Original article

Improvement of orthoptic students' self-evaluation through repeated practice with rubric-based assessments

Haruo Toda, Noriaki Murata, Fumiatsu Maeda, Hokuto Ubukata, Ryosuke Kawachi, Yoshiko Tani, Haruki Abe

Department of Orthoptics and Visual Sciences, Faculty of Medical Technology, Niigata University of Health and Welfare, Niigata, Japan

Key words: rubric assessment, orthoptics, student practice

Received: 11 October 2017/Accepted: 16 January 2018

Abstract

The discipline of orthoptics covers a wide range of ophthalmological examinations and generates supports for visually disabled persons; consequently, orthoptic educators are continually trying to improve practical training. One problem is students' inadequate abilities to self-evaluate and consistency among teacher evaluations. In the United States of America, rubric assessments have been used widely, from elementary through higher educational levels, since the late 1970's. Because a rubric contains the explicit and descriptive criteria for scoring, assessments with rubric can help students and instructors consistently assess student performance. In a previous study, rubric assessments for off-campus clinical orthoptic internships were found useful in standardizing evaluations among multiple instructors. However, there have been no empirical studies on the long-term effects of orthoptic training with rubrics. While writing accurate reports on the practice is important for acquisition of clinical skills through practical course, students cannot obtain report-writing skills without accurate self-evaluation of their own reports. In this study, to investigate the long-term effects of rubric-based training, we repeated practice-assessment-feedback cycles for nine weeks using rubrics, and analyzed the observed disparities between self- and teacher-generated scores over time. All study procedures were performed in-house. We found that, over time, the disparities between the students' and teachers' scores decreased significantly, suggesting that students' self-evaluation accuracy improved through repeated use of rubric. The improvement was category-dependent and larger in students with lower performance. These results indicate the advantages of repeating rubric-based formative assessments in orthoptic practices.

Introduction

In orthoptic education, as well as in education in other medical disciplines, practical training is essential. In Japan, certified orthoptists (COs) perform a wide range of ophthalmological measurements (*e.g.*, fundus photography, measurements of visual acuity, visual field, and number of the corneal endothelial cells) in patients with various ophthalmic diseases. Furthermore, supports and assistance for persons with visual disabilities are developed by COs [1,2]. Given the diverse job duties of COs, practical training in orthoptics can be difficult for students without adequate self-evaluation skills. In reality, Tabuchi and Maeda have

Corresponding author: Haruo Toda

Department of Orthoptics and Visual Sciences, Faculty of Medical Technology, Niigata University of Health and Welfare, 1398 Shimami-cho, Kita-ku, Niigata 950-3198, Japan

TEL/FAX: +81-25-257-4753, E-mail: toda@nuhw.ac.jp

reported that, in Japan, 46% of ophthalmologists think it takes six or more months before they are satisfied with the performances of COs who newly graduated from CO-training faculties or institutes. The data suggest an additional on-job training is needed for newly graduated COs to achieve the clinically expected level of performance because undergraduate training is still incomplete [3]. Therefore, we established requirements for improving practical training and developing students' self-evaluation abilities.

A rubric is a scoring tool for assessing student performance that uses explicit criteria [4]. In the United States of America, rubrics have been used since the late 1970's [5]. Because the scoring criteria are presented to the instructors in an explicit and descriptive way, the rubrics help the instructors consistently assess students' performances on a wide-range of tasks, from elementary through higher-educational levels [6-10]. Similarly, rubrics can be used by students for evaluating their own performance. If the same rubric is shared by both students and their supervisors, repeated assessments using rubrics can develop students' abilities to evaluate their own performance, making practical training more effective.

Maeda et al. introduced rubric assessments for off-campus clinical orthoptic internships, and found them useful for standardizing intern evaluations by both students and instructors [11]. However, the previous study did not evaluate long-term effects of orthoptic training using rubrics because of two obstacles: 1. The observation period of the study was rather short (6 days), 2. Only one rubric assessment was performed.

The ability of writing accurate reports on what they learnt in the practice is one of the bases for acquisition of clinical skills through practical course. Students cannot obtain report-writing skills without accurate self-evaluation of their own reports; therefore, in this study, we employed the accuracy of self-evaluation of practical reports to observe the long-term educational effects of rubric-based assessment in the clinical practice. We used nine repeated rubric-based practice-assessment-feedback cycles and recorded the observed disparities between students' self-evaluation scores and teachers' scores over time during an in-house practice course.

Materials and Methods

1. Subjects

This study was approved by the Research Ethics Committee of Niigata University of Health and Welfare (Number:17827-170829). The data were obtained from 56 second-year undergraduate students enrolled in the "Visual Function II" (practical training of ophthalmic examinations) course in the Department of Orthoptics and Visual Sciences of Niigata University of Health and Welfare. All subjects provided written agreement to participate (except for one student who withdrew from the university after the course). Forty-two were female, 14 were male, and the average age was 19.6 years at the beginning of the course. The teachers consisted of five COs, one ophthalmologist, and one Doctor (PhD) of Engineering, all from our department.

2. Practice Schedule and Rubric

The practices were carried out once a week from April 9 to July 16, 2015. The rubric assessments covered the nine topics of the course under the instruction of seven teachers (Table 1). These topics were divided into three parts and every part was performed in one of three periods ("Rounds" 1 to 3) of the course. Students were also divided into three groups, each consisting of 18 or 19 members, and every group performed one of the three topics belonging to the same "Round." In this study, we focused on these nine weekly practices (we refer these weeks as Week 1 to Week 9). A scheme of the practice-assessment-feedback cycles performed during Weeks 1 to 9 is shown in Figure 1. After each practice, every student wrote

Table 1. Topics of the student practice that assessed with a rubric in Weeks 1 to 9.

Topic	Round	Teacher(s)[c]
Visual acuity 1[a]	1	A, G
Goldmann perimeter 1[b]	1	B, C, G
Contact lens	1	F, E, G
Visual acuity 2[a]	2	C, D, G
Humphrey field analyzer	2	B, G
Fundus photography	2	F, E, G
Near visual acuity	3	D
Goldmann perimeter 2[b]	3	B, E
Lens meter	3	C, G

a, b: In these topics, different contents are trained in 1 and 2.
c: Teachers A–E are certified orthoptists, F is an ophthalmologist, G is a PhD in Engineering.

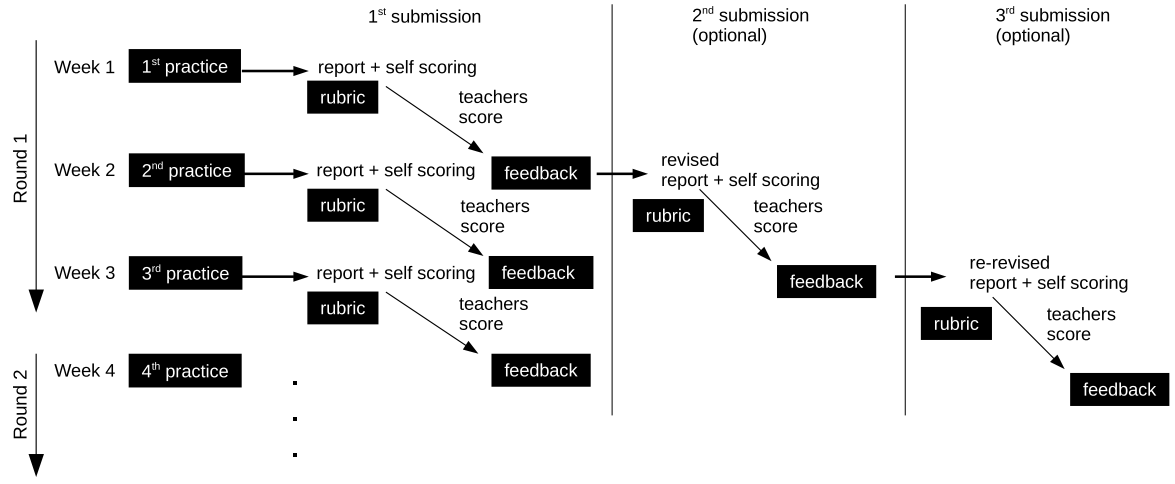

Figure 1. Schedule of practices, report writing, rubric scoring, and feedback from teachers who share the same rubric. See Table 1 for "Rounds." Rounds 2 and 3 are omitted in this scheme.

a report on what he or she had learned and scored his or her report using a rubric. Using the same rubric, the teacher scored each student's report and returned the scores to the student by the next week. Students could revise reports up to two times. The same rubric was used for the revised reports. Before, during, and after Weeks 1 to 9, all students received introductions, guidance, and course examinations (not shown). Regretfully, all data in Week 7 and data from one student in Week 9 were lacking.

3. Evaluation of Disagreement and Statistics

To quantify disagreements between the self- and teacher-evaluated scores for practice reports, we assigned scores 0, 1, and 2 to "poor," "not so good," and "good" (instead, for the sake of encouragement, we used scores 0, 1, and 3 for feedback to students) for every criterion of the rubric. We calculated the distance between the vectors for self- and teacher-evaluated scores in a five-dimensional Euclidean space. To pool distances, we calculated the root square sum of the

distances across the target criteria and students as:

$$SummedDistance = \sqrt{\sum_{criteria,student} (Self\ Score_{criteria,student} - Teacher\ Score_{criteria,student})^2}$$

where Self Score and Teacher Score reflects the scores submitted by students and teachers, respectively. The re-submission count for each student was simple summation of the number of re-submitted reports across the whole practice period. Statistical analyses were performed with GNU R (http://www.r-project.org/) on a Macintosh computer.

Results

Regretfully, all data generated during Week 7 and data from one student during Week 9 were missing and therefore not included in the final analyses.

Improvement in self-evaluation with repeated rubric-based assessment

The disparities between self- and teacher-evaluated rubric scores decreased over the study weeks. As shown in Figure 2, the total disparities (calculated across all five criteria and for all students) decreased in a significantly monotonic fashion ("All," $p = 0.031$, Kendall's Tau). This result suggests that repeating practice-assessment-feedback cycles improved students' self-evaluation. The improvement was most obvious in the "Report format" criterion ($p = 0.014$, Kendall's Tau) and in the "Data-based discussion" criterion, although the p-value did not reach the level of statistical significance ($p = 0.061$, Kendall's Tau). On the other hand, there was no obvious improvement in the "Writing style" and "Presentation and understanding of the task" criteria. The disagreements in the "Reference" criterion were small throughout Weeks 1 to 9.

Greater improvement in lower-performance students

Next, we investigated the relationship between teachers' evaluations and reductions in the disagreements between self- and teacher evaluations (Figure 3). In Week 1 (left), there was a strongly significant and negative correlation between disagreement and the teachers' scores ($p = 4.23 \times 10^{-6}$, Spearman's Rank-Order Correlation), suggesting students with lower reporting performance tended to over-evaluate their reports during the early stages of the practical course. Though the negative correlations between the disagreements and the teachers' scores were still significant by Weeks 5 and 9 (mid and right, $p = 0.048$ and 0.039 for Weeks 5 and 9, respectively, Spearman's Rank-Order Correlation), fewer

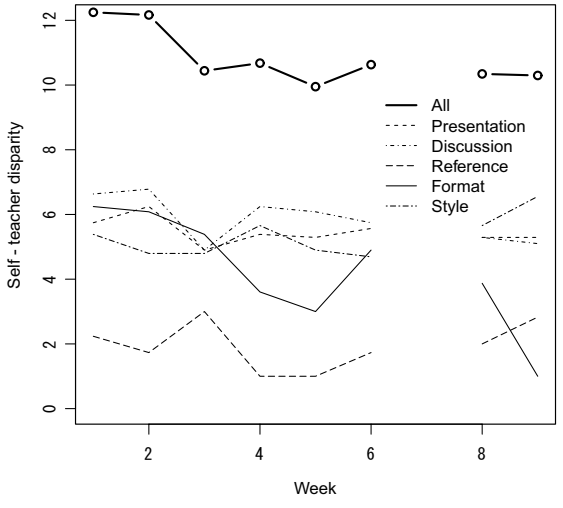

Figure 2. Time-courses of disparities between the self- and teachers' scores. The criterion/criteria used to summate these disparities is/are indicated by line widths and styles. All: distances are summed across all five criteria and all students. Presentation–Style: distances are summed across all students in the criteria of "Presentation and understanding of the task," "Data-based discussion," "Reference," "Report format," and "Writing style" (See Table 2).

Table 2. The rubric used in this study (a tentative English translation).

	good (3)[a]	not so good (1)[a]	poor (0)[a]
Presentation and understanding of the task	Purpose, Subjects, Methods, Results, and Discussion match well to each other. Understanding of the task is good.	Purpose, Subjects, Methods, Results, and Discussion match partially to each other. Understanding of the task is incomplete.	Purpose, Subjects, Methods, Results, and Discussion are not corresponding to each other. Understanding of the task is poor.
Data-based discussion	Interpretation of the data is appropriate. Descriptions are objective and logically ordered.	Interpretation of the data is partially inappropriate. Discussion refers the data, but incomplete.	Interpretation of the data are inappropriate. Discussion does not base on the data. Discussion is totally subjective.
References	References to academic contents are adequately made in accordance with JJOS[b] format.	References to academic resources are adequately made, but not in accordance with JJOS[b] format.	No references to academic resources. References to non-certificated Internet contents.
Report format	Carefully written as a scientific report.	Fonts are too large. Poor handwriting. The format partially deviates from scientific reports.	Not formatted as a scientific report. The length of the report is too short.
Writing style	Concise, concrete, academic wording.	Sentences tend to be long. Less concrete. A few colloquial words. A few typos.	Sentences are too long. Readability is not considered. Not concrete. Colloquial or prose. Three or more typos.

a: Numbers in the parentheses are scores being fed back to students.
b: The Journal of Japan Ophthalmological Society (Nippon Ganka Gakkai Zasshi).

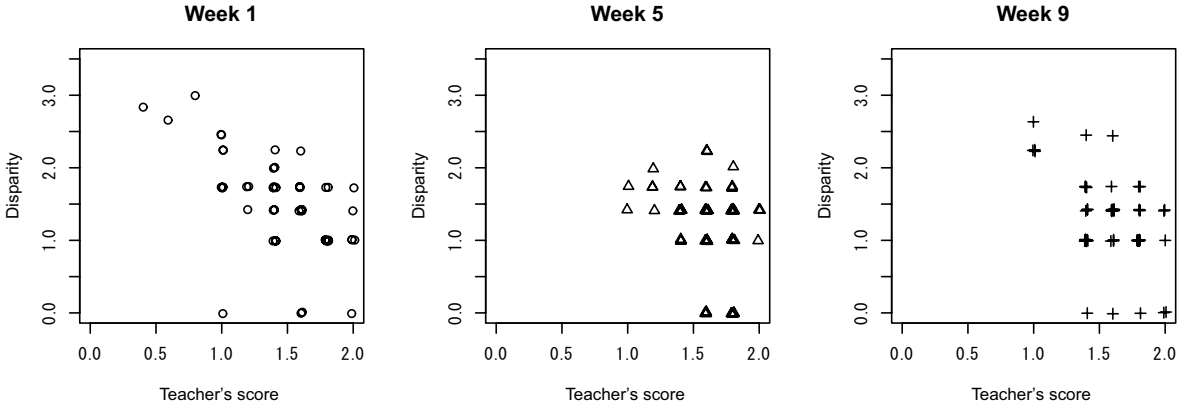

Figure 3. The disparities between self- and teachers' scores are plotted against teachers' scores in Weeks 1, 5, and 9, for the left, mid, and right panels, respectively. The disparities between scores for all criteria are summed. Each symbol represents a student. Jitter is added to each symbol position for display purposes.

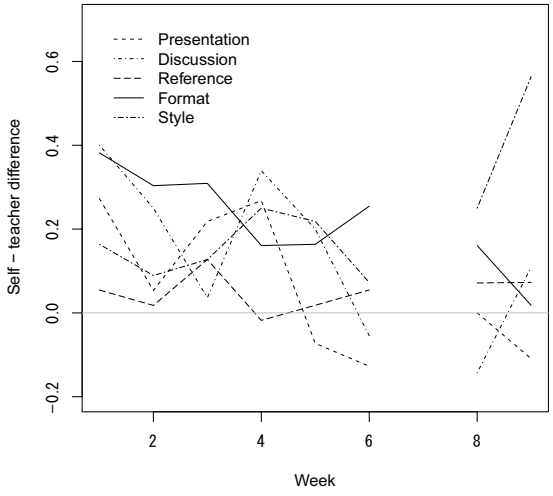

Figure 4. Average difference (*i.e.*, simple subtraction) between self- and teachers' scores in each criterion. The format is same as Figure 2.

students belonged to the upper-left quadrant. This result indicates that improvement of self-evaluation, as well as reports themselves, occurred mainly in students with lower reporting performance and higher self-evaluation scores. The interpretation was supported by two other results. Firstly, in Week 1, the self-evaluated scores were higher than teacher-evaluated scores in all criteria, whereas the self-evaluated scores approximated teacher-evaluated scores in Week 9 except for the "Writing style" criterion (Figure 4). Secondary, the relationships between the disagreement in Week 1 to those in later Weeks were weak; a significant correlation was observed only when all data from Weeks 4 to 9 were pooled (Figure 5).

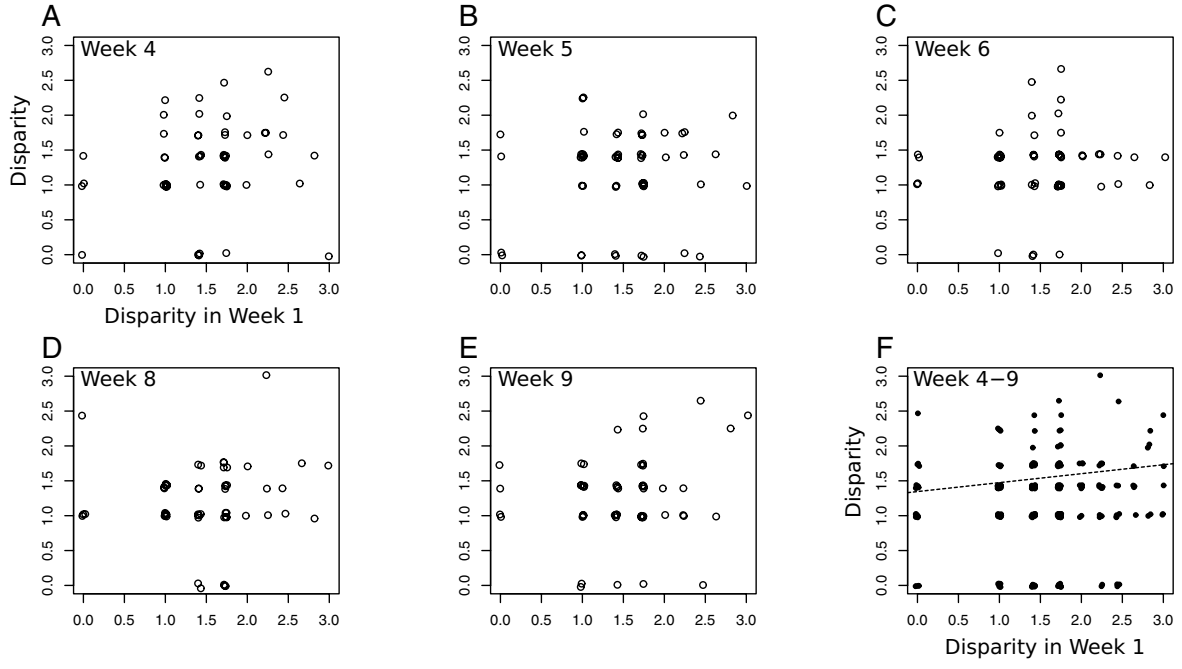

Figure 5.

A – E: The disparities between self- and teachers' scores in Weeks 4, 5, 6, 8, and 9 are plotted against those in Week 1, respectively. The distances for all criteria were summed. Each symbol represents a student. Jitter is added to each symbol position for display purposes.

F: same as A – E, except that the disparities in Weeks 4 to 9 are pooled. Dotted line: a regression line for the Week 1 data and the data pooled across Weeks 4 to 9 ($y = 0.148 x + 1.32$, rho = 0.13, $p = 0.030$, Spearman's Rank-Order Correlation).

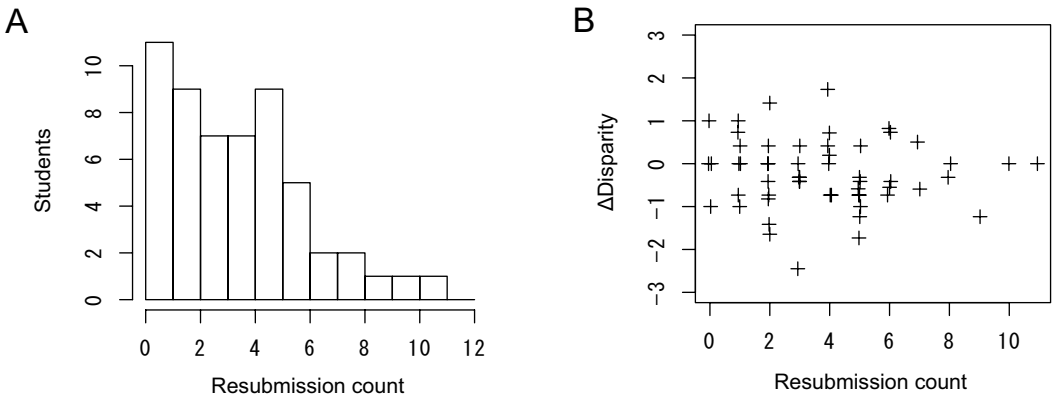

Figure 6.
A: A distribution of the re-submission counts.
B: Changes in disparities between self- and teacher- evaluated scores from Week 1 to Week 9 are plotted against the re-submission counts. Each symbol represents a student. Jitter is added to each symbol position for display purposes.

Effects of re-evaluation in the same topic

In this study, students were allowed to revise their reports up to twice and each re-submission was followed by feedback from a teacher according to the rubric (Figures 1 and 6A). If the improvements in self-evaluation were believed to result from repeated feedback *per se*, then observed improvements should correlate with each student's number of re-submission. To test the hypothesis, we plotted changes in the disparities between self- and teacher-evaluated scores against re-submission counts. As shown in Figure 6B, there was no significant correlation between these variables, indicating that re-evaluation on the same subject is not sufficient to improve self-evaluation using rubric assessments. Perhaps re-evaluation on the same subject lacked novel elements required for generalized learning of the report-writing skills.

Discussion

Adequate self-evaluation is one of the keys to successfully learning orthoptic skills. In this study, we found that students' self-evaluation of practice reports improved through rubric-based practice-assessment-feedback cycles (Figure 2), despite that the students were trained in the nine topics and scored by seven teachers (Table 1). Standardization among multiple assessments secondary to the concrete and explicit descriptions included in the rubric might assist in improving this complex practice. However, the teaching methods must be further refined because the improvements depended upon the criteria. For instance, although the improvements were found in the criteria related to the report format (*e.g.* "Report format"), the improvements were not obvious in choosing academic words ("Writing style"), nor understanding the task ("Presentation and understanding of the task"). Improvements in these criteria may require specific training for technical writing and lectures on the tasks.

Improvements of self-evaluation occurred mainly in students with lower performance (Figure 3). These students might become aware of their weak-points through the practice-assessment-feedback cycles. This result is favorable for our orthoptic education, because the purpose of our department

is to foster as many COs as possible, including students with lower performance. Supports for these students are key to achieving this objective. The students tended to over-evaluate their reports (Figure 3) in all criteria (Figure 4) at the beginning of present study, whereas the "Motivation and Attitude for Practice" criterion was under-evaluated by students in the previous study [11]. Unlike the present study, in which the students evaluated their reports, in the previous study the students evaluated their clinical performance. The discrepancy between results obtained during the previous and present studies might be due to differences in the subjects that were assessed.

A series of cohort studies with randomized control trials (RCTs) would be required to directly investigate the effect of the rubric-based assessments on students achievement in the clinical practice. However, it is difficult to employ RCTs in undergraduate clinical courses, in which equal educational opportunities must be highly considered. Instead, we are planning to compare educational efficiencies with students of another year, who were not conducted self-evaluation of practical reports with rubric-based assessments. In addition, we will examine whether there were savings-effects in learning-curves within each of advanced courses among the same students, who continued self-evaluation of practical reports with rubric-based assessments in these courses after the target period of this study.

To meet the OECD Key Competencies and the request of the Ministry of Education, Culture, Sports, Science and Technology for the evolution of higher education, use of rubric-based assessments are spreading in Japanese higher educational institutions [12]. For instance, Chuo University, one of the largest universities in Japan, introduced rubric-based assessments for education in 2011 [13]. In the Faculty of Science and Engineering of that university, which had precedingly introduced rubrics, reconstructed the practical courses through the efforts to extract keywords for the rubrics [14]. Since our department is much smaller in the student number and more specified for orthoptic education, we expect that improvement not only in report writing but also in the clinical competence will be achieved rather easily than in those diversified and large-sized faculty by further refining and expanding our use of rubric-based assessments.

Acknowledgements

We thank Ms. Rina Toyoshima and Ms. Yumiko Kawakami for secretary works.

Conflicts of interest

The authors have no potential conflicts of interest to disclose.

References

1. Article 17 of the Orthoptists Act. Available from: http://law.e-gov.go.jp/htmldata/S46/S46HO064.html (accessed July 27, 2017) (in Japanese)
2. Japanese Association of Certified Orthoptists. 2015 Survey of Certified Orthoptists; 2015. Available from: http://www.jaco.or.jp/pdf/top/2015survey.pdf (accessed November 7, 2017) (in Japanese)
3. Tabuchi A, Maeda F. School education for certified orthoptists. In: Maruo T, editor. Practical Ophthalmology 29. Tokyo: Bunko-do; 2009; 215-224. (in Japanese)
4. Wolf K, Stevens E. The Role of rubrics in advancing and assessing student learning. The Journal of Effective Teaching. 2007; 7: 3-14.
5. Goggins Selke MJ. Preface. In: Goggins Selke MJ. Rubric Assessment Goes to College. Rowman & Littlefirld Education, Lanham, UK. 2013.
6. Brookhart SM. Chapter 1. What are rubrics and why are they important? In: Brookhart SM. How to create and use rubrics for formative assessment and grading. Association for Supervision and Curriculum Development,

2013, ISBN: 1416615075.
7. Rubrics (explanations and examples) page in University of Florida. Available at http://citt.ufl.edu/tools/rubrics/ (accessed July 20, 2017)
8. Developing Rubrics for Assessment. Available at http://assess.pages.tcnj.edu/files/2011/06/Developing-Rubrics.pdf (accessed July 20, 2017)
9. Rubrics for Assessment. Available at http://www.uwstout.edu/soe/profdev/rubrics.cfm (accessed July 20, 2017)
10. Sample Rubrics. Available at http://course1.winona.edu/shatfield/air/rubrics.htm (accessed July 20, 2017)
11. Maeda F, Oka M, Yamashita T, Kobayashi Y, Tabuchi A. Assessment of an evaluative rubric for orthoptic students' performance in a clinical training. Jap Orthoptic J. 2012; 41: 229-234. (in Japanese)
12. Oki H. How to introduce rubrics into japanese universities: Aiming at fairness, objectivity and rigor in assessment of performance. Ritsumeikan Higher Educational Studies. 2014; 14: 71-90. (in Japanese)
13. Hirano H. Rubric-based Competency. Chuo Online. June 30, 2014. Available at http://www.yomiuri.co.jp/adv/chuo/dy/opinion/20140630.html (accessed July 20, 2017)
14. Makino M. Seeking Education for Assuring Knowledge and Ability through Competency Development —Human Resource Development in the Faculty of Science and Engineering: selected as a new Education GP by the MEXT—. Chuo Online. December 1, 2009. Available at http://www.yomiuri.co.jp/adv/chuo/dy/education/20091201.html (accessed December 12, 2017)