A Case Study of University-level Immersion: One Class for Two Goals[*]

Haemoon Lee

(Sungkyunkwan University)

Lee, Haemoon. (2006). A case study of university-level immersion: One class for two goals. *English Teaching, 61*(3), 145-168.

This paper examines the effect of English immersion of university mathematics class in Korea in terms of English proficiency gain and mathematics content learning. One hundred and seventy students in three math classes were the subjects for the pre and post design research at the 9 weeks interval. The results show first that the immersion students' content-learning was superior to the non-immersed L1 math class. Second, the immersion students' language proficiency gain in content-obligatory and content-compatible English was significant in contrast to the non-immersed counterpart, though their general English proficiency did not improve like their non-immersed counterparts. It is concluded that university level immersion does not reduce content learning and that the language proficiency gain is limited to the content-relevant linguistic features, during the 9 weeks interval.

I. INTRODUCTION

Content-based language instruction (CBI) has been approached from two purposes: learning the content knowledge and learning the target second or foreign language. Depending on the degree of the emphasis between these two goals, a variety of different forms of CBI have been developed in a continuum ranging from the most content-driven approach to the most language-driven approach (Met, 1991, 2004). Content-driven approaches emphasize learning of content as a primary goal as in the subject matter courses taught in the second/foreign language by content specialists, not language teachers. At the center of the continuum of CBI, the equal emphasis is given to content and language and the instruction is given by both content and language specialists. At the other end of the continuum, language-driven approaches emphasize language learning as the primary course objective. Any topic or content that can motivate the language learners are selected

by the language teachers and used as authentic materials. At the content-driven end of the continuum lies "immersion," which originated from Canadian French immersion program at the elementary education level and has been adapted widely in a variety of approaches to secondary and tertiary level education in North America.

College level case studies are based on the three basic types of university level CBI, sheltered, adjunct and theme-based, and their modified versions. Since all of them provide the learners with the language support in addition to the content to some extent in distinction from the Canadian immersion program that was incorporated into the regular curriculum without additional language support at least during the full immersion, the complete immersion or the complete merging of language and content learning into one class has rarely been reported on the university level. Therefore, acquiring both knowledge and second/foreign language competence simultaneously through one content class seems, at present, to be a goal rather than is a reality at least on college level.

Nevertheless, it is noteworthy that many universities in the EFL setting including Korea are promoting partial immersion program by teaching a number of regular subject matter courses through English as part of the university curriculum. The present study focuses on the question of whether EFL immersion class at the college level is effective for both the development of English and the content knowledge by examining a university mathematics class in Korea for one semester through pre- and post tests.

II. PREVIOUS STUDIES ON CBI

1. Canadian Immersion

Canadian immersion started in 1965 in the elementary level education (Omaggio Hadley, 2000). There were three types: early, delayed and late immersion. Early immersion was conducted for three years from the second to the fourth grade, delayed immersion was conducted for two years, fourth and fifth years, and late immersion was conducted during the seventh or eighth grade. During the designated periods, the entire schooling was done completely in French, i.e., three years in early immersion, two years in delayed and one year in the late immersion. Early immersion was followed by a gradual incorporation of English-language instruction up until sixth grade, when instruction in the two languages is evenly divided. The delayed immersion was followed by a reintegration into the regular English language curriculum. The late immersion began all-French instruction after one year of core French, a daily period of language instruction. All the students who completed any one of the immersion programs chose to take several subjects in French if they desire in the next level education. In sum, the Canadian immersion program provided the three models with varying age of onset and the varying period of the program. The youngest students in the program received the longest period of immersion education and the oldest

students in the program received the shortest period of immersion by the common principle that all the instructions were provided in French, i.e., total immersion during the program. However, before and after the program, the treatment of the two languages was gradually incorporated either to the immersion or to the regular English instruction (Omaggio Hadley, 2000; Swain, 1991).

Among the large number of research results, the major findings are summarized as (1) that early immersion learners did not surpass late-immersion students in literacy skills (i.e. reading and writing), (Sternfeld, 1988; Swain & Lapkin, 1989) (2) that whereas the immersion learners generally improved in listening up to the native-speaker level, their production accuracy in speaking and writing was below the native-speaker level and (3) that their content learning was par to the students in the regular content class while their target language proficiency surpassed that of the students in the conventional program in which French was taught in a language class (Omaggio Haddley, 2000).

The first finding that the late-immersion students were equal to the early-immersion counterparts in spite of the shorter period of the program was interpreted that the older learners' cognitive maturity may be facilitating the simultaneous learning of language and content and as suggesting the promising potential of the university level immersion (Sternfeld, 1988). Regarding the second finding, the unbalance between the immersion students' listening comprehension and production proficiency in particular, Swain (1985, 1991) pointed out that the immersion students were significantly poorer than the native counterparts in grammar and the discourse/sociolinguistic appropriateness that heavily rely on grammar as its base. Discourse and sociolinguistic features involving formulaic chunks that are not grammar-based were equally well produced among the native and immersion students. Swain drew the conclusion that the immersion students had enough amount of input, mostly native teacher input with little peer input, and that what the immersion students lacked was not the comprehensible input, but the amount of interaction that pushes the learners to produce precise output. She proposed the incorporation of more output production and interaction as a means for providing output opportunities. The third finding on the success in both content learning and overall language learning that surpassed the language class inspired the broader application of immersion program in the various forms and to various age groups (Stoller, 2004).

In sum, the three major findings suggest first, that learning/teaching content knowledge and the second language through one integrated immersion class is feasible without sacrificing either area, though the grammatical accuracy in the second language production did not reach the native-speaker level. Second they suggest that the older age group may similarly or even better benefit from the immersion.

2. CBI Beyond the Elementary Level

Other than Canadian immersion studies, most of the successful studies of CBI beyond

the elementary level have been about the adjunct or sheltered model, whereas the case studies of true immersion in the university level are rare (Brinton, et al., 2006; Parkinson, 2000; Stoller, 2004).

In the continuum between the most content-driven and the most language-driven program, Brinton, Snow and Wesche (1989, 2006) distinguish three types of content-based language courses; Theme-based (the language class organized around content modules), sheltered courses (subject matter courses taught to a segregated group of second language learners), adjunct courses (learners are enrolled concurrently in two linked courses - a language course and a content course) (also from Parkinson, 2000, p. 373). Among the university level CBI case studies, the immersion French civilization classes at UCLA (Lafayette & Buscaglia, 1985), the sheltered and adjunct model at University of Ottawa (Hauptman, Wesche, & Ready, 1988) and the sheltered humanity classes in Korea (AeJin Kang, 2005) are the well-documented examples of CBI, closest to the content-driven end of the continuum on the university level.

A university level immersion was examined by Lafayette and Buscaglia (1985, cited from Sternfeld, 1988). After three French language courses, at the fourth semester the experimental students were taught French Civilization course in French, without language focus in class at all. Their French proficiency was compared with the control group students in the skill-based French language course as their fourth semester. Experimental group significantly improved in speaking, listening and writing whereas the control group did in listening and writing. Of the two groups, the experimental group outperformed the control group in speaking and the control group outperformed the experimental group in writing. It suggests that if it is appropriately prepared and timed, the college level immersion seems effective as much as conventional language class. Sternfeld (1988) suggests the further research on how many language courses are needed before the true immersion on university level. Though the study showed the students' gain in French language proficiency that is comparable to that of the students in the regular language class, the other goal of the immersion, i.e., content learning, was not measured in comparison to the equivalent content class taught in L1. Therefore, it is not clear whether language proficiency gain cost content learning or not.

A combination of sheltered instruction and an adjunct model was carried out at the University of Ottawa. Hauptman et al. (1988) reported a sheltered and adjunct psychology classes in the University of Ottawa. The students took the first semester psychology in their L1 and then some of the volunteers among them took their second semester psychology in their L2, either English or French on condition that through such a choice they are exempted from the required second semester L2 language course. The L2 psychology classes were of the two types: sheltered and adjunct. In the sheltered class, there was no more class hour to attend than in the regular L1 class but among approximately the 3 hours of class, the 30 minutes for graduate TA sessions were replaced by EFL instructors who helped L2 students separately with the language problem through the content material. In

the adjunct class, additional one and a half hours were spent with the ESL instructors. The results were that both types of the experimental subjects mastered the subject matter as much as the regular L1 students and gained their L2 proficiency as much as the 45 hour L2 language class students. It was interpreted that the CBI in the sequenced class enabled the success by making the students familiar with the class format, textbook, and etc., and the care for the language perspectives coordinated by the language instructor and the content instructor who attended each other's session was another reason for the success. In sum, in case of sheltered and adjunct model of CBI, it is concluded that if language focus is incorporated into the subject course appropriately, there seems to be a significant gain in language proficiency without sacrificing content-learning.

In one of the few university level CBI studies in Korea, AeJin Kang (2005) introduced Sheltered Instruction Observation Protocol (SIOP) model which aims at evaluating sheltered instruction (SI) class as well as guiding the SI instructors. Since it was originally devised for secondary school settings, she adapted the SIOP protocol for college setting through the instructors and the language specialists' discussions in the training sessions. Her study showed that the liberal art subjects, such as English literature and English linguistics, taught in English at a university in Korea by the instructors trained through SIOP resulted in the significant progress in speaking in general, and no improvement in writing in general after 11-12 weeks of interval between pre and post tests. However, the similar subject course, English and International Relations, taught by the non-trained instructor resulted in no improvement in speaking. The examination of the content-learning conducted through the students' self-evaluation compared the two types of classes without measuring the actual amount of learning that occurred during the semester. The class of the trained instructor resulted in higher student evaluation than the one taught by the non-trained instructor regarding content-learning. Although AeJin Kang's study aimed at comparing the trained and non-trained instructors of SI, for the perspective of the present study, it is more noteworthy that the significant improvement in speaking occurred in one of the SI classes taught by the trained instructor in such a short period of time, 11-12 weeks. However, because there has not been any objectively measured improvement in content learning in comparison to the L1 class, it is not clear whether such language improvement cost content learning or not. Therefore, studies on Korean university setting critically warrants further study.

In conclusion, among the above major previous studies of CBI, only the sheltered and adjunct models of University of Ottawa showed the positive results in both content learning and language proficiency gains. The immersion in which there is no language focus at all in the content classes has been both rarely and insufficiently examined since the content learning was not measured in comparison to L1 regular content class (Lafayette & Buscaglia, 1985). Likewise, though Korean sheltered class improved significantly in English proficiency, the proper examination of the content learning was not conducted. Therefore, further studies are needed to examine whether the content learning is not

reduced at the cost of language proficiency gain in the immersion classes on university level.

3. Processes of SLA in CBI

Regarding the mechanisms of simultaneous learning of second language and content knowledge, two theoretical positions have been proposed: interactionist view and congnitivist view. In line with Swain (1985, 1995) who emphasized the role of interaction and output modification opportunities in immersion classes, Pica (2002) observed the teacher-class interaction from the interactionist perspective. Pica's conclusion was that the exclusive reliance on the content in the theme-based CBI results in no improvement in the unacquired low salient linguistics features.

She examined the six classes in the theme-based pre-academic high intermediate ESL classes at a university based language institute. Without measuring the pre and post test differences in terms of content knowledge and language proficiency, she focused on the low salience grammatical features and found that the language focused instruction, such as meaning negotiation, form-focused intervention and metalinguistic instruction, composed only 1% of the total utterances. Regarding whether teacher-student interaction provides attention to the target language features, it was found that teachers negotiated for the already target-like form of the student utterances more than their non-target forms. By modifying their already target-like form in a slightly different context, the teachers provided the positive evidence in their clarification requests and thereby confirmed and reinforced the students' target like forms. Pica analyzed that it does not seem to have been purposefully provided but it seemed so because meaning exchange was the primary goal. Pica suggests that for the purpose of raising the students' attention to the unacquired form, the non-target forms should be modified in the teachers' negotiation so that the students can notice the difference between the target-like and non-target like forms. Though students' responses to the teacher's negotiation or recasts rarely incorporated target-like modification of their earlier non-target like production, Pica recommended recasts, since recasts did not simply incorporated the already target-like student utterances but also included the correction of the non-target forms of student utterances. In sum, the theme-based college level ESL class that Pica examined is heavily content-focused class in reality with little focus on language, although the themes of the classes, literature and film, provided the conditions for abundant and precise use of the articles and verb markings that were the target linguistic features for examination. Pica suggests the implementation of recasts and students' uptake sequence in the content class for the target linguistic features as shown by Doughty and Varela's study of middle school ESL science class (1998). The ESL science teachers' consistent provision of recasts to the non-target form of the verb tense in oral reporting and in the written reports, eliciting the students' incorporation of the correct form in their responses resulted in significant improvement in verb tense marking

after one semester in Doughty and Varela's study (1998).

Pica showed that even in a theme-based class which is taught by an ESL instructor and is closer to the language-driven end in the continuum, language-focused interaction was extremely rare. Pica's analysis accurately showed the nature of language input and interaction in CBI classes and its probable effects on grammar acquisition in support of Swain's proposal for modified output opportunities. Her studies suggests two issues: First, at the university level, even the ESL instructors rely exclusively on content in their interaction with the class, once the class has set out for content-learning, unlike the middle-school ESL science class in Doughty and Varela's study (1998). Second, the feasibility of incorporating any form of deliberate language focused interaction sequence in the university setting immersion content class seems very low unless the immersion classes are specially designed for particular target forms with additional care for language focus as in the adjunct model. In conclusion, the interactionist proposal that interaction is the mechanism for second language acquisition within CBI classes seems to predict the improvement in content-relevant language rather than low saliency grammatical features unless the instructors deliberately attempt language focused interaction.

On the other hand, from a cognitive perspective, Snow, et al. (1989) proposed that the integration of content and language instruction implied the integration of higher order thinking skills into the language classroom. This is particularly likely in cases in which the content is based on the academic curriculum. Use of higher order thinking is desirable because it can stimulate learners' interest in the content and therefore in language. More specifically, they proposed that higher order thinking skills promotes higher order language skills or advanced levels of language proficiency, to the extent that higher order thinking skills require more complex or elaborate language skills in more cognitively demanding tasks. Therefore, cognitively demanding content learning is required for pushing language proficiency forward.

Not only Snow et al. (1989), but also the cognitive theorists such as Skehan (1998) and Robinson (2001, 2003) argue that the more complex concept demands the more complex language. Robinson (2001, 2005) suggests providing the learners with the increasingly complex tasks so that the demands on language complexity increase according to the learners' IL level. Whereas Skehan (1998) and Robinson (2001) listed the task-based research findings that support the positive relationship between cognitive complexity of the task and the linguistic complexity of adult learners' IL, Snow et al. (1989) similarly proposed, mainly regarding the younger learners, that increasingly advanced levels of language should be incorporated into the content areas successively in the higher grades. They pointed out the plateau in the middle or late elementary grades in the immersion program, explaining that it is because continuously increasingly demanding content classes were not offered to them in the target language, despite the fact that the students are yet of limited language proficiency. In sum, CBI researchers and the cognitivists agreed on that the degree of complexity in content-learning or task content drives language learning.

In a similar vein, VanPatten's (1990) finding that the content-relevant vocabulary was better remembered than the function words in the listening comprehension tasks since human beings are given the limited cognitive capacity is in well accordance with Snow et al.'s (1989) classification of two language objectives in CBI classes. Snow, et al. (1989) proposed that by integrating language and content teaching, the two objectives are established: content-obligatory language (language essential to an understanding of content material) and content-compatible language (language that can be taught naturally within the context of a particular subject matter and that students require additional practice with). The first is specific genre register while the second is the more general language competence that enables the discussion of the particular topic. It is expected by both groups of researchers that content-obligatory language and content-compatible language are better acquired than general language features since the content of the class attracts attention first within the learners' limited cognitive capacity.

In conclusion, Pica's (2002) empirical finding that the interaction in the CBI classes almost exclusively relies on content area is in accordance to the claim that the content-relevant linguistic features attract learners' attention or cognitive resources and eventually are better acquired. Though the specially targeted low saliency linguistic features may not be the focus of interaction in the CBI class in general, the same linguistic features could be the focus of attention and interaction if they are content –relevant in a specially designed CBI class, in the same way as the content words or the word chunks are the focus of interaction.

III.   RESEARCH QUESTIONS

Based on the review of the previous studies, the present study focuses on the immersion on the university level which is rare, and on mathematics class. The goal of the present study is to measure both content learning and language proficiency gain in comparison to the L1 math class, in order to examine if there is any trade-off effect. The language proficiency gain to be measured is of the three kinds: the two language objectives that are closely relevant to the content area and the general language proficiency that is not directly relevant to the content of the class but has been measured in most of the previous immersion or CBI. University mathematics class was chosen because the three types of language proficiency gain are rather obviously classified in the non-humanity fields because its genre specific register is rarely used in everyday routine. Therefore, the following research questions have been formulated:

(1) Does university mathematics immersion class promote general English proficiency more than its regular L1 counterpart?

(2) Does university mathematics immersion class promote content-obligatory

English more than its regular L1 counterpart?

(3) Does university mathematics immersion class promote content-compatible English more than its regular L1 counterpart?

(4) Does university mathematics immersion class promote content learning as much as the regular L1 mathematics class?

IV. METHODS

1. Participants

The subjects of the study were one hundred and seventy students in the three intact mathematics classes in the department of mathematics at a major university in Seoul. One of the subject courses, named Set Theory, was divided into two sections due to the excessive number of students, more than 120 students, registered. Although the class was originally designed for the lecture in Korean language, as the class was divided into two, one of the two sections was decided to be given the lecture in English, according to the university policy of promoting English lecture in the subject matter courses. The two classes, lecturing in English and Korean respectively, were the target of comparison in terms of their learning in the subject area and in English. Set Theory is one of the elective courses but due to its importance, most of the second year students in the department, practically in their first year of the mathematics B.A. program, take the course. In addition to these two classes, one Linear Algebra class participated the study as a control group. Since Linear Algebra was designed for the upper class students in the university, the students in the Linear Algebra were more advanced in mathematics in general including the Set Theory and had the higher probabilities that they had taken English lecture of mathematics in the past, since the university had been offering English lectures for several years at the time of data collection. Both Set Theory and Linear Algebra have little overlapping content with the students' high school math curriculum so that the achievement in the two subject classes are hardly influenced by the students' knowledge they have already acquired in the past.

The two Set Theory classes studied the same English textbook under the same syllabus. However, their mid-term and final examinations were written and conducted separately by each professor. And the final grades were given by the percentage curve.

The three classes were taught by three different Korean professors in the department. The three professors had similar backgrounds in terms of their experience in English. They all earned their Ph.D. degrees in the United States and worked as TAs and/or lecturers for a few years in the States. However, their expertise varied, and none of them had Set Theory as their main research topic. Though the professors' teaching style may affect the results, the present study did not include the factor for analysis, considering that

mathematics class is typically lecture-driven, and the lecture is composed of solving the problems.

2. Test Materials

The three sets of pre and post tests in general English, content-relevant English, and content learning in Set Theory were developed. First, general English proficiency was measured by the shortened TOEIC practice tests developed by a commercial TOEIC practice test company, by randomly selecting the questions from listening (15 questions), structure and reading sections (15 questions). Pre and post tests were different set of the same format. Second, content-relevant English was measured by the task of translating the two Set Theoretic texts written in Korean into English. In the original English texts, content-obligatory and content-compatible English as well as general English expressions were identified. The number of expressions from each category that the student produced in his/her translation was converted to a percentage score. Third, the students' content learning in Set Theory was measured by 10 true/false mathematical questions written in English.

The professor of the control group who had taught Set Theory before wrote the math content test and made the selection of the texts for translation from outside the textbook for the course. He was also consulted for the identification of the content-obligatory, content-compatible English expressions in the texts. The pre and post test of translation and math content were the same (see Appendix).

3. Procedure

At the end of March, the three pretests were given to the students in a regular class meeting. The general English test took 20 minutes followed by content-relevant English and math content test together for another 20 minutes. The post-test was given in the beginning of June after the final exam was taken at 9 weeks' interval. One caveat is that though the pretest was given during the regular class, the post test was given during the post-final-exam class meeting when the students were confirmed with their final grades by the professor. The participation rate for the post test was lower than the pretest and especially the English test was not taken seriously by the students as shown by the number of blank answer sheets submitted to the researcher.

4. Analysis

The participants who took both pretest and post test were considered for the data analysis. Therefore, the total number of students analyzed was different in the three test sets. Also the number of students analyzed in each group was different for the same

reason. Each within-group comparison between pre and post test was conducted by a paired t test using the raw scores of the students who took both tests. However, for the between-group comparison, the multivariate GLM procedure adjusted the unbalanced number of subjects in three treatment groups from SPSS 10.0 package so that it produced the estimated mean.

General English test was scored by the number of correct items among the 30 questions. The 3 content-obligatory and 6 content-compatible English expressions were scored correct only if the students wrote the exact words or phrases, whereas the 6 general English expressions were scored correct if the students produced sensible and grammatical varieties. Mathematics test was scored by the correct number of answers among the 10 questions. The treatment (English vs. Korean lecture) was set as the independent variable, and the three sets of scores as the dependent variables. In the analysis of the post test, the pretest scores were set as a covariate so that the effect of the between-group difference in the pretest scores is controlled.

V.   RESULTS

In the results section, the three groups are named by number as Groups 1, 2, and 3 as shown below in Table 1.

TABLE 1
Groups in the Experiment

Group	Treatment	Subject and Language
Group 1	Control group	Linear Algebra in Korean
Group 2	Experimental Group	Set Theory in English
Group 3	Comparison Group	Set Theory in Korean

1. General English Proficiency

Among the 30 questions in General English test, the mean numbers of correctly answered questions by the three groups were 21.4, 20.5, and 20.2 in the pretest and 18.4, 17.8, and 18.2 in the post test as shown in Table 2 and Figure 1. There was no difference among the three groups both in the pretest and the post test as shown in Table 3. However, the paired t-test showed that the post test scores were significantly lower than those of the pre-test except in Group 3 as shown in Table 2. In sum, the math immersion was not particularly helpful to improve General English proficiency in comparison to other math classes lectured in Korean, though it needs to be explained regarding why all three groups scored lower in the post test, Groups 1 and 2 in particular.

TABLE 2
General English Proficiency: Within Group Comparison

Treatment	N	Pretest		Post test		Paired t test		
		Mean	SD	Mean	SD	t	df	Sig (2-tailed)
1	46	21.4	3.3	18.4	4.8	4.38	45	.000
2	38	20.5	3.1	17.8	3.5	5.00	37	.000
3	13	20.2	4.0	18.2	5.9	1.44	12	.176
total	97	20.9	3.3	18.1	4.5	6.41	96	.000

FIGURE 1
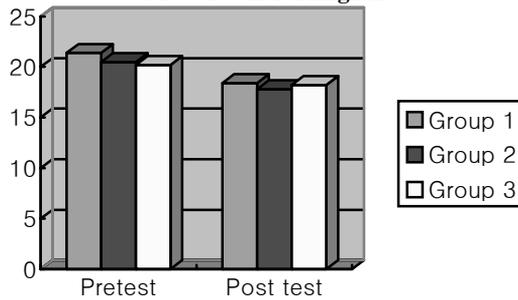**Scores of General English**

TABLE 3
General English Proficiency: Between Group Comparison

	Treatment	N	Estimated Mean	Std Error	Multivariate GLM		Pairwise comparison		
					F	Sig	treat	treat	Sig
pretest	1	44	21.49	.494	.989	.379	1	2	.193
	2	34	20.50	.569			1	3	.396
	3	10	20.50	1.05			2	3	1.00
	Total	89	20.83	.43					
post test	1	45	17.96	.693	.164	.849	1	2	.576
	2	34	18.16	.735			1	3	.961
	3	10	18.95	1.56			2	3	.759
	Total	89	18.36	.587					

2. Content-relevant English Proficiency

The content-relevant English proficiency was compared among the three groups in three ways: content-required English, content-compatible English, and the total content-relevant English that is the combination of the two. Also other than the two types of content-relevant English, general English expressions within the texts were measured separately. The items for each category are shown in Table 4.

TABLE 4
The Categories of Content-relevant and General English in Translation Test

Category	Linguistic items
Content-required	*axiom, set theory, subset (3)*
Content-compatible	*axiom of arithmetic (calculus), "A is a subset of B,"*
	set theoretic relation, relation of inclusion, relation of implication,
	If A, then B. (6)
General English	*sharp line, smallest big number or biggest small number, increase by a*
	sixth one, seem to be the same as logic, logical relation. (6)

1) Content-required English Test

Three one-word or two-word fixed expressions were identified as content-required English. The mean numbers of correctly produced expressions by the three groups were 2.1, 2.2, and 1.4 respectively in the pretest and 2.2, 2.3, and 1.6 in the post test respectively, as shown in Table 5. The paired t test shows that only Group 2 approached the significant improvement ($p=.08$) during the 9 weeks interval as shown in Table 5. The between-group comparison showed that Group 1 and 2 both scored significantly higher than Group 3 in the pretest as shown in Table 6. In the post test, however, there was no significant difference among the groups except that the difference between Group 2 and 3 approached the significant level ($p=.09$). Considering that the pretest scores were controlled as a covariate, the near significant difference between Group 2 and Group 3 in the post test is meaningful.

TABLE 5
Content-required English: Within Group Comparison

Treatment	N	Pretest		Post test		Paired t test		
		Mean	SD	Mean	SD	t	df	Sig (2-tailed)
1	44	2.09	.960	2.23	.910	-.948	43	.349
2	49	2.22	.941	2.43	.816	-1.808	48	.077
3	11	1.36	1.120	1.55	1.214	-.690	10	.506
total	104	2.08	.992	2.25	.932	-2.04	103	.044

TABLE 6
Content-required English: Between Group Comparison

	Treatment	N	Estimated Mean	Std Error	Multivariate GLM		Pairwise comparison		
					F	Sig	treat	treat	Sig
pretest	1	44	2.091	.146	3.558	.032	1	2	.508
	2	49	2.224	.138			1	3	.028
	3	11	1.364	.292			2	3	.009
	Total	104	1.893	.118					
post test	1	44	2.262	.117	1.472	.234	1	2	.711
	2	49	2.323	.110			1	3	.150
	3	11	1.878	.234			2	3	.090
	Total	104	2.154	.093					

2) Content-compatible English Test

In the translating pretest, among the 6 content-compatible expressions, the three groups produced 1.23, .88, and .45 in Groups 1, 2, and 3, respectively. In the post test, they produced 1.02, 1.06, and .64 so that Group 2 ranked the highest as shown in Table 7. The paired t test showed that only Group 2 improved approaching the significant level ($p=.07$). The between-group comparison of pretest showed that Group 1 scored significantly higher than Groups 2 and 3. However, in the post test, the three groups were not different as shown in Table 8. Though the direct between-group comparison did not show the significant difference, the paired t test results distinguishes Group 2 from other groups in that only Group 2 approached the significant level in their improvement in the post test in contrast to the other two groups.

TABLE 7
Content-compatible English: Within Group Comparison

Treatment	N	Pretest		Post test		Paired t test		
		Mean	SD	Mean	SD	t	df	Sig (2-tailed)
1	44	1.23	.859	1.02	.821	1.220	43	.229
2	49	.88	.634	1.06	.747	-1.844	48	.071
3	11	.45	.522	.64	.674	-1.000	10	.341
total	104	.98	.763	1.00	.776	-.217	103	.828

TABLE 8
Content-compatible English: Between Group Comparison

	Treatment	N	Estimated Mean	Std Error	Multivariate GLM		Pairwise comparison		
					F	Sig	treat	treat	Sig
pretest	1	44	1.227	.110	5.868	.004	1	2	.023
	2	49	.878	.104			1	3	.002
	3	11	.455	.220			2	3	.085
	Total	104	.853	.089					
post test	1	44	.988	.114	.191	.826	1	2	.761
	2	49	1.037	.107			1	3	.687
	3	11	.884	.228			2	3	.548
	Total	104	.970	.090					

3) Content-relevant English Test in Total

By combining the content-required and the content-compatible English scores, the three groups scored 3.32, 3.10, and 1.82 in the pretest in the order of Group 1, 2 and 3, respectively. However, in the post test, they scored 3.25, 3.49, and 2.18, with Group 2

ranking the top, as shown in Table 9 and Figure 2. The paired t-test shows that only Group 2 improved significantly over the interval and ranked the highest among the three groups as shown in Table 9. The between-group comparison showed that Groups 1 and 2 scored significantly higher than Group 3 in the pretest but such a difference disappeared in the post test as shown in Table 10. Although Group 2 was not superior to the other groups in the direct between-group comparison, the superiority of Group 2 is indirectly suggested by the paired t test results that distinguishes Group 2 from other groups in that only Group 2 improved significantly in the post test in contrast to the other two groups.

TABLE 9
Content-relevant English in Total: Within Group

Treatment	N	Pretest		Post test		Paired t test		
		Mean	SD	Mean	SD	t	df	Sig (2-tailed)
1	44	3.32	1.639	3.25	1.557	.238	43	.813
2	49	3.10	1.403	3.49	1.386	-2.395	48	.021
3	11	1.82	1.471	2.18	1.601	-.886	10	.397
total	104	3.06	1.563	3.25	1.518	-1.284	103	.202

FIGURE 2
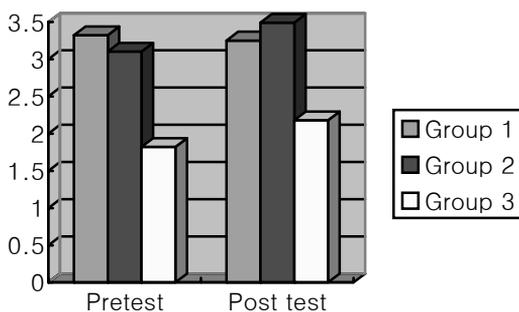**Content-relevant English**

TABLE 10
Content-relevant English in Total: Between Group Comparison

	Treatment	N	Estimated Mean	Std Error	Multivariate GLM		Pairwise comparison		
					F	Sig	treat	treat	Sig
pretest	1	44	3.318	.228	4.356	.015	1	2	.494
	2	49	3.102	.216			1	3	.004
	3	11	1.818	.457			2	3	.013
	Total	104	2.746	.185					
post test	1	44	3.250	.200	.904	.408	1	2	.697
	2	49	3.360	.188			1	3	.283
	3	11	2.762	.401			2	3	.182
	Total	104	3.124	.158					

4) General English Expressions

Among the six general English expressions, the three groups produced 2.11, 1.84, and 1.82 expressions correctly in the preset in the order of Group 1, 2, and 3. In the post test, they produced 2.00, 1.78 and 1.73 expressions correctly as shown in Table 11. The paired t test showed that none of the three groups improved in the post test as shown in Table 11. The between-group comparison showed no significant difference among the groups either in the pretest or in the post test as shown in Table 12.

TABLE 11
General English Expressions in Translation: Within Group Comparison

Treatment	N	Pretest		Post test		Paired t test		
		Mean	SD	Mean	SD	T	df	Sig (2-tailed)
1	44	2.11	1.351	2.00	1.258	.454	43	.652
2	49	1.84	1.375	1.78	1.195	.326	48	.746
3	11	1.82	1.834	1.73	1.555	.232	10	.821
total	104	1.95	1.410	1.87	1.255	.606	103	.546

TABLE 12
General English Expressions in Translation: Between Group Comparison

|	Treatment	N	Estimated Mean	Std Error	Multivariate GLM		Pairwise comparison		
					F	Sig	treat	treat	Sig
pretest	1	44	2.114	.214	.498	.610	1	2	.349
	2	49	1.837	.202			1	3	.538
	3	11	1.818	.427			2	3	.969
	Total	104	1.923	.173					
post test	1	44	2.022	.174	.896	.411	1	2	.197
	2	49	1.702	.164			1	3	.893
	3	11	1.968	.350			2	3	.495
	Total	104	1.897	.138					

3. Content-learning in Mathematics

Among the 10 true/false questions in the math test, the mean scores of the three groups were 6.4, 5.1, and 3.1 in the pretest and 7.0, 6.8, and 4.6 in the post test as shown in Table 13 and Figure 3. The paired t-test showed that all three groups improved in their math test significantly.

TABLE 13
Mathematics Content: Within Group Comparison

Treatment	N	Pretest		Post test		Paired t test		
		Mean	SD	Mean	SD	T	df	Sig (2-tailed)
1	46	6.4	1.0	7.0	1.3	-2.8	45	.008
2	49	5.1	1.3	6.8	1.3	-6.9	48	.000
3	11	3.1	1.8	4.6	1.5	-2.5	10	.034
total	106	5.5	1.6	6.7	1.5	-7.2	105	.000

FIGURE 3
Math Content-learning
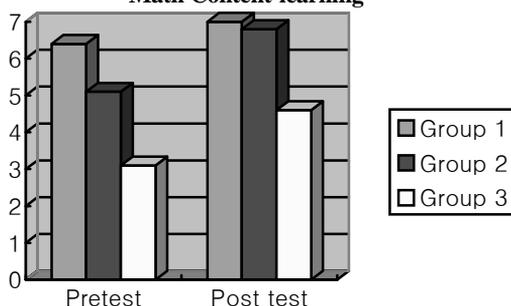
The between-group comparison by Multivariate GLM showed that the three groups were significantly different from one another in their pretest in the order of Group 1, 2, and 3 as shown in Table 14. However, in the post-test the difference between Group 1 and Group 2 disappeared due to the greatest score increase in Group 2 in their post test scores whereas the difference between Group 2 and Group 3 remained significantly. In the post test, the scores of Group 2 was the highest followed by Group 1 and then by Group 3. In consideration of the covariate analysis of the pretest scores, it can be safely interpreted that Group 1 and Group 2 learned the subject matter more than Group 3. This means that, learning the subject matter in English does not reduce the amount of the subject matter learning at all.

TABLE 14
Mathematics Content: Between Group Comparison

	Treatment	N	Estimated Mean	Std Error	Multivariate GLM		Pairwise comparison		
					F	Sig	treat	treat	Sig
pretest	1	44	6.38	.188	27.30	.000	1	2	.000
	2	34	5.15	.216			1	3	.000
	3	10	3.30	.399			2	3	.000
	Total	89	4.94	.164					
post test	1	45	6.79	.199	20.36	.000	1	2	.877
	2	34	7.22	.211			1	3	.000
	3	10	5.10	.445			2	3	.000
	Total	89	6.37	.168					

VI. DISCUSSION

The four research questions were answered as follows:

(1) Does university mathematics immersion class promote general English proficiency more than its regular L1 counterpart?   *No.*
(2) Does university mathematics immersion class promote content-obligatory English more than its regular L1 counterpart?   *Yes (indirectly through t test comparison).*
(3) Does university mathematics immersion class promote content-compatible English more than its regular L1 counterpart?   *Yes (indirectly through t test comparison).*
(4) Does university mathematics immersion class promote content learning as much as the regular L1 mathematics class?   *Yes.*

In General English Proficiency test and general English expression scores in translation, the three groups were not different in pre and post test, and none of the groups improved in the post test. Rather, in all three groups, General English test scores decreased in the post test, significantly in Group 1 and 2, which needs explanation.

In the pretest of two types of content-relevant English, Group 1 scored higher than Group 3 with a significant difference in content-required English, and with a significant difference from both Group 2 and 3 in content-compatible English. The superiority of Group 1 seems due to their relatively advanced level of math learning as well as their previous math classes taught in English, which probably showed cumulative effect over time at the starting point of the present study. However, the highest mean was achieved by Group 2 in the post test of the content-relevant English in total because only Group 2 improved in content-relevant English significantly in the post test. Though the between-group difference in the post test was not significant, the significant improvement within Group 2, no improvement within Group 1, and the medium improvement within Group 3 suggest that the longer term results might show the significant difference in the delayed post test.

The contrast between content-relevant English scores and General English proficiency scores suggests that the effect of immersion math class lies in the improvement in content-relevant English, though not in General English proficiency. The scores of the general English expressions in translation resulted exactly in the same pattern as General English Proficiency test: No between-group difference in both pre and post tests and no improvement in any of the three groups in the post test.   In comparison to the most of the previous humanity subject matter CBI classes that showed the improvement in general English, the effect of mathematics class is somewhat restricted to the content-relevant linguistic features, especially in a relatively short period of time. Therefore, genre-sensitive effect of CBI warrants further study.

In the math content test, the three groups were significantly different at the pretest in the

order of Group 1, 2, and 3. Similar to the superiority of Group 1 in content-relevant English, their superiority in math content may be explained by their relatively more advanced level of math knowledge. However, the reason why Group 2 excelled Group 3 at the pretest is not clear since the two groups were assigned according to the student number, regardless of the students' preference.[1] It is noteworthy that all three groups improved in math content test, suggesting that even Group 1 improved in math knowledge in general, probably because of the overall maturity in mathematics. What is even more noteworthy is that Group 2 improved the greatest, narrowing the gap with Group 1 and widening the gap with Group 3 so that Group 1 and 2 both were significantly better than Group 3 in the post test. Another indirect statistical index that shows the difference between Group 2 and the other groups is the p-value of the t-test: Though all the three groups improved in the post test of math content, Group 2 improved with the smallest p-value, ($p=.000$) among the three groups, which means that the greatest improvement in raw scores of Group 2 is statistically indexed as well. Therefore, at least, it can be safely concluded that math immersion class does not cause any loss in content-learning in comparison to the equivalent math class lectured in Korean language. Considering that such improvement occurred during the 9 weeks of the experiment, it could be cautiously inferred that the initial difference between Group 2 and Group 3 might be also due to the difference in treatment variable that might have already occurred during the past 4 weeks prior to the pretest which was conducted at the end of March.

In sum, the content-relevant English was improved significantly only by the immersion group and the math content-learning was significantly greater by the immersion group than by the non-immersed group. The fact that the between-group comparison in the post test of content-relevant English did not show any significant difference suggests probably that the benefit of the immersion class on content-relevant English has not been substantial yet during the 9 weeks interval enough to be confirmed by the direct between-group comparison. Math content-learning of the immersion group was significantly greater than that of the non-immersed counterpart both directly through the between-group comparison and indirectly through the t-test index of p value. On the other hand, the fact that the three groups already performed differently at the pretest of content-relevant English and math content is hardly explained than by suggesting that learning in those two areas has already begun at the outset of the semester which was 4 weeks prior to the pretest, based on the fact that the following 9 weeks' interval showed the between-group differences indirectly through t-test and directly through multivariate analysis. In contrast to content-relevant English and math content, General English proficiency test and general English expression in translation test were shown not to be affected by immersion at all. In consideration of

---

[1] Though individual factors were considered as random effect in the present study, the precise information about the individual students' background such as their experience of attending English lectures prior to the Set Theory class or their attitude toward the English lecture might be of help for the better explanation.

the shortness of the experimental period and also of the fact that there was absolutely no additional care for language objectives, even the smallest gain in content-relevant English proficiency by the immersion group sheds a promise to the immersion class because there was no loss in content learning.[2]

From a theoretical perspective, as Snow et al. (1989), Skehan (1998), and Robinson (2001, 2005) suggest, language learning seemed to have occurred as a result of academic content learning which requires the content-relevant language. Little change in General English proficiency suggests that English proficiency gains occur probably in the order of content-relevant to general, taking more time in the order.

The success in the simultaneous learning of both may depend on how the students are prepared for the immersion class, which Sternfeld (1988) suggests as an issue subject to empirical studies. The present study did not collect the in-depth data on the students' individual or school based background in English education. At present, as Snow et. al (1989) suggest, timely provided content-driven CBI seems not only feasible but also necessary for advanced level proficiency in language, without compromising the content learning.

Regarding the factors that enabled such a positive result, a couple of suggestions are proposed as the additional topics for the further study: first, the type and the amount of language courses the students took prior to the immersion, and second, the students' study hours outside the class in the three groups as a factor secondary to the students' attitude and motivation which could also be influenced by the treatment variable.

VII. FINAL REMARKS

One of the caveats of the present study is the fact that the general English test scores decreased significantly in the post test. The post test was given to the students on June 4, when the courses were actually over and the instructor met the students to confirm their final grades. The class meeting was not as serious as other regular class in terms of students' attention. Though the analysis of the present study was conducted because all three groups showed the consistent pattern in the post test of General English Proficiency, a better controlled study is required in the future.[3]

---

[2] For more valid interpretation of the results, as a way of examining the test reliability, Pearson correlation coefficient was produced for the pre and post test pairs of all the three tests - English, math content, and translation. The results showed that all three pairs showed strongly significant correlation as .441 ($p=.000$), .427 ($p=.000$), and .529 ($p=.000$).

[3] It is assumed that the students did not take the post tests seriously both in English and Set Theory. However, the consistent drop in all three groups in the English test suggests that all three groups were equally uncooperative and such a factor was revealed by the strong correlation between pre and post tests of all three tests. Therefore, the maximum interpretation of the post test results was conducted under the assumption that the three groups are under the same condition at the time of post test.

Another caveat is that the math content test was provided only in English. It could raise the possibility that the content learning has been measured in favor of the experimental group. For the more objective examination, the test should be provided in two parts, half in Korean and half in English. For the present study, the researcher and one of the participating math professors, the instructor of the Linear Algebra who wrote the math content test, confirmed that the math content test does not include rarely used or heavy loaded mathematical jargons, or complex syntax so that the students may understand the math questions as clearly and easily as written in Korean. Similarly, content-relevant English may be measured not only through translation, but also through writing the answer to the math question in English for the more precise and direct access to the students' English ability in their own academic field.

Lastly, since the math content classes typically have very little teacher-class interaction, the interaction data has not been observed. Since the interaction is proposed as one of the major mechanisms for the success of immersion program, studies that incorporate the interaction effect would be needed in the future.

Immersion receives great amount of attention recently in EFL countries throughout all age group EFL learners. However, in each level, it is conducted in a short period without a long-term institutionalized program. Though the present study examined the university level only, with the promising results, it is not clear what previous preparation of the students enabled it and whether the seed of effect shown in the present study could continue for more substantial results. Through a thorough conduct of test battery, a long-term large-scale systematic observation is strongly needed in order to examine what combination or sequence of language and content-based instruction best facilitates the advanced level language proficiency that enables the students to communicate in the expertise knowledge area.

REFERENCES

Brinton, D., Snow, M., & Wesche, M. (1989). *Content-based second language instruction*. Boston: Heinle & Heinle.

Brinton, D., Snow, M., & Wesche, M. (2006). *Content-based second language instruction*. Ann Arbor: The University of Michigan Press.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). Cambridge: Cambridge University Press.

Hauptman, P., Wesche, M., & Ready, D. (1988). Second-language acquisition through subject-matter learning: A follow-up study at the University of Ottawa. *Language Learning, 38*(3)*,* 433-475.

Kang, AeJin. (2005). How to promote comprehension and participation in CBI courses:

The SIOP model. *English Teaching, 60*(4), 159-196.

Lafayett, R. C., & Buscaglia, M. (1985). Students learn language via a civilization course: A comparison of second language classroom environments. *Studies in Second Language Acquisition, 7*(2), 323-342.

Met, M. (1991). Learning language through content: Learning content through language. *Foreign Language Annals, 24*(2), 281-295.

Met, M. (2004). Content-based instruction. In M. Byram (Ed.), *Routledge encyclopedia of language teaching and learning* (pp. 137-140). UK: Routledge

Omaggio Hadley, A. (2000). *Teaching language in context.* Boston: Heinle & Heinle.

Parkinson, J. (2000). Acquiring scientific literacy through content and genre: A theme-based language course for science students. *English for Specific Purposes, 19,* 369-387.

Pica, T. (2002). Subject-matter content: How does it assist the interactional and linguistic needs of classroom language learners? *The Modern Language Journal, 86*(1), 1-19.

Robinson, P. (2001) Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge: Cambridge University Press.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. Long (Eds.), *The Handbook of second language acquisition* (pp. 631-678). Malden, MA: Blackwelll Publishing.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics, 43*(1), 1-32.

Skehan, P. (1998). *A Cognitive approach to language learning*. Oxford: Oxford University Press.

Snow, C., Met, M., & Genesse, F. (1989). A conceptual framework for the integration of language and content in second/foreign language instruction. *TESOL Quarterly, 23*(2), 201-219.

Sternfeld, S. (1988). The applicability of the immersion approach to college foreign language instruction. *Foreign Language Annals, 21*(3), 221-226.

Stoller, F. (2004). Content-based instruction: Perspectives on curriculum planning. *Annual Review of Applied Linguistics, 24*(3), 261-283.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). Rowley, MA: Newbury House Publishers.

Swain, M. (1991). French immersion and its offshoots: Getting two for one. In B. Freed (Ed.), *Foreign language acquisition research and the classroom* (pp. 91-103). Lexington, MA: D. C. Heath and Company.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Siedhofer (Eds.), *Principle and practice in applied linguistics: studies in honor of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1989). Canadian immersion and adult second language teaching: what's the connection? *The Modern Language Journal, 73*(2), 150-159.

VanPatten, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second language Acquisition, 12*(3), 287-301.

APPENDIX
Test of Mathematics and Translation

ywk	TEST
050328	
St. Number:	Name:

◎ **(1-10)** Read each statement and mark True(**T**) or False(**F**).

(1) Every nonempty subset of $\mathbf{Z}^{+}$ has a smallest element, where $\mathbf{Z}^{+}$ is the set of positive integers. (   )

(2) $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$. (   )

(3) A countable union of countable sets is countable. (   )

(4) $\mathbf{R}$ is countable, where $\mathbf{R}$ is the set of real numbers. (   )

(5) If A is a set, then there exists an order relation on A which is a well-ordering. (   )

(6) There exists a set A such that there is an injective map $f : P(A) \to A$, where $P(A)$ is the power set of A. (   )

(7) Let $\sim$ be an equivalence relation defined on X. There may exist equivalence classes G_x and G_y such that $G_x \neq G_y$ and $G_x \cap G_y = \varnothing$. (   )

(8) The power set of a countable set is countable. (   )

(9) The class of natural numbers, ordered in the usual way, is a well-ordered class. (   )

(10) There is an inductive set which does not have a maximal element. (   )

◎ Translate the following paragraphs from Korean into English:

● 처음에 집합론은 논리학과 같은 것처럼 보였다. `$A$ 는 $B$ 의 부분 집합이다.'라는 집합론적인 포함 관계는 `A 이면 B 이다.'라는 논리학적 함의 관계와 같다.

At first, set theory seemed to be the same as logic. The set theoretic relation of inclusion, "A is a subset of B," is the same as the logical relation of implication, "If A, then B."

Hersh, R. (1997). What is mathematics, really? (p. 76). New York: Oxford University Press; Korean translation by Min Her (2003). Kyung Moon.

● 작고 큼을 구분하는 뚜렷한 경계선은 없다. 가장 작은 큰 수 또는 가장 큰 작은

수는 없다. 당신은 스스로 선택한 *M* 을 사용할 수 있고, 산술에 관한 페아노
(Peano)의 다섯 개의 공리에 다음을 여섯째 공리로 추가할 수 있다.

There is no sharp line between small and large, no smallest big number or biggest small
number. You could use the *M* you just picked, and increase Peano's five axioms of arithmetic
by a sixth one.

Hersh, R. (1997). What is mathematics, really? (p. 147). New York: Oxford University Press;
Korean translation by Min Her (2003). Kyung Moon.

Haemoon Lee
Dept. of English Language and Literature
Sungkyunkwan University
3-53, Myungryun-dong, Jongro-gu
Seoul 110-745, Korea
Email: haemoon@skku.edu