

Outcome-based self-assessment on a team-teaching subject in the medical school

Sang Pil Yoon, Sa Sun Cho

Department of Anatomy, Jeju National University School of Medicine, Jeju, Korea

Abstract: We attempted to investigate the reason why the students got a worse grade in gross anatomy and the way how we can improve upon the teaching method since there were gaps between teaching and learning under recently changed integration curriculum. General characteristics of students and exploratory factors to testify the validity were compared between year 2011 and 2012. Students were asked to complete a short survey with a Likert scale. The results were as follows: although the percentage of acceptable items was similar between professors, professor C preferred questions with adequate item discrimination and inappropriate item difficulty whereas professor Y preferred adequate item discrimination and appropriate item difficulty with statistical significance ($P < 0.01$). The survey revealed that 26.5% of total students gave up the exam on gross anatomy of professor Y irrespective of years. These results suggested that students were affected by the corrected item difficulty rather than item discrimination in order to obtain academic achievement. Therefore, professors in a team-teaching subject should reach a consensus on an item difficulty with proper teaching methods.

Key words: Assessment, Integrated curriculum, Outcome, Team-teaching

Received September 5, 2013; Revised November 13, 2013; Accepted September 17, 2014

Introduction

Anatomy has historically been a cornerstone in medical education regardless of nation or specialty. Throughout the word, recent developments in medical curricula have led to marked changes in the teaching of gross anatomy [1-6]. This change has involved the integration of basic science with clinical cases, the general need to shorten formal anatomy course, and the acknowledged value of dissection. As a result, substantial problems eventually deteriorate the quality of education [3]. Bergman et al. [7] attempted to reveal that students' anatomical knowledge is impaired due to negative

effects from several factors, including teaching by non-medically qualified teachers, diminished use of cadaver dissection as a teaching tool and neglect of vertical integration of anatomy teaching. However, they found there is a lack of sufficient quantity and quality of information within the existing literature to support any of the claims.

Korean medical education is rapidly changing. The general changes are towards reducing the factual load in the curriculum and encouraging students to develop their own learning skills. To accomplish these changes, integrated curriculum and co-teaching has been implemented in medical education [8]. Under these circumstances, co-teaching is indispensable to each medical school. Co-teaching can be defined as "two or more professionals delivering substantive teaching to a diverse or blended group of students in a single physical space" and have four key components as follows: 1) two or more instructors, 2) delivery of meaningful instruction, 3) diverse group of students, and 4) common settings [9]. The trend is based on the confidence that medical students are motivated and enabled to use their out-of-class

Corresponding author:

Sang Pil Yoon

Department of Anatomy, Jeju National University School of Medicine,
102 Jejudaehang-ro, Jeju 690-756, Korea
Tel: +82-64-7543823, Fax: +82-64-7252593, E-mail: spyoon@jejunu.ac.kr

time productively by lessening course hours [2].

The anatomy course at the Jeju National University School of Medicine faced these same challenges as appeared in Table 1 [5, 6, 10]. Over the decades prior to this study, the course was reduced from 272 hours (lecture 85 and dissection 187 in 2001) to 126 hours (lecture 36 and dissection 90 in 2011). We had 46 lecture and 96 dissection hours in 2012, which was similar to that of United States in 2009 [6] but did not reach the recommendation from Korean Association of Anatomists [10]. Besides, Jeju National University chose to restructure its first year medical curriculum in 2012. The anatomy and histology faculties had prior experience in integrating courses, stemming from the successful integration of individual anatomical sciences including gross anatomy and histology into a single course called "Structure of human body (2011)." The integration process was expanded to include the other first year basic science courses including embryology, neuroanatomy and physiology as we progressed toward an integrated course called "Human body: structure and function (2012)."

With regard to this integration course (Human body: structure and function), we gained important insights from summative assessments in 2012. Above all, we learned that there are gaps between accepting a set of assumptions about teaching and learning on gross anatomy. The results of gross anatomy showed poor compared with those of total course, which were caused by the dropped grades on the part of one professor. In this respect, we attempted to investigate the reason why the students got a worse grade in gross anatomy of one professor, and the way how we can improve upon the teaching method.

Table 1. The changes of hours in lecture and dissection for anatomy education in Medical School, Jeju National University (JNU) as compared to those of Republic of Korea (ROK) and United States (USA)

		Past	Recent	Present
JNU ^{a)}	Lecture	85	36	46
	Dissection	187	90	96
ROK ^{b)}	Lecture	96.7±26.4	65.2±32.8	60
	Dissection	154.6±50	83.1±45.9	100-120
USA	Lecture	167±39 ^{c)}	43±19 ^{d)}	NS
	Dissection	NS	94±29	NS

^{a)}Jeju National University set the medical school at 1998, so the data was adjusted as follows: past, 2001; recent, 2011; present, 2012. ^{b)}Data was quoted from Hwang (2010) [10]: past, 1990; recent 2009; present, recommended hours from Korean Association of Anatomists in 2010. Data from United States was quoted from ^{c)}Drake et al. (2002) [5] for 'Past' and ^{d)}Drake et al. (2009) [6] for 'Recent'. Drake et al. (2002) [5] suggested the data in total not dividing the lecture and dissection. Present data were not specified (NS).

Materials and Methods

Item analysis

Item analysis can be distinguished into the qualitative analysis considering of the content validity and the quantitative analysis consisting of the item difficulty and item discrimination [11, 12]. In this study, we analyzed item difficulty and item discrimination based on classical test theory. Data were obtained from retrospective review of course written examination item analysis for year 2012, which were analyzed by the previous reports [13, 14]. We conducted exploratory factor analysis to testify validity: corrected item difficulty and item discrimination. Cronbach's alpha coefficient was used to examine the reliability of tests.

Items were further classified by corrected item difficulty (0.75) and item discrimination (0.2) as suggested by a previous report [13] with slight modification. That is, excellent items meet the condition of appropriate difficulty and adequate discrimination. Good items meet the condition of adequate discrimination in spite of inappropriate discrimination. Considered items meet the condition of inappropriate difficulty and inadequate discrimination, which need to be corrected. Separately, eliminable items meet the condition of inadequate discrimination (<0.2) irrelevant to difficulty, which need to be excluded.

Results on corrected item difficulty and item discrimination were processed by SPSS version 18.0 (SPSS Inc., Chicago, IL, USA).

Survey

Total of 83 participants were enrolled in a gross anatomy course, who were taught during the first semester of the first

Table 2. General characteristics of study participants

Characteristic	No. (%)
Gender	
Male	46 (55.4)
Female	37 (44.6)
Grade	
1st year	41 (49.4)
2nd year	42 (50.6)
Age (y)	
20-29	63 (75.9)
30-39	20 (24.1)
Major prior to entrance	
Humanities	7 (8.4)
Natural science including engineering	63 (75.9)
Education and pedagogy	6 (7.2)
Public health	7 (8.4)

year at Jeju National University School of Medicine: 41 in 2012 and 42 in 2011 (Table 2). All participants had completed the lecture and dissection course on gross anatomy as a part of the standard curriculum. Participating in the survey were voluntary, students were informed that their answers were confidential.

The survey was carried out by “Institute of Liberal Education” after the course completed. All participating students were asked to complete a short questionnaire about their perception of the teaching and learning experiences. Students were asked to anonymously indicate the degree to which they agreed with each statement on a Likert scale ranging from 1 to 5 (1 [strongly disagree] to 5 [strongly agree]) in self-administered questionnaire survey. Immediately after completing the survey, an investigator from Institute of Liberal Education had an interview with randomly selected several students from each grade (first and second year). The quantitative data were processed by SPSS version 18.0. Ordinal variables were compared using chi-square test for trends.

Results

General aspects of course and tests during restructuring the curriculum

According to the recent curriculum, the gross anatomy course had 126 and 142 hours of the first year in 2011 and 2012, respectively (Table 3) [15]. Content quantification regarding the separate courses accounts for 35.7%, 37.3%, and 27.6% in 2011, which was in accord with the text of ‘Human Anatomy.’ The percentage of the separate courses changed to 38.7%, 40.1%, and 21.1% in 2012.

Proportion to the test of professors C and Y were 61.67 and 38.33% in 2011 while 47.54% and 52.46% in 2012, respectively. Table 4 summarizes the number and types of questions regarding the separate courses. Various types of multiple choice questions were used to evaluate the students’ ability such as one-best answer type (type A), multiple true-false type (type K), extended-matching choice (type R), matching, blank and short description in 2011. But a multiple choice question including types A and K was the only method in 2012, except 6 point (4.92%). In total, professor C set the questions as types A and K 30.53% and 69.47% in 2011, and 36.54% and 63.46% in 2012, respectively. Professor Y set the

Table 3. Content quantification of Terminologia Anatomica (TA), American Association of Clinical Anatomists (AACA), and anatomy curriculum at the Jeju National University Medical School

	Textbook (%)	Hours ^{b)}		Recommendation (%)	
	Human anatomy ^{a)}	2011	2012	TA	AACA
Introduction and limbs	33.33	45 (35.7)	55 (38.7)	21.46	26.08
Trunk	39.58	47 (37.3)	57 (40.1)	30.92	39.66
Head and neck	27.08	35 (27.6)	30 (21.1)	47.44	34.26

^{a)}Source: Korean Association of Anatomists (2005), Human Anatomy, 2nd ed. Korea Medical Book Publisher [15]. ^{b)}Values are presented as number (%).

Table 4. Types of questions of the course in 2011 and 2012

Variable	Professor	2011			2012		
		Limbs (n=100)	Trunk (n=100)	Head (n=100)	Limbs (n=50)	Trunk (n=45)	Head (n=27)
A type	C	16	8	5	6	13	-
	Y	-	39	8	15	19	10
K type	C	34	21	11	22	5	6
	Y	-	11	-	7	8	5
R type	C	-	-	-	-	-	-
	Y	-	-	16	-	-	-
Matching	C	22	-	8	-	-	4
	Y	-	-	31	-	-	-
Blank	C	8	4	15	-	-	2
	Y	-	10	-	-	-	-
Short description	C	20	7	6	-	-	-
	Y	-	-	-	-	-	-

A type, one-best answer type of multiple choice question; K type, multiple true-false type of multiple choice question, R type, extended-matching choice of multiple choice question.

questions as types A, K and R 63.51%, 14.87%, and 21.62% in 2011, and types A and K 68.75% and 31.25% in 2012. Overall, professor C preferred types K and Y preferred type A.

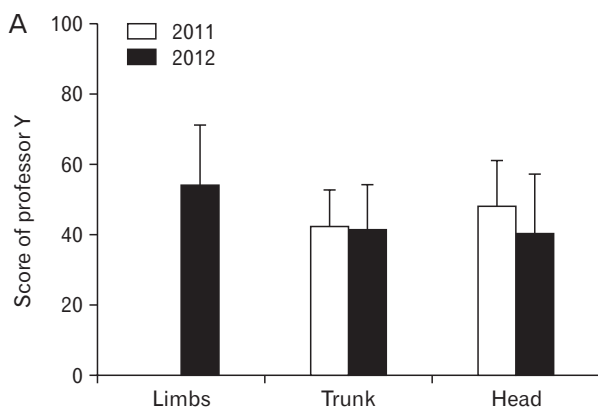
Table 5 summarizes the results of academic scores in 2011 and 2012 regarding the separate courses of gross anatomy. No significant differences in scores were found between two years. In 2011, the mean score for lecture and dissection was 55.07±9.62 and 55.03±13.56, respectively. In 2012, the mean score for lecture and dissection was 57.32±10.11 and 55.0 ±16.03, respectively.

However, the academic achievement scores for lecture in 2012 were considerably changed between two professors. The mean scores of professor C were significantly higher than those of professor Y, although the total scores showed no significant differences comparing 2011 and 2012 (Fig. 1). In 2011, the mean scores of professor C regarding the separate courses were 61.75±14.64, 64.98±11.73, and 52.93±11.93 while the scores of professor Y were 0±0 (not assessed in

Table 5. Comparison of scores regarding the separate courses of introduction and limbs, trunk, and head and neck between groups

Variable	2011	2012	Note ^{a)}
Introduction and limbs			
Lecture	61.75±14.64	62.76±11.90	73.77±10.38
Dissection	36.9±11.60	55.33±14.10	-
Trunk			
Lecture	51.58±10.76	53.18±10.67	75.46±11.61
Dissection	58.93±11.88	56.5±15.47	-
Head and neck			
Lecture	50.49±11.01	54.15±14.22	78.55±11.28
Dissection	69.25±17.20	53.27±18.53	-

^{a)}Mean scores of the course (Human body: structure and function) except gross anatomy.



introduction and limbs part), 42.65±10.12, and 48.47±12.78. In 2012, the mean scores of professor C regarding the separate courses were 69.04±11.89, 70.22±13.67, and 71.42±15.42 while the scores of professor Y were 54.77±16.68, 42.82±12.56, and 40.33±16.67.

Item analysis between professors C and Y in 2012

Item analysis was done to investigate whether there were problems of the problem between professors in 2012 (Fig. 2). The Cronbach's alpha for each test was 0.77, 0.64, and 0.66 regarding introduction and limbs, trunk, and head and neck section, respectively.

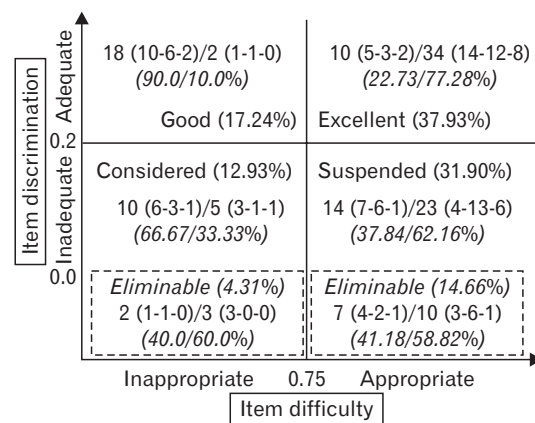


Fig. 2. Classification of items according to item difficulty and item discrimination in 2012. Although the percentage of excellent and good items of each professor was similar, professor C preferred questions with adequate item discrimination and inappropriate item difficulty whereas professor Y preferred adequate item discrimination and appropriate item difficulty with statistical significance ($P<0.01$). The number of items was appeared in order as professor C and professor Y.

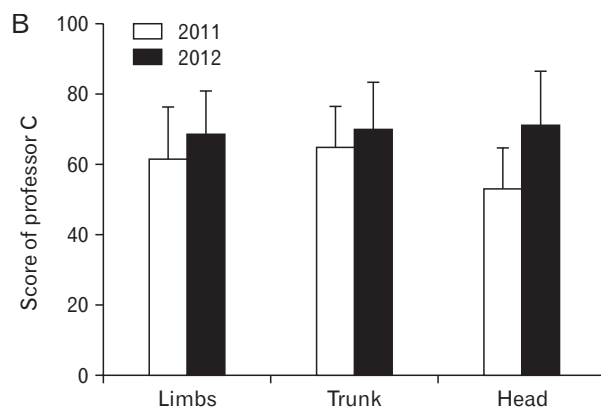


Fig. 1. The academic achievement scores of professor Y (A) and professor C (B) in 2011 and 2012. Although the total scores showed no significant differences comparing 2011 and 2012 (see Table 4 also), there was a wide gap between two professors in 2012. Limbs, introduction and limbs; Head, head and neck.

Items were further classified by corrected item difficulty (0.75) and item discrimination (0.2) as described above: excellent, good, suspended, considered, and eliminable. In total of 116 questions, excellent, good, suspended, considered, and eliminable items accounted for 44 (37.93%), 20 (17.24%), 37 (31.90%), 15 (12.93%), and 22 (18.97%), respectively. The academic problems contain 55.17% of acceptable items, which consist of excellent and good items.

Item analysis was further divided by each professor. In 52 questions of professor C, excellent, good, suspended, considered, and eliminable items accounted for 10 (19.23%), 18 (34.62%), 14 (26.92%), 10 (19.23%), and 9 (17.31%), respectively. The academic problems of professor C contained 53.85% of acceptable items. In 64 questions of professor Y, excellent, good, suspended, considered, and eliminable items accounted for 34 (53.13%), 2 (3.13%), 23 (35.94%), 5 (7.81%), and 13 (20.31%), respectively. The academic problems of professor Y contained 56.26% of acceptable items.

Cross tabulation on the item discrimination and difficulty showed significant differences based on each professor (Table 6). In case of adequate item discrimination, professor C preferred inadequate item difficulty whereas professor Y

preferred adequate item difficulty. The tendency observed uniformly in each separate test, while there was no considerable discrepancies on the item discrimination and difficulty of each professor during separate test.

To investigate whether students were not adapted for a certain professor, the number of upper grade and lower grade were compared in each test. As appeared in Fig. 3, the number in upper grade overlapped 4 (36.4%), 5 (41.7%), and 3 (21.4%), and in lower grade 5 (38.5%), 4 (33.3%), and 6 (50.0%), respectively. The number of students who gained upper grade in one professor also contained in lower grade in another professor was 1, 3, and 0, respectively.

Survey on the gross anatomy course in 2012

We included the self-assessment score regarding the learning objectives, academic achievement, and the students' perception of the learning experience as outcome variables to evaluate the effects of the teaching methods between years. The Cronbach's alpha for this questionnaire was 0.81.

The survey sought to answer the research questions as appeared in Table 7. No significant difference was found in each objective except "Various knowledge needed to be acquired through contents is dealt with in class," which showed a P-value of 0.032. On the assumption that the exam did not considerably differ between years, we checked which objectives affected the perception on the exam (Table 8). All objectives were significantly associated with the exam except the following objectives: "Gives appropriate examples for an important concept and connects it to the prior knowledge (P=0.120)" and "I am capable of understanding the class (P=0.136)."

Free description and mini-interview sought to answer the following questions: 1) What was the most difficult part to understand in class?; 2) At what point do you think

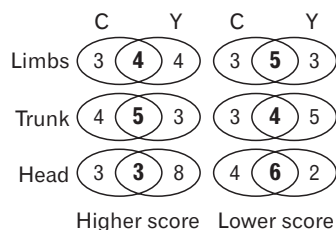


Fig. 3. The number of students in upper and lower grades in each test. To investigate whether students were not adapted for a certain professor, there was no significant difference between groups and among tests.

Table 6. Cross tabulation on the item discrimination and difficulty based on each professor C and Y

Discrimination			Professor		Total	χ ²	P-value
			C	Y			
Adequate	Difficulty	IA	18 (90.0)	2 (10.0)	20	25.29	0.00
		A	10 (22.7)	34 (77.3)	44		
		Total	28 (43.8)	36 (56.3)	64		
Inadequate	Difficulty	IA	10 (66.7)	5 (33.3)	15	3.57	0.07
		A	14 (37.8)	23 (62.2)	37		
		Total	24 (46.2)	28 (53.8)	52		
Total	Difficulty	IA	28 (80.0)	7 (20.0)	35	25.07	0.00
		A	24 (29.6)	57 (70.4)	81		
		Total	52 (44.8)	64 (55.2)	116		

Values are presented as number (%). IA, inappropriate; A, appropriate.

Table 7. The survey on the team-teaching, gross anatomy and professor

Objectives	Year	Likert scale					χ^2	P-value
		1	2	3	4	5		
Explains the learning contents clearly and concisely.	2011	0	6	16	14	5	4.46	0.35
	2012	2	4	15	19	2		
Made introduction to interesting things about the topic.	2011	0	2	6	22	11	6.26	0.18
	2012	1	3	14	18	6		
Gives appropriate examples for an important concept and connects it to the prior knowledge.	2011	1	5	6	17	12	5.07	0.28
	2012	1	2	9	24	6		
Encourages students to express their own opinions freely in class.	2011	0	3	10	12	15	4.13	0.25
	2012	3	3	11	20	8		
I am capable of understanding the class.	2011	8	13	17	3	0	2.33	0.68
	2012	5	12	19	5	1		
Various knowledge needed to be acquired through contents is dealt with in class.	2011	0	1	5	17	16	10.53	0.03
	2012	2	1	10	24	5		
The questions in the exam are based on what were covered and regarded as important in class.	2011	0	6	12	16	7	4.90	0.30
	2012	1	1	14	18	8		
Make sure that students can understand the class here and there in class.	2011	3	13	9	9	7	1.69	0.79
	2012	5	10	12	10	5		
Are you satisfied with the class to some extent?	2011	0	9	10	22	0	8.04	0.09
	2012	1	6	11	18	6		
Are you satisfied with a team teaching made according to the contents of classes?	2011	0	3	11	21	6	6.56	0.16
	2012	2	6	17	13	4		
Are you satisfied with teaching methods of professors in a team teaching?	2011	0	6	12	21	2	3.05	0.55
	2012	1	6	17	15	3		

Table 8. Cross tabulation between the objectives on the exam and other objectives

The questions in the exam are based on what were covered and regarded as important in class	χ^2	P-value
Explains the learning contents clearly and concisely.	28.51	0.03
Made introduction to interesting things about the topic.	32.00	0.01
Gives appropriate examples for an important concept and connects it to the prior knowledge.	22.78	0.12
Encourages students to express their own opinions freely in class.	21.98	0.04
I am capable of understanding the class.	22.23	0.14
Various knowledge needed to be acquired through contents is dealt with in class.	56.27	0.00
Make sure that students can understand the class here and there in class.	50.97	0.00
Are you satisfied with the class to some extent?	34.55	0.01
Are you satisfied with a team teaching made according to the contents of classes?	65.80	0.00
Are you satisfied with teaching methods of professors in a team teaching?	107.59	0.00

the professor's teaching is the best?; 3) What part of the curriculum or classes of professors are you less satisfied with?; and 4) Give suggestions to improve the level of the interest and the effectiveness of next class. Twenty-two of total students (26.5%) appealed that gross anatomy was not familiar with them, not easy to study, and too much volume to study in the same period. As a result, they could not help but lay aside their studies on gross anatomy.

Discussion

Although there is less time to teach anatomy, almost all

anatomists believe that gross anatomy has an important role to play in clinical medicine [1]. A recent review [16] also revealed that gross anatomy graded with the highest relevance by students and clinicians. But current anatomical course of medical students is inadequate and even below the minimum essential for safe medical practice. Although there are a variety of co-teaching options including team-teaching, the impact of such procedures on student outcomes is unclear. The previous meta-analysis [17, 18] revealed that a descriptive summary in co-teaching is frequently stated qualitatively using "improved" or "better," but there has rarely been a quantitative analysis of research in this area. There is

a continued impetus for collaboration, yet practitioners and researchers have not made, or cannot make, empirically based claims that their cooperative efforts have been effective in terms of student outcomes [17]. Therefore, we attempted to provide a synthesis of the quantitative and the qualitative data on the effectiveness and shortcomings of team-teaching in the anatomy course. Student performance using descriptive quantitative statistics was obtained through the scores on curriculum-based multiple choice questions and practical test with cadavers in a given academic area. Data for attitudinal and affective information was gathered by surveys on team-teaching and by an interview by an external investigator for validity.

The outcomes of the gross anatomy in 2012 were evaluated by multiple choice questions. Elements of the assessment process to be evaluated are item analysis, reliability and validity, survey and mini-interview. An external investigator from "Institute of Liberal Education" did the item analysis, reliability and validity, survey and mini-interview, respectively. Although the percentage of professor Y in 2012 was increased by 14.13% compared to 2011, the results showed significantly poor compared with another professor in gross anatomy.

Item analysis showed following features. First, professor Y preferred type A multiple choice question whereas professor C preferred type K. This preference did not change throughout the course. Second, Although the percentage of acceptable items of each professor was similar between professors C (53.85%) and Y (56.26%), professor C preferred questions with adequate item discrimination and inappropriate item difficulty whereas professor Y preferred adequate item discrimination and appropriate item difficulty with statistical significance ($P < 0.01$). Finally, further analysis including differences between grades and mal-adaption for a certain professor did not show that there were problems of the problem between professors and/or among the tests. These results suggested that students may come under the influence of corrected item difficulty rather than item discrimination, although the latter is known to be more important than the former in item analysis [13]. We wish to call particular attention to the following fact that the mean scores of professor Y were significantly lower even though the percentage of professor Y exceeded 60% in second and third test. These results showed that students look for the way how to achieve good record not to fail in a subject. Under these circumstances of team-teaching, more difficult part of one team-teaching subject can easily be abandoned although

medical students are known to be motivated and enabled to use their out-of-class time productively by lessening course hours [2].

The survey showed no significant difference between years except one objective, which means that main cause of academic achievement in 2012 might not be instructor's class operation and teaching methods but a lack of lesson time. In addition, free description and mini-interview revealed that 26.5% of total students gave up the exam on gross anatomy of professor Y. These results may be relevant to the "Medical Education Eligibility Test" in Republic of Korea. That is, parts of embryology, histology and physiology are included in the "Medical Education Eligibility Test," but not gross anatomy. As a result, students in first grade appealed that lecture on gross anatomy was difficult to understand because it was so disjointed. We think that abandonment might lead to the cause of the difference between professors on gross anatomy.

An external investigator from 'Institute of Liberal Education' suggested following reform measures to professor Y, although the interaction between professor Y and students was relatively good: Professor Y should make his lecture handout clearly to improve the context, adjust his speed of lecture, and consider other tests including formative test to improve the inconsistency of students' intelligibility and lecturer's progress.

It is important to assess the function and procedure of team-teaching through summative evaluation procedures. The collection of data on student outcomes will also allow researchers to provide teachers with a more cogent understanding of when, where, and with whom team-teaching is best implemented [17, 18]. Although the innovative educational programs including co-teaching were embraced in a multimodal system implemented worldwide, the various flexible medical curricula can not affect student learning and satisfaction [19], and medical students view team-based learning favorably irrespective of their grades [20] under recent educational environments. But, examination scores remained significantly higher for learners receiving the instruction [21] and the level of challenge in the material that is presented to learners is an important factor to consider in the evaluation [22]. As outcome-based medical education come to the fore, assessment and evaluation hold the central place [23]. In addition to qualitative methods, student performance using descriptive quantitative statistics obtained through the scores on curriculum-based assessment can suggest a meaningful data in a given academic area.

In conclusion, this study revealed that students were affected by the item difficulty irrelevant to the quantity in order to obtain academic achievement. Therefore, professors in a team-teaching subject should reach a consensus on an item difficulty and should consider their teaching methods so that students would not to give up the gross anatomy.

Acknowledgements

The authors would like to thank Ms. Hyo Jeong Hong (Institute of Liberal Education, Jeju National University) for statistical analysis and helpful discussion. This work was supported by a research grant from the Jeju National University Hospital Research Fund of Jeju National University in 2013.

References

- Patel KM, Moxham BJ. Attitudes of professional anatomists to curricular change. *Clin Anat* 2006;19:132-41.
- Rizzolo LJ, Rando WC, O'Brien MK, Haims AH, Abrahams JJ, Stewart WB. Design, implementation, and evaluation of an innovative anatomy course. *Anat Sci Educ* 2010;3:109-20.
- Sugand K, Abrahams P, Khurana A. The anatomy of anatomy: a review for its modernization. *Anat Sci Educ* 2010;3:83-93.
- Klement BJ, Paulsen DF, Wineski LE. Anatomy as the backbone of an integrated first year medical curriculum: design and implementation. *Anat Sci Educ* 2011;4:157-69.
- Drake RL, Lowrie DJ Jr, Prewitt CM. Survey of gross anatomy, microscopic anatomy, neuroscience, and embryology courses in medical school curricula in the United States. *Anat Rec* 2002;269:118-22.
- Drake RL, McBride JM, Lachman N, Pawlina W. Medical education in the anatomical sciences: the winds of change continue to blow. *Anat Sci Educ* 2009;2:253-9.
- Bergman EM, van der Vleuten CP, Scherpbier AJ. Why don't they know enough about anatomy? A narrative review. *Med Teach* 2011;33:403-9.
- Oh SA, Chung EK, Rhee JA, Baik YH. An evaluation of integrated curriculum based on students' perspective. *Korean J Med Educ* 2007;19:305-11.
- Cook L, Friend M. Co-teaching: guidelines for creating effective practices. *Focus Except Child* 1995;28:1-16.
- Hwang YI. Reasonable hours of lecture and dissection for anatomy education in medical school. In: 60th Congress of Korean Association of Anatomists; 2010 Oct 20-23; Jeju, Korea. p. S1-1: 22.
- Kulik JA, McKeachie WJ. The evaluation of teachers in higher education. *Rev Res Educ* 1975;3:210-40.
- Kaplan RM, Saccuzzo DP. *Psychological testing: principles, applications, and issues*. 6th ed. Pacific Grove: Brooks/Cole Publishing Co.; 2004.
- Lee YM, So YH, Ahn DS, Rhee KJ, Im H. Psychometric analysis of comprehensive basic medical sciences examination. *Korean J Med Educ* 2002;14:301-6.
- Park JC, Kim KS. A comparison between discrimination indices and item-response theory using the Rasch Model in a clinical course written examination of a medical school. *Korean J Med Educ* 2012;24:15-21.
- Korean Association of Anatomists. *Human anatomy*. 2nd ed. Seoul: Korea Medical Book Publisher; 2005.
- Pabst R. Anatomy curriculum for medical students: what can be learned for future curricula from evaluations and questionnaires completed by students, anatomists and clinicians in different countries? *Ann Anat* 2009;191:541-6.
- Welch M, Brownell K, Sheridan SM. What's the score and game plan on teaming in schools? A review of the literature on team teaching and school-based problem-solving teams. *Remedial Spec Educ* 1999;20:36-49.
- Murawski WW, Swanson HL. A meta-analysis of co-teaching research: where are the data? *Remedial Spec Educ* 2001;22:258-67.
- Jelsing EJ, Lachman N, O'Neil AE, Pawlina W. Can a flexible medical curriculum promote student learning and satisfaction? *Ann Acad Med Singapore* 2007;36:713-8.
- Vasan NS, DeFouw DO, Compton S. A survey of student perceptions of team-based learning in anatomy curriculum: favorable views unrelated to grades. *Anat Sci Educ* 2009;2:150-5.
- Reilly FD. Outcomes from building system courseware for teaching and testing in a discipline-based human structure curriculum. *Anat Sci Educ* 2011;4:190-4.
- Chariker JH, Naaz F, Pani JR. Item difficulty in the evaluation of computer-based instruction: an example from neuroanatomy. *Anat Sci Educ* 2012;5:63-75.
- Gruppen LD. Outcome-based medical education: implications, opportunities, and challenges. *Korean J Med Educ* 2012;24:281-5.