary of issues (proposed)), support documentation, the Ministry of Education, Culture, Sports, Science and Technology
- [2] Hiroshi Iida and Kenichi Goto (2015), "Trial Implementation of Science Classes Using Mutual Evaluation Sheets and Science Class Research at an Upper Secondary School— Focusing on Improved Motivation to Learn —," Journal of Research in Science Education (submitted: scheduled to be published)
- [3] Kenichi Goto (2013) "Developing a study program and evaluation of Science Project Study Focusing on the abilities to judge and express oneself (Research on effects of self-evaluation in chemistry experiments at high school –making use of mutual evaluation sheets-)," Journal of Research in Science Education, Vol. 54, No. 1, 13-26.
- [4] Kenichi Goto, Kenji Matsubara, Yorikazu Nonai, Takuya Miyauchi, Teruhiro Kitagawa, Shosuke Teratani, Shizuo Matsubara (2015) "Consideration toward Practicing Model-based Learning for High School Chemistry" Journal of Science Education In Japan, (submitted)
- [5] Yuriko Minamoto (2007) "Theory and practice of participatory evaluation" Koichi Miyoshi "For People Learning Evaluation Theory" Sekai Bunka Publishing Inc., 95-112
- [6] Koichi Miyoshi and Yayoi Tanaka (2001) Future prospects of participatory evaluation - a study for the concept and practice of participatory evaluation," "Japanese Journal of Evaluation Studies" Vol.1, No.1
- [7] The Ministry of Education, Culture, Sports, Science and Technology (2014), "Appropriate curriculum standards, etc. in elementary and secondary education(consultation)"