

Inter-Evaluator Consistency in Evaluation of Midwifery Students' Records in Clinical Practice for Antenatal Health Check-ups - Effectiveness of a Developed Rubric -

Takayo Yamada , Shizuka Takagi* , Sadanori Higashino** , Keiko Shimada*** , Kimikazu Sugimori****

Abstract

Background

The main concern of clinical practice in Japanese midwifery education is assisting in deliveries. The national guidelines for midwifery schools explicitly require delivery assistance, but with regard to antenatal care, the national guidelines require students to handle ongoing care from the second trimester through postnatal confinement in at least one case. Therefore, it has been left to the discretion of each midwifery school whether clinical practice includes antenatal health check-ups, and relatively little attention has been given to instruction or standards for evaluating midwifery students' records. Therefore, there is a need to develop reliable, objective evaluation methodology for evaluating antenatal health check-ups conducted by students in clinical practice.

Objectives

This study was performed to raise objectivity in the evaluation of midwifery students' records in antenatal health check-ups in clinical practice by (1) measuring disparities in agreement among evaluators and (2) developing a rubric for designing evaluation sheets to minimize disparities between multiple evaluators due to individual views.

Methods

1) Research term

January – August, 2016

2) Subjects

Six midwifery evaluators that had been in charge of clinical practice in performing antenatal health check-ups, including three from the 2-year midwifery program of University A and three from a graduate program at University B.

3) Instruments used for evaluation

(1) Evaluation tool developed based on the pregnancy care checklist in the “Nurse Midwife Training Guide” published by the Japanese Nursing Association; (2) a rubric we developed for evaluation of midwifery students' records from clinical practice in conducting antenatal health check-ups.

4) Student records used in the study

We examined records from antenatal health check-ups in 10 cases conducted by midwifery students who had completed clinical practice sessions of 2-year midwifery school.

5) Analysis method

Two assessment instruments were applied in each case of antenatal health check-ups by midwifery students. The six evaluators consisted three from A University and three from B University. Fleiss' kappa was used to compare the disparities in results between the two groups of evaluators.

Results

Agreement of evaluators' student evaluations with the sheet based on Japanese Nursing

Doctoral Course, Health Sciences, Division of Health Sciences, Graduate School of Medical Science, Kanazawa University

* School of Nursing, University of Shizuoka

** School of Management & Information, University of Shizuoka

*** Division of Health Sciences, Graduate School of Medical Science, Kanazawa University

**** Institute of Liberal Arts and Science, Faculty of Higher Education Research and Development, Kanazawa University

Association pregnancy care checklist: There was “poor agreement” in three cases (30%) in the University A group and in seven cases (70%) in the University B group. There were no cases of relatively high “moderate” agreement. There was “poor agreement” for 12 evaluation items in the University A group (75%) and for nine items in the University B group (56.3%). There was “slight agreement” for two items in the University A group (12.5%) and five in the University B group (31.3%) .

Agreement of evaluators’ student evaluations with the sheet based on our rubric: “Poor agreement” was observed in one case in the University A group (10%) and four cases in the University B group (40%) . Agreement was “slight” in five cases (50%) in the University A group and four cases (40%) in the University B group. “Moderate” agreement was observed for three cases (30%) in the University A group. There was “poor agreement” for two evaluation items in the University A group (25%) and four in the University B group (50%) . Agreement level was “slight” for five items in the University A group (62.5%) and three items (37.2%) in the University B group.

Conclusion

To assess evaluators’ objectivity in evaluating midwifery students’ records of antenatal health check-ups in clinical practice, this study examined levels of disparities among evaluations given by evaluators from two midwifery schools. Evaluators from University A showed a 30% rate of “poor agreement” in their evaluations of the same student records, while the rate was 70% for evaluators at University B. Disparity levels were reduced to 10% and 40%, respectively, when evaluation standards based on our rubric were used. These results suggest that introduction of the rubric as an instrument for evaluation of antenatal health check-ups could be an effective means of improving consistency among evaluators with regard to evaluation of student performance.

KEY WORDS

Midwifery education , Antenatal health check-ups , Clinical practice , Evaluate agreement , Rubric

Introduction

Due to a decrease number of in obstetricians and gynecologists in recent years, there is high expectation for midwives to conduct low-risk antenatal health check-ups on expectant mothers. Establishment of midwife outpatient clinics¹⁾ and enhancement of practical skills in maternity care in midwifery education are required. According to the “attainment level of technical competencies in midwife and nursing education at the time of graduation” as indicated by the Ministry of Health, Labor and Welfare in 2008, in most of the items regarding the diagnosis and care of the health condition of expectant mothers and their families, the rating is “independence possible with minor advice”, indicating a high level of achievement²⁾.

However, in midwifery education in Japan, the focus of clinical practice is placed on deliveries, and the lack of clinical practice for antenatal health checks has long been considered a problem³⁻⁴⁾. According to the national guidance for midwifery schools, “midwifery students must handle 1 or more cases for which continuous clinical practice is implemented from the second trimester of

pregnancy to 1 month post-delivery⁵⁾”. When compared to the designated “approximately 10 cases of normal delivery” in deliveries, the number of cases experienced remains low. Thus, it has been up to each midwifery school whether or not clinical practice includes antenatal health check-ups, with the result that relatively little attention has been given to instruction or to standards for evaluating midwifery students’ records.

There are multiple courses in midwifery education, namely an elective course, a 1-year course and a 2-year course. The Japan Society of Midwifery Education reported that regarding the practical skills in maternity care at the time of completion of midwifery education, “there were differences in the midwifery education course, and thus disparity existed in the attainment level at the time of graduation as set forth by the Ministry of Health, Labor and Welfare⁶⁾. In addition, based on the results of a survey conducted in 2012, the Japan Society of Midwifery Education reported that 60% of midwifery schools evaluated the clinical practice of midwives in antenatal health check-ups based on reports and students’ records⁷⁾.

Survey on the clinical practice records in maternity care will lead to assessment of antenatal health check-ups clinical practice, which in turn will contribute to the improvement in the practical skills of maternity care.

Regarding records made by midwives, Article 42, Clause 1 of the Act of Public Health Nurses, Midwives and Nurses states: "When midwives assist in childbirth, matters related to delivery must be recorded without delay"⁸⁾, so making the recording of maternity care is a midwives' duty. In addition, the "2nd Japan Obstetric Compensation System for Cerebral Palsy: Report on the Prevention of Recurrence" reported a lack of or inadequacies in the entry of medical records⁹⁾. The ability of midwives to records of the conditions of expectant mothers, the fetus and newborns, as well as "the plan for deliveries and accurate recording and assessment of implemented maternity care"¹⁰⁾ is a fundamental skill. For this reason, assessment of whether adequate recording of antenatal health check-ups is conducted from the perspective of midwifery education. It is also important whether there are disparities among multiple evaluators. Moreover, determining whether adequate assessment is conducted and whether disparity exists among multiple raters is vital from the aspect of standardization and fairness in the assessment of antenatal health check-ups clinical practice.

On the other hand, the problem of disparity in the assessment among the multiple raters was examined and reported on as the degree of agreement in the evaluation among raters when the Objective Structured Clinical Examination (OSCE) was introduced in medical education¹¹⁻¹²⁾, dental education¹³⁻¹⁵⁾ and pharmaceuticals¹⁶⁾. Ito reported that when there were two evaluators, there was concern for decline in objectivity due to variations in assessment¹⁵⁾.

Therefore, in this study, a comparison was made between the degree of agreement in the evaluation by the three evaluators at the A University with regard to the antenatal health check-ups clinical practice records by midwifery students studying in the graduate course of midwifery at the A University and the degree of agreement in the evaluation by three evaluators at the B University as an external assessment, in order to examine the degree of agreement among raters in midwifery education.

Additionally, aside from nursing clinical practice,

the evaluation of midwifery clinical practice including antenatal health check-ups clinical practice is a comprehensive learning assessment which includes knowledge, skills and attitude. Presently, rubric is regarded as being appropriate to assess comprehensive learning¹⁷⁻¹⁸⁾. In Japan, rubric was an evaluation instrument introduced in 2003 by Kawai, which was suited to objectively assess "comprehensive learning"¹⁹⁾. Rubric is defined as an "instrument which positions specific matters for which attainment is desirable with regard to a task", or a "tool which divides a certain task into several elements, explaining in detail the level which was required to attain the assessment criteria according to the element"¹⁹⁾. In medical education and nursing education, studies of educational assessment using rubric have been implemented. Hence, introduction thereof in midwifery education requires consideration.

The purpose of this study was set out to raise objectivity in the evaluation of midwifery students' records in antenatal health check-ups clinical practice by (1) measuring disparities in agreement among evaluators' evaluations and (2) developing a rubric as the basis for creating evaluation sheets that tend to minimize disparities between multiple evaluators' evaluations due to individual views.

Method

1. Research design: Quantitative comparative descriptive research

2. Subject

Six midwifery evaluators who had been in charge of clinical practice in performing antenatal health check-ups, including three from the two-year midwifery program of A University and three from a graduate program at B University.

3. Term and collection of data

1) Data collection

Data collection was carried out from January to August, 2016.

2) Rubric development period

The rubric was developed from April to July, 2016.

3) Instruments used for evaluation

(1) Evaluation tool created based on the basis of the pregnancy care checklist in the "Midwife Training guide" published by the Japanese Nursing Association; (2) a rubric we developed for evaluation of midwifery students'

records from clinical practice in conducting antenatal health check-ups.

(1) Evaluation tool created based on the basis of the pregnancy care checklist in the "Midwife Training guide"

Included in the evaluation scales regarding pregnancy care currently used was the checklist from <diagnosis of gestation period and care> under maternity care skills in the "Midwife Training guide" published by Japanese Nursing Association, "Technical Evaluation items in Midwifery Education and the Attainment Level at the Time of Graduation" set forth by the Ministry of Health, Labor and Welfare⁴⁾, and the "Core Curriculum in Midwifery Education" proposed by Japan Society of Midwifery Education.

As a record evaluation scale, the "Technical Items in Midwifery Education and the Attainment Level at the Time of Graduation" set forth by the Ministry of Health, Labor and Welfare and the "Core Curriculum in Midwifery Education" proposed by Japan Society of Midwifery Education had 9 evaluation items, respectively, with similar contents. In addition, the "Technical Items in Midwifery Education and the Attainment Level at the Time of Graduation" set forth by the Ministry of Health, Labor and Welfare is an evaluation scale for skills. Therefore, the 2 evaluation scales were excluded.

Moreover, although the checklist from <diagnosis of gestation period and care> under maternity care skills in the "Midwife Training guide" published by Japanese Nursing Association is designed for newly-graduated midwives, it was selected because of it being widely used. Table 1 is an evaluation scale created based on the checklist of pregnancy care in the "Midwife Training guide" published by the Japanese Nursing Association. The original check list had 20 evaluation items, we chose 16 evaluation items which were thought to be related to records were extracted and a 4-point Likert scale was used, consisting of "Highly competent (4 points)", "Competent (3 points)", "Not very competent (2 points)", and "Not competent (1 point)".

(2) A rubric we developed for evaluation of midwifery students' records

Stevens and Levi showed the basic 4 steps in rubric construction²⁰⁾. The 4 steps referred to reflection in the 1st step, creating a list in the 2nd step, grouping and indexing in the 3rd step, and creating a table in the 4th step. Reflection in the 1st step meant taking time to reflect

on "what was required on the part of the students and why this task was presented". Creating of a list in the 2nd step involved "focusing on the specific contents of the task and the learning course objective". Grouping and indexing in the 3rd step involved "summarizing the results of the reflection in the 1st and 2nd steps, grouping various expected results from the task, and lumping together each assessment viewpoint". The creating of a table in the 4th step meant "applying the assessment viewpoint and criteria obtained in the 3rd step into a final format of the rubric".

Steps 1 to 4 in preparing a rubric were conducted by 2 instructors. Both of them had guidance experience in antenatal health check-ups clinical practice more than 5 years. In addition, supervision was provided by midwifery educators with experience in creating a rubric and by instructors in pedagogy at the beginning and completion of the 1st and 2nd steps, when formulating an assessment viewpoint and criteria in the 3rd step, and when creating the final rubric format in the 4th step. They supervised about consistency, contractibility, appropriate weighting and grounds of rubric's technical requirements.

As the reflection in the 1st step, the assessment objects, task, clinical practice content and attainment level of the antenatal health check-ups clinical practice were confirmed, and a review and reflection were made of maternity care based on 2 types of midwifery textbooks.

In preparing the list in the 2nd step, records of antenatal health check-ups clinical practice were received from 6 midwifery students who are expected to complete the graduate course at the A University. From all of the antenatal health check-ups clinical practice records received, assessment viewpoint and criteria on their records, namely "how much of what needs to be entered" was extracted, and all necessary information were written on post-its.

In the grouping, indexing and rubric development in the 3rd step, the contents extracted from the clinical practice records in the 2nd step were grouped and indexed according to assessment viewpoints and criteria.

In the preparation of the final format of the rubric in the 4th step, supervision was provided for the completed rubric by midwifery educators and experts in pedagogy who are knowledgeable in the development of rubrics, after which multiple modifications were made.

4) Student records used in the study

Table 1. Evaluation scale for assessing midwifery students' records of antenatal health check-ups clinical practice

Content of Record about maternity care competency (behavioural objectives given in Japanese Nursing Association manual)			Highly competent	Competent	Not very competent	Not competent
			4 points	3 points	2 points	1 point
Diagnosis and Care During Pregnancy	1	Understands the woman's needs				
	2	Can perform a pregnancy health examination				
		(a) Can perform physical assessment (weight gain, blood pressure, NT scan, proteinuria and glycosuria)				
		(b) Can diagnose normal and abnormal progress of pregnancies				
	3	Can conduct a foetal health examination				
		(a) Can assess foetal growth and health				
	4	Can prepare a maternal care plan proposal				
	5	Can consult or provide health education based on a maternal care plan				
		(a) Can give a pregnancy consultation				
		(b) Can document high-risk factors for the development of maternal–infant attachment and child abuse and prepare appropriate responses				
	6	After pregnancy, can evaluate its effect				
	7	Can cooperate in caregiving and take an ongoing role				
	8	Understands birth plans				
	9	Understands the needs of breast feeding				
	10	Can facilitate breast feeding and conduct classes or manage consultations				
	11	Understands vital signs and test results				
	12	Understands the anatomical and physiological aspects of pregnancy				
	13	Understands the pathology and physiology of basic perinatal disorders; knows the tests and can recognise their signs by observation				
		(threatened miscarriage, hyperemesis, pregnancy-induced hypertension, gestational diabetes, placenta previa and multiple gestation pregnancy)				
	14	Can read a fetal monitor				

Evaluation scale for assessing midwifery students' records of antenatal health check-ups clinical practice, created using excerpts from Japanese Nursing Association manual for new midwives.

We examined records from antenatal health check-ups in ten cases conducted by midwifery students who had completed clinical practice sessions of two-year midwifery schools.

5) Recruit procedure for research participants

Requests for research cooperation were made to three evaluators from the A University and three evaluators from the B University who were research subjects, and to midwifery students who provided evaluation objective and records for rubric development. Requests made toward research subjects consisted of researchers or joint researchers briefing the research subject evaluators regarding the research objectives both orally and in writing and receiving consent therefrom.

Regarding midwifery students who provided evaluation objective and records for rubric preparation, approval was obtained from the supervisor of the midwifery course before briefing them. Of the seven midwifery students who received both oral and written explanations of the research objective by the researchers, six students who consented provided antenatal health check-ups clinical practice records.

4. Analysis method

With two assessment instruments were applied in each case of antenatal health check-ups by midwifery students. The six evaluators conducted three evaluators from A University and three evaluators from B University, and Fleiss' kappa was used to compare the disparities of evaluation results from the two evaluators groups. In addition, the average score, standard deviation and Fleiss' kappa values for each of the 10 records of the expectant mothers were calculated. Our data were analyzed using the National Council for the Social Studies (NCSS) statistical analysis software, version R2.8.1. To measure agreement among the evaluators and within each group, we calculated Fleiss' Kappa values for each student record. The Kappa statistics is a good indicator of agreement for results from two or more evaluators measuring two or more subjects or taking two or more measurements, and it is applicable to data on either ordinal or nominal scales²¹⁾. Cohen' was applied to data from two evaluators, and Fleiss' kappa was applied to data from three or more evaluators. We examined the agreement among evaluators as indicated by the statistics.

The interpretation of Fleiss' kappa values is $-1 \leq \text{kappa} \leq 1$, and the interpretation thereof was the standard for

Feinstein. In other words, $-1 - 0$ indicates poor agreement, $0 - 0.20$ indicates slight agreement, $0.21 - 0.40$ means fair agreement, $0.41 - 0.60$ means moderate agreement, $0.61 - 0.80$ means substantial agreement, and $0.81 - 1.00$ means almost perfect agreement²²⁻²³⁾.

Evaluation items 10, 11, 12 and 16 which were extracted from assessment tool (1), <diagnosis of gestation period and care> under maternity care skills in the "Midwife training guide" were considered evaluation items in the third trimester, and six cases in the latter half of pregnancy were selected as analysis objective, excluding the clinical practice records of the first and second trimester from the analysis.

5. Ethical considerations

At the beginning of our study, ethical approval was obtained from the appropriate committees of participating universities whose students provided clinical practice records and midwifery institutions to which the evaluators belonged. Permission to conduct this study was obtained from the Ethics Committee of Kanazawa University Graduate School Of Medical Science (Review No. 653-1) and University of Shizuoka Research Ethics Committee (Receipt No. 28-1). After conveying the research objectives to the research subjects, it was explained orally and in writing that anonymity would be maintained so that individuals could not be identified from the assessment results of the records, data entry would be carried out in an unlikable manner, consideration would be given to personal privacy, cooperation toward the survey would be based on free will, participation might be withdrawn during the course of the survey, no disadvantage for the research subject would be incurred due to participation or non-participation in the survey, cooperation might be determined based on the free will of the research subject themselves, cooperation with the survey might be called off at any time during the course of the survey, and the results would not be used for any other purpose than for academic purposes. Evaluators consented to the above conditions signed the consent form.

To midwifery students who provided the records which would become assessment objective, after conveying the research objective, it was explained both orally and in writing that the academic evaluation for the antenatal health check-ups clinical practice had already been determined and that it would not affect the grades in their midwifery clinical practice course, records were made

anonymous so that individuals could not be identified, data entry of the assessment results would be carried out in an unlinkable manner, consideration would be given to personal privacy, cooperation toward the survey would be based on free will, and participation might be withdrawn during the course of the survey, no disadvantage for the research subject would be incurred due to participation or non-participation in the survey, cooperation might be determined based on the free will of the research subject themselves, cooperation with the survey might be called off at any time during the course of the survey, and the results would not be used for any other purpose than for academic purposes. Students who consented to the above conditions signed the consent form.

Results

1. Student records used in the study: There were 54 antenatal health check-ups clinical practice records of expectant mothers which were provided by midwifery students at the A University Graduate School of nursing. Of these, there were 10 records of expectant mothers who were examined twice or more times in the clinical practice and whose records became evaluation objective by the evaluators.

The antenatal health check-ups clinical practice records which became evaluation objective were as follows: 2 in the first trimester (7-11 weeks, 8-12 weeks), 2 in the second trimester (2 in 25-28 weeks), 6 in the third trimester (28-30 weeks, 32-36 weeks, 34-38 weeks, 35-

36 weeks, 35-38 weeks, 36-37 weeks), with a total of 10 subjects. Additionally, in step 2 of rubric construction, clinical practice records for all 54 expectant mothers were used for list creation.

2. Evaluators who conducted assessments

The numbers of years of midwifery experience the evaluators who conducted the assessments had ranged from 4 to 31 years, with an average of 13.3 years.

3. Agreement of evaluators' student evaluations with the evaluation sheet based on Japanese Nursing Association pregnancy care checklist: Table 2 indicates the assessment results by the A University evaluators group and B University evaluators group according to the clinical practice records based on "Midwife training guide" published by Japanese Nursing Association. When examining the kappa values indicating the agreement among the raters according to the clinical practice record, the item which showed the highest kappa values among the A University instructor group was "slight agreement" for clinical practice record No. 10 with 0.20. The lowest kappa value was "poor agreement" for No. 3 with -0.22. In the clinical practice records assessed by the A University instructor group, in addition to record No. 10, "slight agreement" was seen in 7 records (70%), namely No.2, No.4, No.5, No.6, No.7 and No.9. There were 3 clinical practice records (30%) for which there was "poor agreement", and kappa values of 0 or below, namely No.1 and No.8, in addition to No.3.

The highest kappa values observed among the B

Table 2. Agreement of evaluations by midwifery students' records when using the evaluation scale for assessing midwifery students' records

The number of students' record	Week of Pregnancy	A university evaluators Group					B university evaluators Group				
		Mean ^a	SD ^b	K	Agreement	p	Mean ^a	SD ^b	K	Agreement	p
1	7-11	1.83	0.70	-0.09	Poor agreement	0.74	2.06	0.71	-0.11	Poor agreement	0.80
2	8-12	2.28	0.74	0.06	Slight agreement	0.32	2.50	0.56	-0.22	Poor agreement	0.91
3	25-28	2.58	0.87	-0.22	Poor agreement	0.97	2.86	0.43	0.03	Slight agreement	0.47
4	25-28	2.36	0.64	0.18	Slight agreement	0.12	2.64	0.54	-0.02	Poor agreement	0.55
5	28-30	2.14	0.76	0.18	Slight agreement	0.07	2.50	0.56	-0.17	Poor agreement	0.85
6	32-36	1.96	0.81	0.16	Slight agreement	0.06	2.35	0.77	0.23	Fair agreement	0.06
7	34-38	2.13	0.76	0.04	Slight agreement	0.37	2.69	0.55	-0.01	Poor agreement	0.51
8	35-36	2.13	0.90	-0.19	Poor agreement	0.93	2.48	0.65	-0.02	Poor agreement	0.56
9	35-38	2.73	0.85	0.06	Slight agreement	0.40	2.86	0.68	-0.11	Poor agreement	0.64
10	36-37	2.45	0.80	0.20	Slight agreement	0.13	2.67	0.56	0.14	Slight agreement	0.24

^a Mean; ^b SD (standard deviation); K = Fleiss' kappa; p = p-value (probability)

University instructor group was record No. 6 with “fair agreement” at 0.23. The lowest kappa value was clinical practice record No. 2 with “poor agreement” at -0.22. The only record with “fair agreement” among the B University instructor group was No.6. 2 clinical practice records (20%), namely No.3 and No. 10 were assessed as having “slight agreement”. In addition to No.2, 7 records (70%) which indicated “poor agreement” with kappa values of 0 or less were No. 1, No.4, No.5, No.6, No.7, No.8 and No.9.

Records in which “slight agreement” was reached among both instructor groups of the 2 universities was No.10, and there were 2 records in which both groups indicated “poor agreement”, namely No.1 and No.8.

Table 3 shows the assessment results of the A University instructor group and the B University instructor group according to the evaluation item based on the “Midwife training guide” published by the Japanese Nursing Association. The evaluation item which had the

highest kappa values among the A University instructor group was “Understands the needs of breast feeding”, with “Fair agreement” and a score of 0.33. The item with the lowest kappa values was “Can read a fetal monitor” with “poor agreement” and a score of -0.33. There were 2 evaluation items (12.5%) in which “fair agreement” were reached concerning the clinical practice records assessed by the A University instructor group, namely “Understands the needs of breast feeding” and “Understands the woman’s needs”. “Slight agreement” was reached for 2 evaluation items (12.5%), which were “Understand birth plan” and “Understands vital signs and test results”. The other 12 evaluation items (75%) indicated “poor agreement” with kappa values of 0 or less.

The evaluation item which had the highest kappa values among the B University instructor group were “Understands the woman’s needs” and “Understands the needs of breast feeding” with “fair agreement” and a score

Table 3. Agreement of evaluations by evaluation items when using the evaluation scale for assessing midwifery students' records

Evaluation items	A university evaluators Group					B university evaluators Group				
	Mean ^a	SD ^b	K	Agreement	p	Mean ^a	SD ^b	K	Agreement	p
1 Understands the woman's needs	2.33	0.61	0.21	Fair agreement	0.11	2.76	0.44	0.23	Fair agreement	0.21
2 Can perform physical assessment	2.73	0.64	-0.15	Poor agreement	0.77	2.77	0.57	-0.03	Poor agreement	0.55
3 Can diagnose normal and abnormal progress of pregnancies	2.77	0.63	-0.03	Poor agreement	0.54	2.72	0.53	0.05	Slight agreement	0.42
4 Can assess foetal growth and health	2.67	0.55	-0.14	Poor agreement	0.72	2.83	0.38	0.04	Slight agreement	0.46
5 Can prepare a maternal care plan proposal	2.21	0.62	-0.08	Poor agreement	0.61	2.73	0.45	0.15	Slight agreement	0.29
6 Can give pregnancy consultation	1.97	0.78	-0.19	Poor agreement	0.91	2.67	0.48	-0.20	Poor agreement	0.81
7 Can document high-risk factors for development of maternal–infant attachment and child abuse and prepare appropriate responses	1.72	0.75	-0.08	Poor agreement	0.70	2.00	0.74	-0.30	Poor agreement	0.98
8 After giving pregnancy care, can evaluate its effect	1.67	0.80	-0.16	Poor agreement	0.84	2.43	0.63	0.16	Slight agreement	0.17
9 Can cooperate in caregiving and take a role in ongoing caregiving	1.93	0.91	-0.13	Poor agreement	0.84	2.67	0.55	-0.14	Poor agreement	0.72
10 Understands birth plans	2.33	0.82	0.12	Slight agreement	0.30	2.73	0.59	-0.18	Poor agreement	0.65
11 Understands the needs of breast feeding	2.33	1.05	0.33	Fair agreement	0.08	2.14	0.86	0.23	Fair agreement	0.14
12 Can facilitate breast feeding and conduct classes or give consultation on breast feeding	1.67	0.98	-0.04	Poor agreement	0.56	2.21	0.89	0.07	Slight agreement	0.39
13 Understands vital signs and test result values	2.73	0.58	0.12	Slight agreement	0.31	2.76	0.64	-0.01	Poor agreement	0.51
14 Understands the anatomical and physiological aspects of pregnancy	2.27	0.83	-0.15	Poor agreement	0.87	2.53	0.57	-0.07	Poor agreement	0.65
15 Understands the pathology and physiology of basic perinatal disorders; knows the tests and can recognise their signs by observation	2.55	0.63	-0.11	Poor agreement	0.73	2.52	0.57	-0.18	Poor agreement	0.84
16 Can read a fetal monitor	1.58	0.90	-0.33	Poor agreement	0.79	1.77	0.93	-0.21	Poor agreement	0.79

^a Mean; ^b SD (standard deviation); K = Fleiss' kappa; p = p-value (probability)

of 0.23. The evaluation item with the lowest kappa values was “Can read a fetal monitor” with “poor agreement” and a score of -0.21. There were 5 evaluation items (31.3%) in which “slight agreement” were “Can diagnose normal and abnormal progress of pregnancies”, “Can assess foetal growth and health”, “Can prepare a maternal care plan proposal”, “After giving pregnancy care, can evaluate its effect”, and “Can facilitate breast feeding and conduct classes or give consultation on breast feeding” and 5 evaluation items which indicated. The other 9 evaluation items had kappa values of 0 or less, indicating “poor agreement”.

4. Rubric evaluation table

The rubric that was developed into an evaluation table consisted of 8 elements of evaluation and 4-step assessment criteria (Table 4). The evaluation items were: “Clinical examination essentials”, “Getting the right information about the woman's stage of pregnancy”, “Information acquisition (communication skills)”, “Pregnancy diagnosis (evidence, scale, basis for diagnosis and use of information sources)”, “Implementing a care plan appropriate to a diagnosis”, “Midwifery practice: drawing up and revising

a care plan”, “Appropriateness of records (ethical, objectivity, terminology, clarity and wording), Reporting, communicating information and discussing issues (contact with midwives, doctors and instructors) ”.

5. Agreement of evaluators' student evaluations with the evaluation sheet based our rubric: Table 5 shows the assessment results according to the clinical practice record based on the rubric of the A University instructor group and B University instructor group. Examining the kappa values indicating the degree of agreement among the raters according to the clinical practice records in the A University instructor group, the highest kappa values was seen in record No.5 with “moderate agreement” and a score of 0.67. The lowest kappa value was seen in record No. 7 with “poor agreement” and a score of -0.77. Of the clinical practice records assessed by the A University instructor group, “moderate agreement” was observed in 3 records (30%), namely No.3, No. 5 and No.6. “Fair agreement” was observed in 1 clinical practice record (10%), which was No.4. “Slight agreement” was seen in 5 records (50%), which were No.1, No.2, No.8, No.9 and No.10. Only No.7 showed “poor agreement” (10%) with

Table 4. Rubric assessment sheet

Aims: Competence to conduct health examinations and provide midwifery care from early pregnancy through the postnatal period.
Practice goals: Ability to conduct antenatal health examination independently.

	Elements of examination	A (4 points)	B (3 points)	C (2 points)	D (1 point)
1	Clinical examination essentials	Adequate examination, conducted appropriately for both the objectives of the examination and stage of pregnancy	Some inadequacy but basically conducted appropriately for the objectives and stage of pregnancy	Examination is procedurally inadequate for the objectives and stage of pregnancy	Student does not understand how to examine appropriately based on the objectives and stage of pregnancy
2	Getting the right information about the woman's stage of pregnancy	Information obtained about the stage of pregnancy is almost adequate and systematically organised; selection and ordering of priorities accurately reflect the urgency level	Information obtained about the stage of pregnancy is almost adequate and systematically organised; selection and ordering of priorities mostly reflect the urgency level	Information obtained about the stage of pregnancy is inadequate and not systematically organised; selection and ordering of priorities do not reflect the urgency level	Necessary information for the stage of pregnancy is not selected or obtained
3	Information acquisition (communication skill)	Has a good grasp of subjective information on aspects of a woman's living situation that will affect the pregnancy; information on the woman's preferences is adequate for use in assessment and caregiving	Has received subjective information from the woman that can be used in assessment and caregiving	Has received subjective information from the woman, but it is inadequate for use in assessment and caregiving	Has not received basic subjective information from the woman, leaving a major gap in available information
4	Pregnancy diagnosis (evidence, scale, basis for diagnosis and use of information sources)	Has assessed maternal health and foetal development, made use of evidence in a clinical interpretation and given a diagnosis	Has assessed maternal health and foetal development, made a clinical interpretation and given a diagnosis, but use of evidence is limited	Considerable insufficiency of evidence use in assessing maternal health and foetal development, making a clinical interpretation and giving a diagnosis	There were errors in assessing maternal health and foetal development, making a clinical interpretation and giving a diagnosis
5	Implementing a care plan appropriate to a diagnosis	Implementation of a care plan based on diagnosis was adequate and consistent	Implementation of a care plan based on diagnosis was partially inadequate and showed some inconsistency	Implementation of a care plan based on diagnosis was very inadequate and inconsistent	There was either no care plan based on diagnosis or it was not in written form
6	Midwifery practice: drawing up and revising a care plan	Adequate evaluation and revision of a care plan suited to the progress of the pregnancy and changes in the mother; the anticipated progress of the pregnancy has been forecast	Evaluation and revision of a care plan suited to the progress of the pregnancy and changes in the mother has been conducted	A care plan suited to the progress of the pregnancy has been evaluated but not revised based on changes in the mother's status	There has been no evaluation or revision of a care plan suited to the progress of the pregnancy or changes in the mother's status
7	Appropriateness of records (ethical, objectivity, terminology, clarity and wording)	No problems with ethics, objectivity or terminology; recording of examinations/care is accurate, complete and understandable	No problems with ethics, objectivity or terminology; there is a record of examinations/care	Problems with ethics, objectivity and/or terminology; some examination/care records are not understandable	No ethical consideration and/or objectivity; errors in the use of terminology; no documentation about examinations/care
8	Reporting, communicating information and discussing issues (contact with midwives, doctors and instructors) *	Communicates information clearly to the concerned parties, verbally or in written form, and shows full understanding of record-keeping responsibilities	Communication in verbal and written forms is satisfactory	Communication in verbal and written forms is partially unsatisfactory	No verbal or written communication to rate

*We consider it necessary that caregiving provided during clinical practice, and information obtained thereby, should be communicated to the midwives, doctors and instructors on duty in verbal and written forms.

kappa values of 0 or below.

The highest kappa values manifested among the B University instructor group was record No. 5 with “fair agreement” and a score of 0.36. The lowest kappa value was record No. 9 with “poor agreement” and a score of -0.17. Including record No. 5, the clinical practice records with a “fair agreement” assessment among the B University instructor group totaled 4 (40%), namely No.1, No.3, No.5 and No.6 and 2 had “slight agreement” (20%), which were records No.7 and No.8. “Poor agreement” in

which the kappa values was 0 or below were seen in 4 records (40%) in addition to No.2, namely No.4, No.9 and No.10. There were 5 clinical practice records in which there were “slight agreement” or higher among the instructor groups from both universities which were No.1, No.3, No.5, No.6 and No.8. There was no record with “poor agreement” in both of the groups.

Table 6 shows the assessment results according to the evaluation items by the rubric of the A University instructor group and the B University instructor group.

Table 5. Agreement of evaluations by midwifery students' records when using the newly developed rubric

The number of students' record	Week of Pregnancy	A university evaluators Group					B university evaluators Group				
		Mean ^a	SD ^b	K	Agreement	p	Mean ^a	SD ^b	K	Agreement	p
1	7–11	2.13	0.74	0.02	Slight agreement	0.46	2.58	0.50	0.31	Fair agreement	0.07
2	8–12	2.29	0.69	0.17	Slight agreement	0.17	2.38	0.49	-0.07	Poor agreement	0.61
3	25–28	2.67	0.70	0.50	Moderate agreement	0.02	2.75	0.53	0.29	Fair agreement	0.12
4	25–28	2.33	0.70	0.31	Fair agreement	0.04	2.58	0.50	-0.03	Poor agreement	0.55
5	28–30	2.00	0.59	0.67	Moderate agreement	0.00	2.67	0.56	0.36	Fair agreement	0.05
6	32–36	2.13	0.61	0.53	Moderate agreement	0.01	2.75	0.53	0.29	Fair agreement	0.12
7	34–38	2.50	0.72	-0.11	Poor agreement	0.73	3.04	0.46	0.05	Slight agreement	0.45
8	35–36	2.13	0.68	0.02	Slight agreement	0.46	2.67	0.48	0.06	Slight agreement	0.40
9	35–38	2.88	0.90	0.16	Slight agreement	0.21	3.17	0.76	-0.17	Poor agreement	0.86
10	36–37	2.50	0.72	0.06	Slight agreement	0.40	3.25	0.74	0.00	Poor agreement	0.50

^a Mean; ^b SD (standard deviation); K = Fleiss' kappa; p = p-value (probability)

Table 6. Agreement of evaluations by items when using the newly developed rubric

Rubric evaluation items	A university evaluators Group					B university evaluators Group				
	Mean ^a	SD ^b	K	Agreement	p	Mean ^a	SD ^b	K	Agreement	p
1 Gets the essentials that meet examination objectives	2.93	0.52	-0.02	Poor agreement	0.53	3.27	0.52	0.28	Fair agreement	0.10
2 Gets the information appropriate to the stage of pregnancy	2.33	0.66	0.05	Slight agreement	0.39	2.63	0.61	0.04	Slight agreement	0.40
3 Communication skills for getting information from woman	2.37	0.67	0.07	Slight agreement	0.33	2.90	0.61	0.11	Slight agreement	0.29
4 Diagnosis of pregnancy progress (acquires evidence, applies scale and uses diagnostic standards and references)	2.53	0.57	0.06	Slight agreement	0.37	2.93	0.52	-0.02	Poor agreement	0.53
5 Makes and implements a care plan appropriate to the diagnosis	2.50	0.51	0.20	Slight agreement	0.14	2.87	0.51	-0.04	Poor agreement	0.56
6 Evaluation of midwifery practice, planning and record-keeping	2.47	0.57	0.03	Slight agreement	0.43	2.70	0.53	0.18	Slight agreement	0.20
7 Appropriateness of records (logic, objectivity, terminology, clarity and expressive ability)	2.60	0.56	-0.01	Poor agreement	0.53	2.67	0.61	-0.03	Poor agreement	0.56
8 Reporting, communicating information and discussing issues (the student's mutual contacts with midwives, doctors and instructors)	1.10	0.31	0.26	Fair agreement	0.31	2.30	0.60	-0.05	Poor agreement	0.57

^a Mean; ^b SD (standard deviation); K = Fleiss' kappa; p = p-value (probability)

The highest kappa values shown among the A University instructor group was "Reporting, communicating information and discussing issues" with "fair agreement" and a score of 0.26. The evaluation item with the lowest kappa values was "Gets the essentials that meet examination objectives", with "poor agreement" and a score of -0.02. There were 5 evaluation items for which "slight agreement" were observed among the A University instructor group according to the records assessed, namely "Gets the information appropriate to the stage of pregnancy", "Communication skills for getting information from woman", "Diagnosis of pregnancy progress", "Makes and implements a care plan appropriate to the diagnosis", and "Evaluation of midwifery practice, planning and record-keeping". "Poor agreement" was observed for 2 evaluation items, which were "Gets the essentials that meet examination objectives" and "Appropriateness of records".

The evaluation item with the highest kappa value among the B University instructor group was "Gets the essentials that meet examination objectives" with "fair agreement" and a score of 0.28. The evaluation item with the lowest kappa values was "Reporting, communicating information and discussing issues" with "poor agreement" and a score of -0.05. Among the clinical practice records rated by the B University instructor group with an assessment of "fair agreement" was just "Gets the essentials that meet examination objectives". There were 3 evaluation items for which there was "slight agreement", namely "Gets the information appropriate to the stage of pregnancy", "Communication skills for getting information from woman", and "Evaluation of midwifery practice, planning and record-keeping". The 4 evaluation items for which there was "poor agreement" were "Diagnosis of pregnancy progress", "Makes and implements a care plan appropriate to the diagnosis", "Appropriateness of records" and "Reporting, communicating information and discussing issues".

The 3 evaluation items for which both university instructor groups reached "slight agreement" were "Gets the information appropriate to the stage of pregnancy", "Communication skills for getting information from woman" and "Evaluation of midwifery practice, planning and record-keeping". The evaluation item in which both university instructor groups indicated "poor agreement" was "Appropriateness of records".

Discussion

The results of this study was examined from 3 aspects, i.e., the degree of agreement among the 3 raters regarding the clinical practice records, the degree of agreement among the 3 raters regarding the evaluation items and the development of the rubric.

1. Degree of agreement among the 3 raters regarding the clinical practice records

As to the degree of agreement regarding the clinical practice records, it had been noted that assessment among the instructors towards the same records were not uniform, that there was "poor agreement" in the clinical practice assessment of the students, resulting in the risk of disparity in the course of providing clinical practice guidance. According to the research results as discussed above, the degree of agreement among the 3 raters based on the evaluation scale which was excerpted from <diagnosis of gestation period and care> under maternity care skills in the "Midwives Training guide", there were "poor agreements" for 3 clinical practice (30%) records in the A University instructor group and 7 (70%) in the B University instructor group.

Regarding clinical practice assessment, Tanner stated that "Nursing clinical practice evaluation is more of a subjective process than an objective one"²⁴. Oermann also explained that in the course of clinical practice evaluation, the instructor was required to confirm his/her own tendency, values, attitude and beliefs²⁵. From the results of this study, it was discerned that assessment of the clinical practice records differs according to the subjectivity of the instructors.

It had been reported that even in OSCE, the degree of agreement differed in a "medical interview" setting¹¹⁻¹⁵. The assessment of a "medical interview" involves 2 preliminary meetings where confirmation regarding evaluation items with regard to skills was made. Although the actual "medical interview" was approximately 5 minutes, there was still disparity in the assessment among the raters, which had been considered a problem. From the results of this study, it was thought that the assessment criteria according to the subjectivity of the instructors had a major impact on the assessment results.

On the other hand, in the case of a rubric, perhaps due to the detailed divisions in the assessment criteria and items, the impact of subjectivity on the part of the

raters was kept to a minimum. According to a report by Japan Society of Midwifery Education, clinical practice for maternity care was implemented at a 2-year midwifery education institution. Regarding instructors at universities, Kabeyama indicated that in vocational schools, there was a standard targeting the students based on the guidelines for full-time instructor clinical practice course, thus achieving equalization in the quality of nursing education. However, with instructors at universities, there was no restrictions and the recruiting of instructors was entrusted to educational institutions. This result in fluctuations in the minimum clinical capabilities required as instructors and basic knowledge in order to impart education²⁶⁾. Therefore, the results revealed that fluctuations were observed even in the assessment of maternity care which is carried out in a 2-year course depending on the instructor.

2. Degree of agreement in the assessment among raters regarding evaluation items

In this study, not only the degree of agreement in the assessment of the clinical practice records but also the degree of agreement in evaluation items was also considered. As a result, with regard also to the evaluation items, the degree of agreement with regard to the clinical practice records among the 3 raters based on the evaluation scale developed from the <diagnosis of gestation period and care> under maternity care skills in the “Midwives Training guide” indicated “poor agreement” in most of the evaluation items among both the A University instructor group and B University instructor group.

However, regarding the 2 evaluation items, namely “Understands the woman’s needs” and “Understands the needs of breast feeding”, both the A University instructor group and B University instructor group had the same result of “fair agreement”. Concerning these 2 evaluation items, it is surmised that regardless of the educational institution, there was a measure of uniformity in the standard of thinking of instructors engaged in midwifery education. Regarding the other evaluation items, there may be a possibility that the assessment varied according to the weeks of pregnancy care.

In the degree of agreement for the evaluation items according to the rubric, “fair agreement” was indicated for “Reporting, communicating information and discussing issues” among the A University instructor group.

This was thought to be attributed to the fact that the evaluation item was not extracted from the student clinical practice record, but rather created based on advice from a supervisor. Thus, it was not recorded in the midwifery clinical practice record, resulting in a higher degree of agreement. However, as proposed in the Obstetric Compensation System report for Cerebral Palsy⁹⁾, when implementing maternity care henceforth, “Reporting, communicating information and discussing issues” must be recorded in the midwifery records, highlighting the importance of this item.

Furthermore, there was disparity in the degree of agreement among the A University instructor group and the B University instructor group according to the evaluation item. Regarding items with a low degree of agreement, the disparity in the assessment would serve as a useful reference for discussion.

3. Development of a rubric

The preparation of a rubric in the course of this study was thought to have led to a measure of improvement in the degree of agreement among the raters. However, in implementing guidance for midwifery students using a rubric in the actual antenatal health check-ups clinical practice, the task remained of consistency in assessment with the instructors, including whether the expressions were easily understandable for the students. Moreover, rubrics were used in simulations in nursing education, and rubrics which were used in combination with checklists were also being developed²⁷⁾. Oermann stated that “nursing clinical practice and simulation are 2 different things”²⁵⁾. However, it was thought that in order to attain a higher degree of agreement, modification of the rubric, it was used in combination with a checklist, and holding meetings using a rubric to achieve agreement in assessment among instructors are necessary.

Limitations of research

This study revealed that the instructors of the A University who were affiliated with the Graduate School of Nursing who provided the clinical practice records and the instructors of the B University had differing educational objectives with regard to antenatal health check-ups clinical practice. Subsequently, the differences thereof might have affected the assessment results. In addition, the number of survey subjects was limited. Therefore, in order to enhance the objectivity of the

assessment in antenatal health check-ups clinical practice in the future, it was necessary for clinical practice to be conducted in other educational institutions, and further studies with consideration to the term of clinical practice and generalization were required.

Conclusion

In order to assess evaluators' objectivity in evaluating midwifery students' records of antenatal health check-ups in clinical practice, the study examined levels of disparities among evaluations given by evaluators from two midwifery schools. Evaluators from the University A showed a 30% rate of "poor agreement" in their evaluations of the same student records, while the rate was 70% for evaluators at the University B. Disparity levels were reduced to 10% and 40%, respectively, when

evaluation standards based on our rubric were used. The results suggest that introduction of the rubric as an instrument for evaluation of antenatal health check-ups can be effective for improvement of consistency among evaluators' evaluations of student performance.

Acknowledgements

We would like to express our heartfelt gratitude to the research subjects who willingly cooperated with this study. We would also like to thank the graduates of the Graduate School of Nursing at University of Shizuoka who provided the clinical practice records of antenatal health check-ups for the purpose of this study, as well as the instructors at University of Shizuoka and Kanazawa University who conducted assessment of the clinical practice records using the 2 evaluation tools.

References

- 1) Takagi S, Suzuki M, Matsuura M: Problems Involved in development of midwifery clinical skills; A consideration from the standpoint of current state of in-hospital midwifery clinical care. *Shizuoka Journal of Maternal Health* 3 (1): 33-41, 2013
- 2) The Ministry of Health, Labor and Welfare : Attainment level of technical competencies in midwife and nursing education at the time of graduation, 0208001, 2008 http://www.hospital.or.jp/pdf/15_20080208_01.pdf, accessed on 2016.09.20
- 3) Hirasawa M, Sindo S, Naito Y et al: A Study of the process of care ability acquirement and factors influencing the process; Especially the ability of care of women in pregnancy and childbearing. *Journal of Japan Academy of midwifery* 2 (1): 21-31, 1988
- 4) Ebata Y: Problems of midwives' technical education from the actual survey. *The Japanese Journal for midwives* 58 (3): 204-210, 2004
- 5) Supervised by Japanese Nursing Association: New edition, *Midwife Work Handbook – Auxiliary Version*, Japanese Nursing Association Press, Tokyo, pp 280-281, 2008
- 6) Japan society of midwifery education: The annual report of 2012, Japan society of midwifery education Press, Tokyo, pp 47-58, 2013
- 7) Japan society of midwifery education: The annual report of 2012, Japan society of midwifery education Press, Tokyo, pp 37-46, 2013
- 8) Supervised by Japanese Nursing Association: New edition, *Midwife Work Handbook – Auxiliary Version*, Japanese Nursing Association Press, Tokyo, pp 250-257, 2008
- 9) Japan Obstetric Compensation System for Cerebral Palsy: The second, Report on the Prevention of Recurrence, Japan Council for Quality Health Care Press, Tokyo, pp 63-70, 2012
- 10) Japanese Nursing Association: *Midwife Training guide*, Japanese Nursing Association Press, Tokyo, pp 40, 2012
- 11) Murakami J, Takenaka H, Horikoshi A et al: Problems in the evaluation of medical interviewing skill with objectives structured clinical examinations: How can reasonable objectivity be ensured?. *Medical Education* 32 (4): 231-237, 2001
- 12) Suzuki E, Ito M, Aoyagi Y et al: Study of the suitability of evaluation of initial objectives structured clinical examinations at the Niigata University School of Medicine. *Medical Education* 34 (1): 37-44, 2003
- 13) Suganuma T, Kakuta S, Hasegawa T et al: Evaluation agreement in objective structured clinical examination. *The Journal of Japanese Dental Education Association* 21 (1): 3-9, 2005
- 14) Aizawa F, Yomemitsu M, Mizuki H: Relations between reliability/ validity in OECE evaluation and "word" used for evaluation item sentence. *The Journal of Japanese Dental Education Association* 21 (2): 123-132, 2005
- 15) Ito T, Aoki S, Oomine H et al: OSCE evaluation discrepancies in medical interview settings –Based on analysis of Human errors-. *The Journal of Japanese Dental Education Association* 21 (1): 21-30, 2005
- 16) Kubota Y, Yano Y, Seki S et al: A study of the evaluation of information- gathering capability in pharmaceutical objective structured clinical examination. *Medical Education* 41 (4): 273-279, 2010
- 17) Matsushita K: Assessment of the quality of learning through performance assessment: based on the analysis of types of learning assessment. *Kyoto University Higher Education Research bulletin* 18: 75-114, 2012
- 18) Suskie L: *Assessing Student Learning –A Common Sense Guide- second edition*, Jossey-bass Press, San Francisco, pp 137-154, 2009
- 19) Kawai H: Research and development of Rubric aiming at objective evaluation: The report of Grant-in-Aid for Scientific Research (C), National Institute for Educational Policy Research, Tokyo, pp 1-33, 2003
- 20) Stevens D D, Levi A: *Introduction to rubric: an assessment tool save grading time, conveys effective feedback, and promote student learning*, Stylus Publishing, Virginia, pp 29-46, 2004
- 21) Tsushima E: Application of coefficient of reliability to the studies of physical therapy. *Rigakuryoho Kagaku* 17 (3): 181-187, 2002
- 22) Fleiss J L: Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76 (5): 378-382, 1971
- 23) Sim J, Wright C C: The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy* 85 (3): 257-268
- 24) Tanner C A: Professional judgment in evaluation. *Journal of Nursing Education*, 33: 243-244, 1994
- 25) Oermann M H, Gaberson K B: *Evaluation and testing in nursing education- Fourth edition-*, Springer Publishing Company, New York, pp 233-248, 2014
- 26) Kabeyama K, Yagi K, Taniguchi H et al: Teaching ability of instructors belonging to midwifery school. *Health science: annual reports of School of Health Sciences, Faculty of Medicine, Kyoto University* 8: 1-18, 2013
- 27) Kim S J, Kim S: Development of a simulation evaluation tool for assessing nursing students' clinical judgment in caring for children with dehydration. *Nurse Education Today* 37: 45-52, 2016.

妊婦健康診査実習における助産学生の実習記録の評価の評価者間の一致度の検討 ～開発したルーブリックの有用性～

山田 貴代, 高木 静*, 東野 定律**, 島田 啓子***, 杉本 公一****

要 旨

背景

日本の助産師教育において、実習の中心は分娩介助である。助産師養成所指定規則では分娩介助について明記されているが、妊婦ケアについては妊娠中期から産褥まで継続して一例以上受け持つとされているのみで、そのため、妊婦健康診査の実習は各養成校に柔軟に任されてきた。ゆえに妊婦健康診査の実習は遅れて始まり、その評価方法も基準の検討も十分でなかった。そこで、妊婦健康診査の実習評価について客観的に信頼性の高い評価方法を考案する必要があった。

目的

本研究は、妊婦健康診査実習の実習記録に対する評価の客観性を高めるため、複数の評価者間の評価の一致度から評価者による評価の差異を明らかにし、さらにルーブリックの作成により、評価者間の評価の差異が少ない評価表の作成を目的とした。

研究方法

1) 研究期間

平成 28 年 1 月～8 月

2) 研究対象

2 年課程の助産師養成校、A 大学看護学研究科で妊婦健康診査実習の指導経験がある教員 3 名と、B 大学の看護学研究科博士前期課程で妊婦健康診査実習の指導経験がある教員 3 名

3) 評価用具

- (1) 日本看護協会「新卒助産師研修ガイド」の妊婦ケアのチェックリストから作成した評価表と、
- (2) 妊婦健康診査実習の実習記録の評価のために開発したルーブリック

4) 評価対象

2 年課程の助産師養成校で妊婦健康診査実習を終えた助産師学生の妊婦 10 事例についての妊婦健康診査の実習記録

5) 分析方法

2 つの評価用具を用いて、A 大学教員 3 名と B 大学教員 3 名が助産師学生の妊婦健康診査実習の実習記録の評価を事例ごとに行い、各大学教員 3 名の評価の結果について、項目別、実習記録別にフライスの κ 係数を用いて各大学での教員間の差異を比較した。

結果

評価用具「新卒助産師研修ガイド」の妊婦ケアのチェックリストから作成した評価表を用いた評価結果の三人の評価者間の実習記録に対する一致度は、A 大学教員群で「ほとんど一致しない」は 3 事例 (30%) であり、B 大学教員群では 7 事例 (70%) であった。「わずかな一致」は A 大学教員群では 7 事例 (70%) であり、B 大学教員群では 2 事例 (20%) で、より一致度の高い「中等度の一致」は認められなかった。評価項目については、A 大学教員群で「ほとんど一致しない」は 12 項目 (92.3%) で、B 大学では 9 項目 (56.3%) であった。「わずかな一致」を示したのは A 大学教員群では 2 項目 (12.5%)、B 大学教員群では 5 項目 (31.3%) であった。

一方、開発したルーブリックを用いた評価結果では、「ほとんど一致しない」は A 大学教員群で 1 事例 (10%)、B 大学教員群では 4 事例 (40%) であった。「わずかな一致」は、A 大学教員群で 5 事例 (50%)、B 大学教員群では 4 事例 (40%) であった。また、A 大学教員群では、より一致度の高い「中等度の一致」が 3 事例 (30%) あった。評価項目については、A 大学教員群では、「ほとんど一致しない」を示したのは 2 項目 (25%)、B 大学教員群では 4 項目 (50%) であった。また、「わずかな一致」を示したのは A 大学教員群では 5 項目 (62.5%)、B 大学教員群では 3 項目 (37.2%) であった。

結論

妊婦健康診査実習の評価の客観性をみるため、助産学生の実習記録から 2 つの助産師養成校の評価者間の評価に差異が明示された。同一の実習記録に対する評価の「ほとんど一致しない」の割合は、A 大学教員群 30% と B 大学教員群 70% の差異を認めた。開発したルーブリックを用いた場合、同一の実習記録に対する評価の「ほとんど一致しない」の割合は A 大学教員群 10%、B 大学教員群 40% まで減少した。ルーブリックを用いた評価の方が、教員間の評価の一致度は向上したことにより、ルーブリックは今後の妊婦健康診査実習の評価用具として導入する有用性が示唆された。