Effectiveness of Group Moderation Program for Developing Professional Assessment Skills of Teachers

HIRONORI SASAKI*, YUKA KAWASAKI**
Graduate School
Hyogo University of Teacher Education
942-1 Shimokume, Kato, Hyogo 673-1494
JAPAN

HARUHISA YAMAGUCHI
Department of Education
Okayama University
3-1-1 Tsushimanaka, Okayama 700-8530
JAPAN

YUMI YAMAGUCHI
Sanyo Gakuen University
1-14-1 Hirai, Okayama 703-8501
JAPAN

Abstract: - In this study we created and applied a system of “group moderation” to clarify the criterion of the grades that teachers gave to students. In group moderation, teachers gathered with students’ works and graded them. Then, they examined the reasons why they gave a particular grade and why other teachers graded higher or lower. The purpose of Group Moderation is to maintain consistency in teachers’ judgments in criterion-based assessment. We set Group Moderation emphasizing on how to develop the professional assessment skill of teachers. We have also practiced Group Moderation in a seminar at the Okayama Education Center in Okayama, Japan. The results of this study showed that 1) group moderation improved the assessment ability of teachers to create the feedback on how to instruct students to achieve a higher level. 2) Group Moderation enhanced their ability to maintain the consistency of assessment.

Key-Words: - Group moderation, Teacher training, Assessment, Education center, Seminar

1 Introduction
After the release of the report on assessment by the Curriculum Council in December, 2000, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan released a notification regarding the improvement of Cumulative Guidance Record. In the notification, the assessing of subjects was changed from the norm-referenced assessment system to the criterion-based assessment system. In Japan, teachers are not traditionally accustomed to the criterion-based form of assessment. Therefore, the lacking of objectivity has become a serious issue. Because high schools use the result of assessment from junior high schools to select the freshmen in the entrance examination, it is necessary for the prefectural education centers for teacher education to support teachers in this area.

In this study, we applied “group moderation” that Gipps proposed in 1994. “Here examples of work are discussed by groups of teacher or lecturers; the purpose is to arrive at shared understandings of the criteria in operation and thus both the processes and the products of assessment are considered” [1]. In the process of group moderation, teachers meet to discuss the assessment of their students’ works, set a standard and then agree on a score for the standard. This procedure has been used in some districts in the United Kingdoms, Australia, and the United States. The Senior Secondary Assessment Board of South Australia (SSABSA) has published an assessment and moderation policy on their website [2].
would like to propose a program on the procedure of group moderation to the board of education and the schools in Japan.

The purpose of group moderation is to develop and maintain consistency of teachers’ judgments. Suzuki (2002) indicated that the process of group moderation could assist teachers in developing their professional skills [3]. Effectiveness of developing professional skills for teachers in group moderation has not been discussed so far. We set group moderation putting emphasis on how to develop the professional skill of assessment. The purpose of this study is to examine the development of professional skills of teachers through the development of a group moderation program, and practice it in a seminar at the Education Center in Okayama, Japan.

2. Development of Program
Earlier research (Sasaki, 2003) created and reported a design model of a procedure of group moderation [4]. We will apply it in this study.

2.1 Required Condition
2.1.1 Form
In order to increase each participant’s level of experience, a small group of 3 to 4 teachers meet for a conference. The program of the conference is decided before-hand, to ensure that the opinions of one teacher do not dominate the discussion.

2.1.2 Assessment Methods
Assessment methods can vary from research reports and observations. Therefore, when teachers write their teaching plans, they should set an assessment criteria to use with the assessment. So we used students’ reports as a work sample for the purpose of group moderation.

2.1.3 Criteria in the Lesson
The National Institute for Educational Policy Research in Japan published a set of criteria for the scopes for each subject, but not for individual assessment pieces. So we used a set of criteria that a textbook had suggested because in a city or prefecture, all schools use the same textbooks.

2.1.4 Taking Steps to Guarantee Scholastic Ability
MEXT presented the policy of designation the National Guideline of Japan as a minimum standard. Making sure all students achieve Standard B is the most important issue because it guarantees students’ scholastic ability. We added the step which teachers create feedback to enable students to achieve higher
level, raising the work standard from Standard C to Standard B. The procedure of the program is shown in Figure 1.

![Figure 1](image1)

**2.2 The Goals of the Program**

In this study we assured the professional development consisted of assessment ability of feedback and consistency of teachers’ judgment.

**2.2.1 Creating Feedback**

It is possible to raise the assessment ability to create the feedback on how to instruct students to achieve a higher level.

**2.2.2 Consistency of Teachers’ Judgment**

It is possible to ensure the consistency of assessment concerning the student work, which is used inside the group of the conference.

**3 Experiment**

**3.1 Date, Subjects and Purpose**

Date: June 20th, 2003

Participants: 35 teachers (20 elementary school teachers, 7 junior high school teachers, 7 senior high school teachers and a special education teacher.)

Purpose: We conducted the group moderation program at the seminar and discussed whether the goals were achieved.

**3.2 Practice**

As for the teachers who become the subjects of the program, it was desirable that they were from the same Level of schooling, from the same subject area and furthermore, they use the same pieces of students work. However, this proved difficult as the participants at this seminar came from different levels of schooling. Next, as shown in Figure 2, the piece of work “Automobile Newspaper” was chosen as the assessment piece for discussion. This is the work of a 5th year student from Social Studies at an elementary school. This is a piece, which, was used as an example for group moderation.

At first the participants were divided into 10 groups of 3 or 4 people, 6 elementary school groups, 2 junior high school groups and another group.

At second they used the criteria provided by the text, “Industrial Production of our Country”, to assess the “Automobile Newspaper”. They decided the three specific criteria were 1) Collecting information from automobile company properly 2) Using fundamental materials effectively 3) Devising the expression intelligible.

At third we marked the serial number onto the pieces of work “Automobile Newspaper” from No.1 to No.7 and handed there out to each group. Then they assessed individually with A, B, C according to three criteria, and recorded the score in the card with A into 3, B into 2, C into 1. So the maximum score is 9 and the minimum is 3. It took about 2 minutes to assess a piece of work and rotated in the group.

At fourth they calculated the sum and arranged them on the desk according to the score. Then they talked about individual scores as a group, discussing the reasons why they gave the score and why other teachers graded higher or lower. And they decided the boundaries of A-B and B-C and marked the lines. (Fig. 3)

Furthermore, they also discussed how they might help the student move from a C Standard to a B Standard. Cards were posted on which teachers wrote how to instruct a student to achieve a higher level.

After short break, we handed out another 5 pieces of newspaper which we marked serial numbers from No.8 to No.12. They assessed them and recorded the score. Finally, they looked back and discussed the procedure as a review and then ended the program.

**4 Result**

**4.1 Score of Posted Cards**

The score of the number of the posted cards on which teachers wrote the way of instruction is in Table 1.

<table>
<thead>
<tr>
<th>Group</th>
<th>No.1</th>
<th>No.2</th>
<th>No.3</th>
<th>No.4</th>
<th>No.5</th>
<th>No.6</th>
<th>No.7</th>
<th>Sum</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>0</td>
<td>9</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>27</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>19</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>16</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>0</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>22</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>26</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>22</td>
<td>2</td>
<td>13</td>
<td>18</td>
<td>26</td>
<td>0</td>
<td>122</td>
<td>6.1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Score of Posted Cards
4.2 Score of Assessment
The records of each group and the average scores on pieces of Automobile Newspaper are in the Table 2.

<table>
<thead>
<tr>
<th>Group</th>
<th>No.1</th>
<th>No.2</th>
<th>No.3</th>
<th>No.4</th>
<th>No.5</th>
<th>No.6</th>
<th>No.7</th>
<th>ave</th>
<th>No.8</th>
<th>No.9</th>
<th>No.10</th>
<th>No.11</th>
<th>No.12</th>
<th>ave</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td></td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td></td>
<td></td>
<td>4</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td></td>
<td></td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>STD.DEV</td>
<td>1.15</td>
<td>1.39</td>
<td>0.94</td>
<td>0.97</td>
<td>1.66</td>
<td>1.66</td>
<td>0.89</td>
<td>1.00</td>
<td>0.82</td>
<td>0.92</td>
<td>0.88</td>
<td>0.88</td>
<td>1.03</td>
<td>0.69</td>
</tr>
</tbody>
</table>

We used only the score of elementary school teachers because other teachers have no experience in teaching that area.

5 Discussion
5.1 Creating Feedback
The total number of the posted cards was 122. Otherwise average was 6.1. They wrote various kinds of instructions on theposted cards, for example “write indexes”, “distinguish information and opinion,” “examine the contents in the paragraph”. Many of them were appropriate and suitable for the context of the students learning. Practical thinking of expert teachers was characterized as contextualized thinking [3].

This fact indicated that they raised their ability to create feedback on how to instruct students to achieve a higher grade through the discussion.

5.2 Consistency of Teachers’ Judgment
The largest standard deviation was 1.66 of No.5 between No.1 to No.7. The smallest standard deviation was 0.89 of No.6. The average of standard deviation of No.1 to No.7 was 1.15, which was comparatively larger. The numbers of newspaper that the standard deviation was higher than 1.0 was 4.

The largest standard deviation was 1.03 of No.11 among No.8 to No.12. On the other hand, the smallest standard deviation was 0.69 of No.12. The average of standard deviation of No.8 to No.12 was 0.89, which was smaller compared to 1.15 among No.1 to No.7. The number of newspaper which standard deviation was more than 1.0 was just 1. And difference Standard deviation afterwards in 4 out of 6 groups is smaller than that of before. In 4 groups the variation among teachers in assessing students’ work was smaller.

As a result, we confirmed that the group moderation program enhanced their ability to ensure consistency of assessment.

6 Conclusion
We set group moderation putting emphasis on how to improve the professional skills of teachers. We also practiced it in a seminar at the Okayama Education Center in Okayama, Japan. The results of this study showed that; 1) the group moderation program improved the assessment ability of teachers to create the feedback on how to instruct students to achieve a higher level, 2) the group moderation program enhanced teachers’ ability to develop the consistency of assessment.

We would like to suggest that future researchers should apply statistics to obtain more reliable results and verify the effectiveness of group moderation. We are also planning to run the program after we have selected identical pieces of work from the same subject, grade and unit, as well as the teachers from those areas. In addition, by offering the program to the board of education, the schools, and the research groups, we believe that we can secure objectivity when the teacher assesses the students’ samples for the scopes for each subject.
References: