

영어강의의 효과성에 대한 연구[†]

조장식¹

¹경성대학교 정보통계학과

접수 2012년 10월 18일, 수정 2012년 11월 8일, 게재확정 2012년 11월 13일

요약

요즘 대부분의 대학들이 국제화를 목표로 재학생들의 어학능력 및 국제적 경쟁력 강화와 함께 영어강의의 중요성이 증대되고 있다. 본 연구에서는 강의평가점수를 이용해서 과목특성 변수들과 개인특성 변수들 별로 영어강의와 한국어강의의 효과성을 비교하였다. 또한 로지스틱회귀분석과 의사결정나무분석을 이용하여 어떤 요인들이 영어강의가 한국어강의에 비해서 효과적인지를 주요 과와 상호작용효과 측면에서 분석하였다. 분석결과에 따르면 영어강의의 효과성에 영향을 미치는 변수로는 학년, 계열, 강좌규모, 평균평점, 계열, 전형방법 등으로 나타났다. 또한 영어강의 효과성이 특히 높은 그룹은 1학년이면서 인문계열인 경우, 그리고 1학년이면서 자연 및 예체능계열이고 평균평점이 높은 그룹이 높게 나타났다. 그리고 영어강의의 효과성 비율이 특히 낮은 그룹은 2-3학년 학생이면서 강좌규모가 크고 인문계열인 경우로 나타났다.

주요용어: 데이터마이닝, 로지스틱 회귀분석, 상호작용효과, 영어강의, 의사결정나무분석, 주효과.

1. 서론

최근 각 대학들은 재학생들의 어학능력 및 국제적 경쟁력 강화와 함께 대학의 국제화를 목표로 영어강의 비율을 점차 늘리고 있다. 이를 통해 글로벌 인재에 걸맞은 영어능력이 뛰어난 학생들을 양성하고자 한다. 그러나 현재 국내의 여러 대학에서 영어강의의 효과에 대한 의문이 제기되고 있다. Han 등 (2010)에 의하면, 학생들의 관점에서는 강의내용이 제대로 이해되지 않으며 한국어로 강의할 때 보다 진행속도가 느리다는 의견이 있고, 교수들에게는 영어구사력으로 인하여 수업준비의 어려움이 있다는 의견이 있다. Horwitz 등 (1986)은 성인들은 외국어 학습 자체를 부담스런 경험으로 인식하기 때문에 자발적인 학습참여를 소홀히 하고, 또 실수하지 않으려는 압박감 때문에 새로운 언어형태를 시도하지 않으려는 경향이 있다고 하였다. 그리고 Klassen과 Graaff (2001)는 내용전달의 비효율성, 비언어적 표현법, 학생과의 관계형성의 어려움, 개념형성의 어려움 등을 언급하고 있다. 또한 Lee (2007)은 영어강의 수업을 수강한 학생들을 통해 학생들이 인지하는 영어강의의 효과를 연구하였다. 그리고 Lee (2010)은 대학생들의 영어원어강의에 대한 인식조사를 한 바 있다.

한편 본 연구에서는 영어강의와 한국어강의 교과목들의 강의평가점수를 비교하고, 어떤 요인들이 영어강의의 효과성에 영향을 미치는지를 분석하고자 한다. 일반적으로 영향력 분석에는 판별분석과 로지스틱회귀분석, 그리고 데이터마이닝기법의 의사결정나무분석 등이 연구방법으로 주로 사용되고 있다. 로지스틱 회귀분석은 판별분석에 비해 요구되는 기본가정이 엄격하지 않으며, 다른 분석과 결합하여 다양한 정보를 줄 수 있는 확장성을 지니고 있다. 또한 의사결정나무분석은 예측과 분류에서

[†] 이 논문은 2012학년도 경성대학교 학술연구비지원에 의해 연구되었음.

¹ (608-736) 부산광역시 남구 수영로 309 번지, 경성대학교 이과대학 정보통계학과, 교수.
E-mail: jscho@ks.ac.kr

다른 데이터마이닝기법에 비해 유용성이 높게 평가되고 있으며 다변량분석의 기본가정인 선형성, 정규분포성, 등분산성이 요구되지 않는다. 의사결정나무분석을 활용한 연구로는 Lee와 Hur (2009)과 You와 Park (2009), Cho (2010, 2011) 등이 있다. 따라서 본 연구에서는 로지스틱회귀분석과 의사결정나무분석을 이용하여 영어강의의 효과성에 미치는 영향력을 주효과와 상호작용효과 관점에서 분석하고자 한다.

본 연구의 구성은 다음과 같다. 제 2절에서는 본 연구에 사용되는 자료와 변수를 소개하고, 제 3절에서는 영어강의와 한국어 강의에 대한 평균 강의평가점수를 비교하였다. 제 4절에서는 회귀분석과 의사결정나무분석을 통하여 영어강의의 효과성에 미치는 영향력을 분석하였고, 마지막으로 제 5장에서는 본 연구에 대한 결론을 제시하였다.

2. 자료 및 변수설명

본 연구에 사용된 자료는 모 대학의 1년간 개설된 강좌별 수강생들의 강의평가점수를 포함한 과목특성과 인적특성과 관련된 자료이다. 자료는 한 학생이 여러 교과목을 수강신청을 할 수 있으므로 기본적으로 중복케이스로 구성되어 있다. 본 연구에서는 수강생의 입학년도가 2007년도 이후이고 교과목 개설학년이 1학년부터 3학년까지 자료만 분석에 사용하였다. 왜냐하면 4학년 학생은 1~3학년에 비해서 상대적으로 소수의 교과목을 수강 신청할 뿐만 아니라 강의평가도 저학년에 비해서 상대적으로 후하게 주는 편이기 때문에 분석에서 제외하였다 (Cho 등, 2009). 데이터에 포함된 변수와 기술통계의 결과는 아래 Table 2.1과 같다.

Table 2.1 Descriptive statistic

| variables | categories | categories % | E.class % | variables | categories | categories % | E.class % |
|--------------|------------|--------------|-----------------|----------------|--------------------|--------------|-----------|
| sex | male | 23.5 | 1.3 | double major | yes | 10.4 | 2.0 |
| | female | 76.5 | 1.8 | | no | 89.6 | 1.6 |
| grade | freshmen | 61.7 | 1.2 | changing dept. | yes | 2.1 | 1.6 |
| | sophomore | 27.0 | 2.5 | | no | 97.9 | 1.7 |
| | junior | 11.4 | 2.3 | lecture type | online | 1.9 | 3.3 |
| class size | < 10 | 5.4 | 3.9 | | offline | 98.0 | 1.6 |
| | 10~20 | 19.4 | 2.7 | entrance year | 2007 | 25.7 | 1.7 |
| | 20~30 | 27.5 | 1.6 | | 2008 | 32.9 | 1.9 |
| | 30~40 | 23.0 | 1.5 | | 2009 | 23.9 | 1.5 |
| | 40~50 | 14.8 | .4 | | 2010 | 17.5 | 1.2 |
| > 50 | 9.9 | .8 | evaluation type | relative | 40.1 | .3 | |
| subject type | major | 60.4 | | 2.5 | absolute | 59.9 | 2.6 |
| | credit | liberal art | 37.9 | .4 | scheduled entrance | yes | 45.8 |
| A | | 35.9 | 2.0 | no | | 54.2 | 1.7 |
| B | | 37.3 | 1.4 | dept. category | humanity | 49.5 | 2.3 |
| C | | 17.3 | 1.4 | | nature | 27.1 | 1.1 |
| D | | 5.1 | 1.5 | | art | 23.5 | 1.0 |
| | F | 4.4 | 2.2 | total | | 165,246 | |

1) E.class means english class.

위의 결과로부터 개설된 교과목에 대해서 강의평가 총 건수는 165,246개이며, 특성변수들의 범주별로 영어강의 (E.class)의 수강비율이 상대적으로 높은 특성을 살펴보면 다음과 같다. 성별 (sex)로는 남학생 (male)의 영어강의 수강비율에 비해서 여학생 (female)의 비율이 상대적으로 높게 나타났으며, 개설학년 (grade)별로는 1학년 (freshmen)의 영어강의 수강비율에 비해서 2학년 (sophomore)과 3학년 (junior) 학생의 수강비율이 높게 나타났다. 강의규모 (class size)는 작을수록 높게 나타났고, 과목유형 (subject type)별로는 교양과목 (liberal art)에 비해서 전공과목 (major)이, 학점 (credit)이 대체로 높을수록 영어강의의 수강비율이 높게 나타났다. 그리고 복수전공 (double major)을 하는 학생이 하지 않는 학생에 비해서 높게 나타났고, 강의형태 (lecture type)가 온라인(online)인 경우가

오프라인 (offline)인 경우보다 높게 나타났으며, 입학년도 (entrance year)가 빠를수록, 그리고 평가 방법 (evaluation type)이 상대평가 (relative)에 비해서 절대평가 (absolute)가 높게 나타났다. 그리고 계열 (dept. category)이 인문계열 (humanity)인 경우가 자연계열 (nature) 또는 예체능계열 (art)에 비해서 영어강의 수강비율이 상대적으로 높게 나타났다.

3. 영어강의와 한국어강의 비교

본 연구에서는 영어강의의 효과성을 영어강의 교과목의 평균 강의평가점수가 한국어강의 교과목의 평균 강의평가점수에 비해서 높은 경우로 정의한다. 이 경우 타당하게 영어강의의 효과성을 분석하기 위해서는 수강학생별로 영어강의 과목들의 평균 강의평가점수와 한국어강의 과목들의 평균 강의평가점수를 대응비교를 해야 한다. 왜냐하면 2장에서 사용된 자료는 한 학생이 여러 교과목을 수강신청 하므로 중복케이스를 가지는 데이터 구조로써 케이스의 독립성을 요구하는 일반적인 분석방법으로는 분석할 수가 없다. 또한 영어강의를 수강한 학생들과 한국어강의를 수강한 학생들을 과목별 독립적으로 비교한다면 학생들의 특성 (예를 들어 학점, 학년, 강좌규모, 교과목구분 등) 들을 효과적으로 통제하기 어렵다.

따라서 교과목별로 수강학생들의 중복 케이스를 포함하는 자료로부터 수강학생별로 영어강의 교과목들과 한국어강의 교과목들의 과목특성 변수들로 자료통합을 통해서 재구성하였다.

아래 Table 3.1은 특성변수별로 한국어강의 과목들의 평균 강의평가점수와 영어강의 과목들의 평균 강의평가점수의 상관분석과 대응비교 t-검정 결과를 나타낸 결과이다. 여기서 *, **, ***는 각각 유의수준 0.10, 0.05, 0.01에서 통계적으로 유의함을 의미한다.

Table 3.1 The results of correlation analysis and pairwise t-test

| variables | categories | n | evaluation score | | difference | r | pairwise t-test |
|------------------|-----------------|-------|------------------|---------|------------|--------|-----------------|
| | | | E_score | K_score | | | |
| sex | male | 360 | 4.06 | 3.90 | 0.16 | .53*** | 3.79*** |
| | female | 1,520 | 3.96 | 3.82 | 0.14 | .48*** | 8.12*** |
| grade | freshman | 865 | 3.99 | 3.80 | 0.19 | .48*** | 7.68*** |
| | sophomore | 696 | 3.94 | 3.85 | 0.09 | .49*** | 3.33*** |
| | junior | 319 | 4.05 | 3.91 | 0.14 | .60*** | 4.41*** |
| dept. category | humanity | 1,248 | 4.00 | 3.84 | 0.16 | .50*** | 7.93*** |
| | nature | 366 | 3.96 | 3.80 | 0.16 | .52*** | 4.22*** |
| | art | 266 | 3.92 | 3.83 | 0.09 | .45*** | 1.92* |
| changing dept. | yes | 42 | 4.03 | 3.93 | 0.10 | .36** | .98 |
| | no | 1,838 | 3.98 | 3.83 | 0.15 | .50*** | 8.90*** |
| screening method | SAT | 242 | 4.03 | 3.81 | 0.22 | .48*** | 5.35*** |
| | general | 561 | 3.95 | 3.82 | 0.13 | .53*** | 4.35*** |
| | academic school | 468 | 3.99 | 3.82 | 0.17 | .49*** | 5.47*** |
| entrance year | others | 609 | 3.97 | 3.87 | 0.10 | .48*** | 3.52*** |
| | 2008 or before | 1,166 | 3.98 | 3.83 | 0.15 | .50*** | 7.11*** |
| | 2009 or later | 714 | 3.98 | 3.84 | 0.14 | .49*** | 5.44*** |
| total | | 1,880 | 3.98 | 3.83 | 0.15 | .50*** | 8.95*** |

1) E_score means the average evaluation score of english class and K_score means the average evaluation score of korean class.

2) SAT means screening method by scholastic aptitude test.

3) * : p-value < 0.10, ** : p-value < 0.05, *** : p-value < 0.01.

위의 Table 3.1의 결과에서 모든 특성변수에 대해서 수강생들의 영어강의 과목들과 한국어강의 과목들의 평균 강의평가점수 간에는 통계적으로 유의한 양의 상관관계가 존재함을 알 수 있다. 즉, 한국어강의 과목들의 평균 강의평가점수가 높을수록 통계적으로 유의하게 영어강의 과목들의 평균 강의평가점수도 높음을 알 수 있다. 그리고 전과여부 (changing dept.)를 제외한 모든 특성변수들에 대해서 한국어강의 과목들과 영어강의 과목들의 평균 강의평가 평균점수간에는 통계적으로 유의한 차이가 존재함을 알 수 있다. 즉, 모든 특성변수에 대해서 영어강의 교과목들의 평균 강의평가

점수가 한국어강의 교과목들의 평균 강의평가점수에 비해서 통계적으로 유의하게 높음을 알 수 있다.

4. 회귀분석과 의사결정나무분석

3절의 Table 3.1은 수강생별로 영어강의 과목들과 한국어강의 과목들에 대한 평균 강의평가점수를 단순 비교한 결과로써 다양한 특성변수들(독립변수)이 반영되지 않았고, 또한 특성변수들의 상호작용 효과가 반영되지 않았다는 점에서 한계가 있다.

따라서 본 절에서는 모수적인 방법인 로지스틱 회귀분석을 통해서 영어강의의 효과성을 특성변수들이 포함시킨 상태에서 주효과를 분석하고자 한다. 그러나 회귀분석 방법은 엄격한 가정만 아니라, 다양한 범주형 독립변수들로 구성된 경우 이들에 대한 고차의 상호작용효과를 분석하는데는 다소 한계가 있다. 이를 보완하기 위해서 비모수적인 방법인 의사결정나무 분석을 통하여 특성변수들에 대한 고차의 상호작용효과가 영어강의의 효과성에 미치는 영향력을 분석하고자 한다.

로지스틱 회귀분석과 의사결정나무분석에 사용될 변수목록은 아래 Table 4.1에 제시하였다.

Table 4.1 List of variables

| variables | definition | scale |
|------------------|--|------------|
| effect* | 0=E_score is less than K_score*, 1=otherwise | nominal |
| sex | 0=female*, 1=male | nominal |
| grade | 0=freshman*, 1=sophomore, 2=junior | ordinal |
| dept. category | 0=art*, 1=humanity, 2=nature | nominal |
| changing dept. | 0=no*, 1=yes | nominal |
| class size | the number of students in class | continuous |
| GPA | total GPA | continuous |
| entrance year | 0=2008 or before*, 1=2009 or later | ordinal |
| screening method | 0=others*, 1=SAT, 2=general, 3=academic school | nominal |

1) * means reference category.

Table 4.1에서 영어강의의 효과성(effect)을 나타내는 종속변수는 effect로서 그 값이 1인 경우가 영어강의 교과목들의 평균 강의평가점수가 한국어강의 교과목들의 평균 강의평가점수보다 높게 나타나는 경우이다. 그리고 독립변수로는 계열(dept. category), 성별(sex), 전형방법(screening method), 입학연도(entrance year), 평균평점(GPA), 학년(grade), 전과여부(changing dept.), 강의규모(class size) 등이 있다.

먼저 Table 4.1에서 영어강의의 효과성(effect)에 영향을 미치는 독립변수들의 주효과를 분석하기 위해 로지스틱 회귀분석을 실시한 결과 아래 Table 4.2와 같다.

Table 4.2 Result of logistic regression analysis

| variables | B | S.E. | Wals | p-value | Exp(B) | |
|--------------------------|-----------|---------|------|---------|--------|------|
| male dummy | .05 | .13 | .12 | .72 | 1.05 | |
| grade | sophomore | -.49*** | .11 | 19.86 | .00 | .61 |
| | junior | -.31** | .15 | 4.05 | .04 | .73 |
| dept. category | humanity | .04 | .15 | .07 | .79 | 1.04 |
| | nature | -.08 | .18 | .23 | .63 | .92 |
| changing dept. dummy | .11 | .33 | .12 | .73 | 1.12 | |
| class size | -.01* | .00 | 3.66 | .06 | .99 | |
| GPA | .18** | .09 | 4.17 | .04 | 1.20 | |
| entrance year 2009 dummy | -.04 | .11 | .12 | .73 | .96 | |
| screening method | SAT | .23 | .16 | 1.95 | .16 | 1.25 |
| | general | .10 | .12 | .64 | .42 | 1.11 |
| | academic | .24* | .13 | 3.29 | .07 | 1.27 |
| constant | .09 | .35 | .07 | .80 | 1.09 | |

n=1,880 / $R^2 = .023$ / $\chi^2 = 31.731***$ / d.f.=12

1) GPA means grade point average

위의 Table 4.2의 분석결과를 살펴보면, 영어강의의 효과성에 통계적으로 유의하게 영향을 미치는 변수로는 학년 (grade), 강좌규모 (class size), 평균평점 (GPA), 전형방법 (screening method) 등으로 나타났다. 먼저 학년 (grade)별로는 1학년에 비해서 2학년 (sophomore)과 3학년 (junior)이 통계적으로 유의하게 효과적이지 않은 것으로 나타났다. 즉, 한국어강의에 비해서 영어강의의 효과성이 1학년이 가장 높게 나타났으며, 2, 3학년의 효과성이 떨어짐을 알 수 있다. 그리고 강좌규모 (class size)가 클수록 영어강의의 효과성이 통계적으로 유의하게 낮아짐을 알 수 있고, 평균평점 (GPA)이 높을수록 영어강의의 효과성이 통계적으로 유의하게 높아짐을 알 수 있다. 전형방법 (screening method)별로는 기타전형 (other)에 비해서 인문계고 전형 (academic)이 통계적으로 유의하게 영어강의의 효과성이 높음을 알 수 있다.

다음으로는 비모수적인 방법인 의사결정나무 분석을 이용하여 독립변수들에 대한 고차의 상호작용효과가 영어강의의 효과성에 미치는 영향력을 분석하고자 한다.

의사결정나무분석을 수행하기 위해 지니지수 (Gini index)를 분리기준으로 사용하였으며, 이진분리를 수행하는 CART (classification and regression trees; Breiman 등, 1984) 알고리즘을 사용하였다. 정지규칙으로는 최대나무깊이 (maximum tree depth=3), 최소 케이스 수 (minimum number of cases: parent node = 50, child node=1)로 설정하였으며, 가지치기(pruning)를 병행하였다.

아래 Figure 4.1은 의사결정나무분석의 결과를 나타낸다.

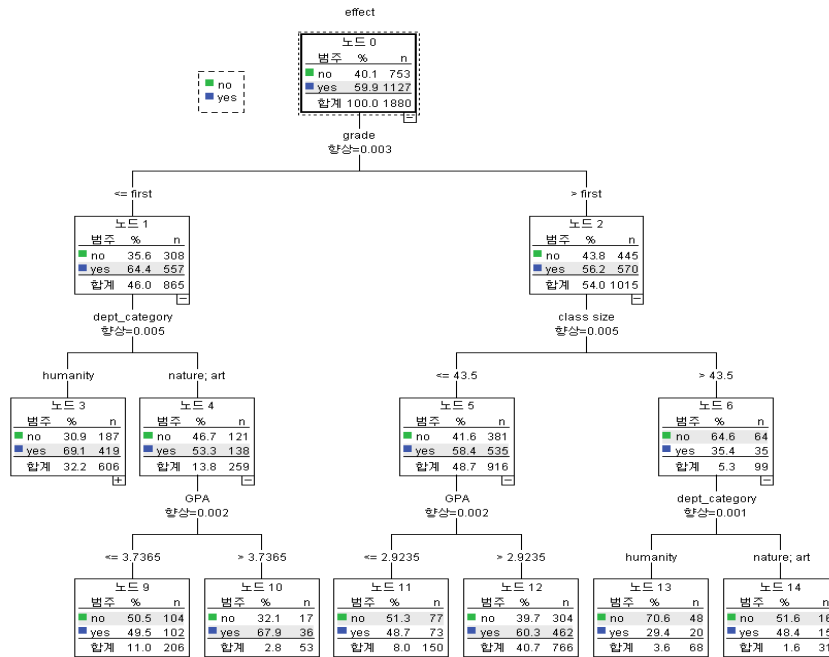


Figure 4.1 Result of decision tree analysis

Figure 4.1로부터 영어강의의 효과성에 가장 많은 영향을 미치는 변수는 학년임을 알 수 있다. 그 다음으로 영향을 미치는 변수들로는 계열, 강좌규모, 평균평점의 순으로 나타났다. 전체 1,880명의 수강생들 중 59.9% (1,127명/1,880명)가 영어강의가 효과적이었으며, 그 중 1학년생의 64.4%가 2,

3학년생의 56.2%에 비해서 영어강의가 더 효과적인 것으로 나타났다. 또한 1학년 학생 중에서는 인문계열이 69.1%로 예체능계열과 자연계열의 53.3%보다 영어강의의 효과성이 상대적으로 높게 나타났다. 그리고 예체능 및 자연계열 중에서는 평균평점이 3.74보다 높은 그룹이 67.9%로 평균평점이 3.74보다 낮은 그룹의 49.5%보다 영어강의의 효과성이 높은 것으로 나타났다.

한편 2, 3학년 학생 중에서는 강좌규모가 43.5명 이하 그룹이 58.4%로 강좌규모가 43.5명 초과 그룹의 35.4%보다 영어강의가 더 효과적인 것으로 나타났다. 또한 강좌규모가 43.5명 이하 그룹 중에서는 평균평점이 2.92를 초과하는 그룹이 60.3%로 평균평점 2.92 이하 그룹의 48.7%보다 높게 나타났다. 그리고 강좌규모가 43.5명을 초과하는 그룹 중에서는 자연계열과 예체능계열 그룹이 48.4%로 인문계열의 29.4%보다 상대적으로 높게 나타나 영어강의가 더 효과적임을 알 수 있다. 이상의 결과로부터 독립변수의 상호작용효과에 의한 영어강의 효과성 비율이 특히 높은 그룹의 특성은 1학년이면서 인문계열인 경우 69.1%로 높게 나타났으며, 1학년이면서 자연 및 예체능계열이고 평균평점이 3.74를 초과하는 그룹이 67.9%로 높게 나타났다. 한편 영어강의의 효과성 비율이 특히 낮은 그룹은 2-3학년 학생이면서 강좌규모가 43.5명을 초과하고 인문계열인 경우 29.4%로 나타났다.

이와 같이 상호작용효과에 초점을 맞춰 분석하는 의사결정나무분석의 결과는 주효과를 주로 분석하는 회귀분석과는 근본적으로 차이가 있지만, 상호 보완적으로 해석을 한다면 보다 유익한 분석 결과를 얻을 수 있다. 위의 결과에서 알 수 있듯이, 의사결정나무분석의 결과는 대체로 회귀분석의 결과와 유사한 결과를 보임을 알 수 있다.

5. 결론

본 논문에서는 모 대학에서 개설되는 과목특성 변수와 수강생들의 개인특성 변수 등이 한국어강의에 비해서 영어강의의 효과성에 미치는 영향력을 분석하기 위해 모수적인 방법인 로지스틱 회귀분석과 비모수적인 방법인 의사결정나무분석을 통해서 분석하였다. 즉 독립변수들의 주효과를 분석하는 로지스틱 회귀분석과 회귀분석 결과에 대한 보완적인 방법으로 독립변수들의 상호작용 효과를 분석하기 위한 의사결정나무모형을 사용 하였다. 분석결과를 요약하면 다음과 같다.

먼저 회귀분석결과 영어강의의 효과성에 통계적으로 유의하게 영향을 미치는 변수로는 학년, 강좌규모, 평균평점, 전형방법 등으로 나타났다. 학년별로는 1학년에 비해서 2학년과 3학년이 상대적으로 효과적이지 않은 것으로 나타났다. 그리고 강좌규모가 클수록 영어강의의 효과성이 낮아짐을 알 수 있고, 평균평점이 높을수록 영어강의의 효과성이 높아짐을 알 수 있다. 전형방법별로는 기타전형에 비해서 인문계고 전형이 영어강의의 효과성이 높음을 알 수 있다.

다음으로 의사결정나무분석 결과에서는 영어강의의 효과성에 가장 많은 영향을 미치는 변수는 학년이고, 그 다음으로 계열, 강좌규모, 평균평점의 순으로 나타났다. 영어강의 효과성 비율이 특히 높은 그룹의 특성은 1학년이면서 인문계열인 경우와, 1학년이면서 자연 및 예체능계열이고 평균평점이 3.74를 초과하는 그룹으로 나타났다. 그리고 영어강의의 효과성 비율이 특히 낮은 그룹은 2-3학년 학생이면서 강좌규모가 43.5명을 초과하고 인문계열인 경우로 나타났다.

이상의 결과로부터 학사관련 담당부서에서는 과목특성과 개인특성을 바탕으로 영어강의 교과목 개설을 확대하기 위한 기초자료로 활용되었으면 한다.

본 연구에서는 수강생들의 학업의지, 어학능력 및 자격증과 같은 개인 스펙 및 가정환경 변수 등 중요한 변수가 누락된 상태에서 수집 가능한 제한된 자료만을 사용하여 영어강의의 효과성에 대한 영향력을 분석했다는 점에서 연구결과를 지나치게 일반화 하는 것은 한계가 있다는 점을 분명히 밝힌다.

참고문헌

- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and regression trees*, Wadsworth, Belmont.
- Cho, J. S., Kang, C. W and Choi, S. B. (2009). Comparison on equating methods for course evaluation. *Journal of the Korean Data & Information Science Society*, **20**, 65-75.
- Cho, J. S. (2010). The influence analysis of admission variables on academic achievements. *Journal of the Korean Data & Information Science Society*, **21**, 729-736.
- Cho, J. S. (2011). Determinants of job finding using student's characteristic information. *Journal of the Korean Data & Information Science Society*, **22**, 849-856.
- Han, K. H., Heo, J. H. and Yun, I. G. (2010). The role and task of english medium instruction(EMI) for educating global engineers. *Journal of Engineering Education Research*, **13**, 53-59.
- Horwitz, E. K., Horwitz, M. B. and Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, **72**, 283-294.
- Klassen, R. G and De Graaff, E. (2001). Facing innovation; Preparing lectures for English-medium instruction in a non-native context. *European Journal of Engineering Education*, **26**, 281-289.
- Lee, J. Y. (2008). *The study of anxiety and risk-taking of university students of english-medium classes in Korea*, Master Thesis, Sookmyung University, Seoul.
- Lee, J. W. (2010). *Study on the university students' perceptions on the English mediated courses*, Master Thesis, Hankuk University of Foreign Studies, Seoul.
- Lee, Y. G. and Hur, J. (2009). A study on the comparison of descriptive variables reduction methods in decision tree induction: A case of prediction models of pension insurance in life insurance company. *Journal of the Korean Data & Information Science Society*, **20**, 179-190.
- You, Y. J. and Park, H. C. (2009). Design and implementation of data mining tool using PHP and WEKA. *Journal of the Korean Data & Information Science Society*, **20**, 425-433.

Study on the effectiveness of english-medium class[†]

Jang Sik Cho¹

¹Department of Informational Statistics, Kyung Sung University

Received 18 October 2012, revised 8 November 2012, accepted 13 November 2012

Abstract

Many universities stress gradually the importance of english-medium class in order to improve the international competitiveness and the internationalization of the university. In this paper, we compare english-medium class with korean class using course evaluation score. Also we analyze the factors that affect the effectiveness of the course evaluation score of english-medium class. First, logistic regression analysis is used to examine the main effects of subjects and individual characteristics. Also, decision tree analysis is used to examine the interaction effects for subjects and individual characteristics. The results of this paper are as follows. Grade, department category, class size, GPA and screening method affect the effectiveness of english-medium class. The highest effectiveness group of english-medium class is that grade is freshmen and department category is humanity. Also the group of the second highest effectiveness group is that grade is freshmen and department category is nature and art and GPA is high.

Keywords: Data mining, decision tree analysis, english teaching, interaction effect, logistic regression analysis, main effect.

[†] This research was supported by Kyung Sung University Research Grants in 2012.

¹ Professor, Department of Informational Statistics, Kyung Sung University, Busan 608-736, Korea.
E-mail: jscho@ks.ac.kr