

## 논문 15

## 교육생산함수의 추정 - 학업성취도에 대한 회귀분석 -

김미란\*

## 1. 서론

본 연구는 수능성적을 종속변수로 하여 교육생산함수(education production function)를 추정함으로써 고등학생의 학업성취도에 영향을 주는 여러 가지 요소들 즉, 학교와 교사특성, 가족배경, 학생개인의 학습노력, 대도시 거주에 대한 영향에 대하여 살펴보고자 한다.

교육정책의 상당 부분은 학교의 교육여건을 개선함으로써 그 성과를 높일 수 있다고 본다. 교육에 대한 투자가 비교적 높은 수준인 미국의 경우 1980년대에 교육생산함수를 통해 실증분석 해 본 결과, 학업성취도에 중요한 영향을 미칠 것으로 기대했던 학교관련 투입요소들이 통계적으로 유의하지 않거나 거의 영향을 주지 못하는 것으로 나타나 교육계와 정책당국이 충격을 받았다. 즉 학교나 교사특성변수보다는 가족배경변수들이 통계적으로 더 유의하게 영향을 주는 것으로 나타난 것이다<sup>1)</sup>. 그러나 이러한 분석결과는 심각한 반론을 불러일으켰는데, 학교나 교사특성변수가 수량으로 측정(measure)하기 어렵고, 학생의 성적에 영향을 미치는 변수가 매우 다양하고 또 그 과정이 복잡하여 계량적 모형으로 구성하기 곤란하다는 점이다. 이런 논쟁은 교육경제학 내에서 아직도 진행 중이지만 우리나라에서는 평준화 효과, 학교효과에 대한 주제로 연구되고 있다.

교육생산함수는 교육투자에 대한 성과를 측정하는 전통적인 접근방법의 하나로서 교육학에서는 학업성취도 분석모형으로도 불린다. 우리나라에서는 아직까지 학업성취도에 대한 통계자료의 접근 자체가 거의 불가능한 상황이다. 그럼에도 불구하고 국제비교 자료인 TIMSS자료를 이용한 것(김진영: 2004, Kang Changhui:2005)과 지역 또는 부분적인 표본 자료를 이용하여 학업성취도를 분석하고 있다. 주요한 주제로는 학업성취도 차이에 대한 학교의 효과(양정호·김경근, 2003, 성기선, 1997), 고교 평준화와 학업성취도의 관계(양정호 2004, 윤종혁 2003, 김태종 외 2004, 성기선 2002, 김태일 1998), 그리고 학업성취도 자체

\* 한국직업능력개발원 전문연구원

1) Hanushek(1995)

에 영향을 주는 요인에 대한 분석(오영수·윤정식, 2003) 등이다. 양정호(2004)에 따르면 기존의 연구들은 전국단위의 패널자료가 없다는 제약, 학교선택효과를 적절히 통제하지 못하는 방법론상의 문제 등으로 인해 그다지 실증적으로 의미있거나 일관성있는 결론을 내지 못하고 있다고 한다. 특히 평준화에 대한 찬반을 위한 이론적 논의보다는 실증적 분석의 중요성을 강조하고 있는데, 구체적으로는 동일한 학생에 대한 3년 이상의 패널자료에 대하여 위계적 선형모형방법(HLM)을 적용하여 분석할 것을 제안하고 있다.

미국의 경우 학업성취도 분석 중 교육생산함수에 대한 연구논문이 1980년대 중반까지 200여편에 이를 정도로 많다. 그러나 많은 연구에서 교사 및 학교특성 변수의 대부분이 통계적으로 유의하지 않았다. 예를 들면 대표적인 투입요소(input factor)인 ‘교사-학생 비율’의 경우, 교육생산함수를 추정하는 152개의 연구논문에 사용되었으나 125개에서는 통계적으로 유의하지 않았고 27개에서만 통계적으로 유의하게 나타났다. 이에 비해 ‘가족적 배경과 관련된 변수’는 많은 연구에서 공통적으로 학생의 학업성취도에 영향을 주는 것으로 나타났다<sup>2)</sup>.

실증적인 분석방법 측면에서 볼 때 교육생산함수는 ‘교육에 대한 성과는 투자에 대비하여 평가해야 한다’는 전형적인 투입-산출분석(input-output analysis)의 시각에 기초하고 있다. 실증적 추정방법과 관련한 논쟁은 한마디로 학교나 교사특성변수가 학업성취도에 통계적으로 유의하지 않은 것으로 나타나는 이유가 ‘이 변수들이 수량적으로 측정하기 곤란하기’ 때문인지 아니면 실제로 학교나 교사특성변수들이 학업성취도(차이)에 ‘그다지 큰 영향을 주지 않기’ 때문인지 알 수 없다는 것이다. 예를 들면 “학급규모변수와 학업성취도는 밀접한 관계가 있음을 실증적으로 규명하고, 이들이 상관관계가 없다는 기존연구(Hanushek)의 결론을 반박하고 있다”<sup>3)</sup>. 기존의 교육생산함수에 대한 많은 실증연구들의 추정모형과 그 결과를 재분석하는 과정에서 매우 다양한 설명변수들이 계량경제학적으로 위 세 가지 문제들에 해당되게 되어 개별적 요소들의 영향이 통계적으로 왜곡되고 있다고 지적하고 있다. 소위 계량경제학에서 말하는 세 가지 문제, 즉 (1) 측정문제, (2) 투입요소변수의 이질성 문제(heterogeneity problem), (3) 적절한 변수의 선택과 구체적이 모형설정에 있어 어려움(specification problem)이 모두 다 관련되어 추정이 매우 곤란한 경우에 해당한다.

여기에서는 방법론에 대한 논의보다는 교육생산함수 추정에 사용되는 다양한 설명변수들에 있어 두번째와 세번째 문제를 고려하여 변수를 선택, 분석하고자 한다. 아울러 다음 두 가지에 초점을 둔다. 첫째 학생의 학업성취도에 영향을 주는 여러 가지 요소들을 모형화한 다음, 각 요소들의 영향을 통제된 상태에서 학업성취도에 중요한 영향을 주는 요소를 식별해 보고자 한다. 즉 ‘학교나 교사특성’과 ‘학생의 가족배경과 학습노력’, 그리고 ‘대도시 거주여부’와 같은 요소들 중에서 어느 것이 가장 중요하게 작용하는지에 대해 분석한다. 두 번째는 성적의 분포형태가 정규분포가 아니고 또 회귀분석으로 추정한 교육생산함수의 ‘오차항이 정규성’가정을 충족하지 못하는 문제에 대하여 분위회귀식(quantile regression)을

2) 반상진(2003)

3) 반상진 (2003)

이용하여 추정하고자 한다.

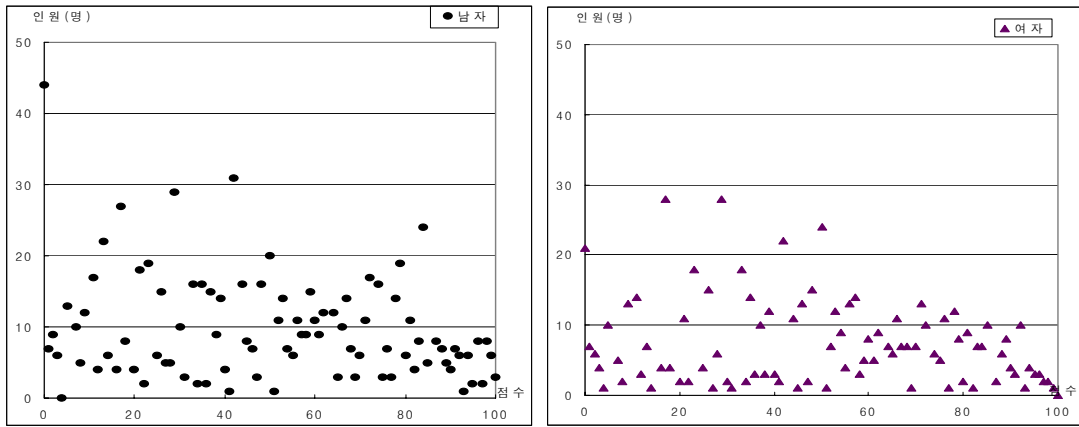
## 2. 자료와 추정식

### 2.1 자료

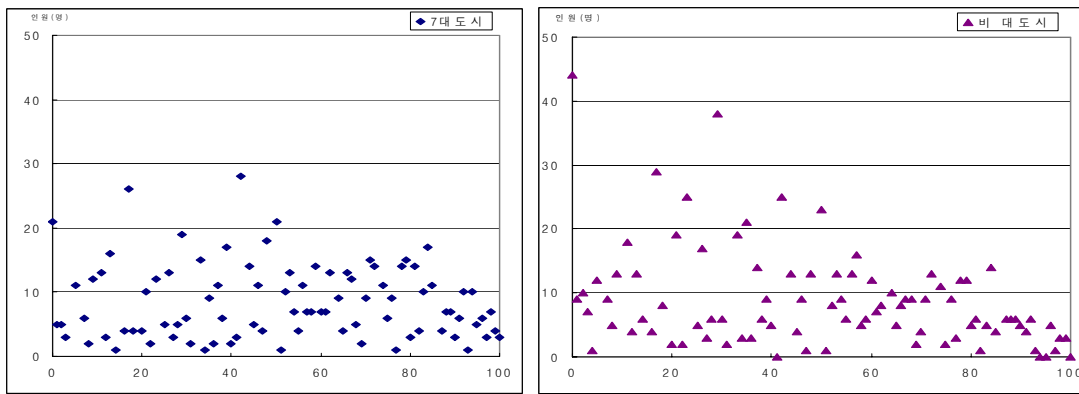
본 연구에서는 한국직업능력개발원의 '한국교육고용패널(KEEP)' 자료의 일반계 고등학교 3학년생의 수학능력시험의 수학점수(표준화)를 교육생산함수 추정의 종속변수로 사용한다. 일반고교생 표본(1,921개)에서 수학을 선택한 1,543명이 도출되며, 이중 설명변수의 미관찰이 있는 106을 제외한 1437명이 분석대상이 된다. 교육생산함수의 산출 지표는 자아개념, 건강한 습관에서부터 언어능력, 시험점수, 학교 탈락률과 졸업률, 개인의 소득 증대, 기술과 능력의 향상, 자선행위 증대에 이르기 까지 매우 다양하다. 특히 학업성취도에 대한 지표(proxy)로는 재학 중 학업성적이나 수능성적, 대학입학을 위한 자격시험을 사용한다. 성기선·강태중(2001)은 경기도지역 재학생의 학업성취도 변화에 대하여 재학 중 성적 변화(고등학교 1학년부터 3학년까지)를 조사하여 사용하였으며, 학교별 또는 학급단위의 성적 평균값과 그 변화분(김영일, 2001, 김태중 외 2004)을 사용하기도 하지만 가장 이상적인 성적 지표는 동일한 개인에 있어 성적의 변화분이 될 것이다.

수학점수의 분포(distribution) 형태를 그림으로 보면 아래와 같다. 세로(Y)축의 값은 각 점수별 인원수(빈도)를 나타내는데, 남녀 모두에서 중간보다 왼쪽에 몰려 있다. 왼쪽으로 왜곡돼(skewed) 있다. 대도시 여부별로 보아도 마찬가지이다(<그림 1>, <그림 2>). 성적의 분포를 10개의 백분위별로 구분하여 누적분포를 보면 <그림 3>과 같다. 남학생과 여학생 간에는 분산이 차이가 별로 없지만, 지역별로는 다르다. 7대도시 지역에 비해 비 대도시 지역에서 성적의 분산이 큰 것으로 나타난다. 이러한 분산의 차이는 대도시지역을 중심으로 한 고교 평준화정책의 결과로 보인다. 성적의 평균값에서도 7대도시 지역은 50.42, 비도시지역은 42.04로 차이가 있다. 요컨대 대도시 여부에 따라 성적의 평균과 분산이 통계적으로 유의하게 다른 것으로 나타나고 있다.

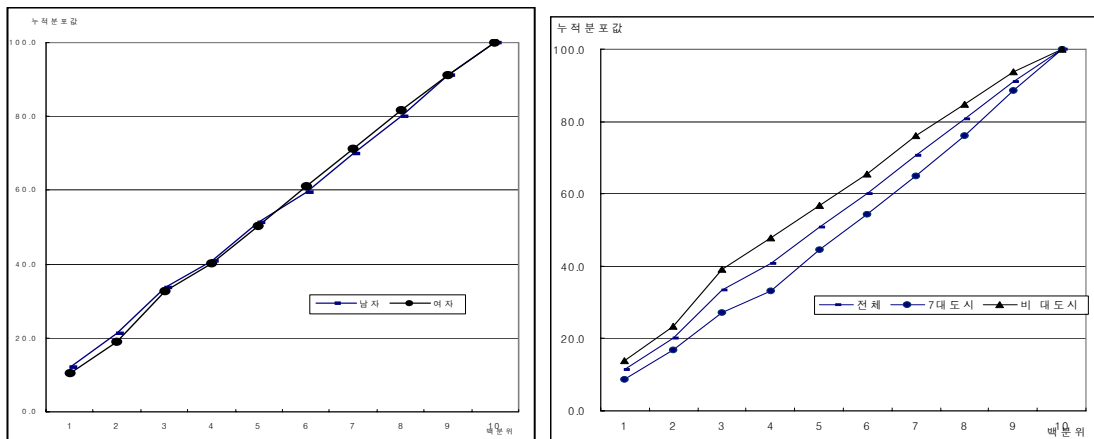
두번째로 수학점수의 분포가 정규분포는 아니더라도 남녀별로 또는 대도시거주 여부에 따라 서로 다를 수 있는데, 이에 대하여 Kolmogorov-Smirnov 검정을 실시한 결과, 수학점수의 분포는 남녀간에는 통계적으로 유의한 차이가 없으나, 대도시거주 여부별로는 동일한 분포로 볼 수 없는 것으로 나타난다(<표 1>).



[그림 1] 수학점수의 분포: 성별



[그림 2] 수학점수의 분포: 대도시 여부별



[그림 3] 수학점수의 분산도: 성별, 대도시 여부별

&lt;표 1&gt; 두 집단 간 수학점수분포의 동일성에 대한 검정

Kolmogorov-Smirnov 검정		
Ho: 남자의 수학성적 분포 = 여자의 수학성적 분포		
집단	D	P-value
여자	0.0339	0.419
남자	-0.0339	0.419
Combined K-S	0.0339	0.776
검정결과: p값이 통계적으로 유의하지 않아 Ho가설을 기각하지 못함. 즉 두 집단의 분포가 같음을 기각하지 못함		
Kolmogorov-Smirnov 검정		
Ho: 7대도시 거주자의 수학성적 분포 = 비 대도시지역 거주자의 수학성적분포		
집단	D	P-value
비 대도시	0.1460	0.000
7대도시	0.0000	1.000
Combined K-S	0.1460	0.000
검정결과: 1%이하의 유의수준에서 Ho가설 기각됨. 즉 두 집단의 분포가 같다는 가설이 기각됨.		

분석에 사용할 변수의 요약통계는 <표 2>와 같다. 전체적으로 보아 남녀 차이보다는 대도시 여부별로 차이가 있다. 주요 특징은 첫째, 성별로 남학생의 ‘이과 비율(51%)’이 여학생(30%)보다 높게 나타나, 여전히 여학생들이 이과선택을 선호하지 않고 있음을 알 수 있다. 둘째, 수학과 국어에서는 남녀 간에 점수 차이가 없으나 영어에서는 남녀 간에 차이가 있는 것으로 나타난다. 즉 여학생이 남학생에 비해 상대적으로 언어능력이 높은 것으로 나타난다. 셋째, 7대도시 거주여부별로 국어, 영어, 수학의 점수 모두에서 통계적으로 유의(5% 이하의 유의수준)하게 차이가 있는 것으로 나타난다(<표 3>). 넷째, 대부분의 설명변수들도 피어슨의 카이제곱검정결과 대도시 여부를 기준으로 볼 때, 그 값이 다른 것으로 나타난다.

이상의 분포형태에 대한 분석은 수학점수에 대한 회귀분석에 대하여 다음 두 가지를 시사한다. 첫째, 설명변수를 사용하여 수학점수에 대해 회귀식을 추정한 후의 잔차항(residuals)의 분포가 정규분포가 아닐 가능성이 있다는 것이다. 그래서 회귀분석 후 잔차항에 대한 정규성 검증을 할 필요가 있다. 만약 잔차항이 정규성을 충족하지 못할 경우 그 왜곡을 보완하기 위해 최소자승추정법(OLS)보다는 분위회귀추정법(quantile regression)을 사용하는 것이 더 적절할 수 있다는 것이다. 그래서 본 연구에서는 먼저 OLS를 사용하여 추정한 다음, 분위회귀추정을 수행한다. 둘째는 대도시 거주여부별로 주요 변수들의 평균치와 분포형태가 다르므로 7대도시 여부별로 구분하여 분석하는 것이 적절함을 시사한다.

&lt;표 2&gt; 기초통계: 성별

변 수	전체			남자		여자		Pearson $\chi^2$ (Pr)
	Obs	Mean	s.d	Mean	s.d	Mean	s.d	
수학점수	1,543	46.14	27.54	46.14	28.21	46.15	26.64	92.06 (0.39)
국어점수	1,543	47.30	27.14	45.62	27.63	49.54	26.33	67.04 (0.37)
영어점수	1,543	47.03	27.55	45.58	28.54	48.95	26.07	95.53 (0.03)
점수합계	1,543	140.47	71.79	137.34	74.06	144.63	68.49	
이과더미	1,543	0.42	0.49	0.51	0.50	0.30	0.46	
공립더미	1,543	0.42	0.49	0.40	0.49	0.45	0.50	
학교연혁	1,543	33.88	22.38	31.73	20.54	36.75	24.35	
학생-교사비율	1,543	15.41	2.45	15.55	2.32	15.23	2.61	
교사 수	1,543	68.32	23.07	67.54	23.34	69.35	22.67	
학생 수	1,543	1086.97	439.84	1079.07	435.32	1097.49	445.89	
수준별 이동수업더미	1,543	0.58	0.49	0.58	0.49	0.57	0.50	
30-45세 교사비율	1,543	0.51	0.12	0.51	0.11	0.52	0.12	
담임 남성더미	1,543	0.79	0.41	0.86	0.35	0.70	0.46	
교사의 경력년수	1,542	16.08	7.05	16.07	7.29	16.08	6.73	
담임교사의 학력년수	1,543	16.98	1.13	17.17	1.17	16.72	1.03	
담임의 석사이상 학위더미	1,543	0.33	0.47	0.40	0.49	0.23	0.42	
담임의 전교조 소속여부	1,543	0.23	0.42	0.21	0.41	0.25	0.43	
비부모 보호자더미	1,543	0.06	0.24	0.06	0.24	0.07	0.25	
아버지 교육년수	1,463	12.85	2.93	12.99	3.02	12.66	2.80	
어머니 교육년수	1,492	11.72	2.47	11.76	2.47	11.67	2.47	
부모 중 하나 전문직 더미	1,543	0.31	0.46	0.31	0.46	0.30	0.46	
아버지 소득	1,379	281.04	167.44	287.26	169.41	272.82	164.58	
어머니 소득	690	144.72	127.27	139.55	112.74	151.36	143.71	
금융 및 재산액	1,518	5.89	2.33	5.96	2.31	5.79	2.36	
같은반학생 모(母)의 평균학력	1,541	5.00	0.58	5.02	0.61	4.97	0.54	
수학 과외시간	1,543	2.62	3.20	2.76	3.46	2.45	2.81	
국어 과외시간	1,543	1.37	2.57	1.38	2.75	1.35	2.32	
영어 과외시간	1,543	1.80	2.58	1.90	2.74	1.66	2.33	
과외시간 합계	1,543	5.80	6.37	6.05	6.83	5.46	5.69	
혼자 공부시간	1,543	12.33	9.98	12.50	10.17	12.11	9.73	
자신감 평균	1,543	3.49	0.54	3.51	0.55	3.46	0.53	
서울더미	1,543	0.22	0.42	0.20	0.40	0.26	0.44	
7대 도시더미	1,543	0.49	0.50	0.51	0.50	0.47	0.50	
표본 수		1,543		881		662		

자료: KEEP

<표 3> 기초통계: 성적상위(20%)/하위(20%)집단별, 7대도시 여부별

변 수	상위 20%		하위 20%		7대도시		비 대도시	
	Mean	s.d	Mean	s.d	Mean	s.d	Mean	s.d
수학점수	84.58	7.25	8.15	6.27	50.42	27.49	42.04	26.98
국어점수	67.36	23.63	24.44	22.16	50.69	26.93	44.05	26.95
영어점수	70.60	21.60	21.84	20.88	51.31	27.29	42.92	27.19
점수합계	222.54	44.06	54.43	43.37	152.42	70.21	129.02	71.47
이과더미	0.51	0.50	0.35	0.48	0.45	0.50	0.39	0.49
공립더미	0.45	0.50	0.39	0.49	0.43	0.50	0.41	0.49
학교연혁	35.13	24.40	30.06	17.81	37.27	25.93	30.64	17.77
학생-교사비율	15.60	2.58	14.86	2.89	16.07	2.21	14.79	2.51
교사수	72.74	20.26	60.74	23.99	77.84	16.67	59.19	24.61
학생수	1155.70	407.66	945.84	455.07	1265.03	350.01	916.37	449.58
수준별 이동수업더미	0.62	0.49	0.57	0.50	0.68	0.47	0.48	0.50
30-45세 교사비율	0.52	0.11	0.51	0.12	0.48	0.11	0.55	0.11
담임 남성더미	0.84	0.37	0.80	0.40	0.83	0.38	0.76	0.43
교사의 경력년수	16.03	7.36	16.01	7.00	17.60	7.20	14.62	6.58
담임교사의 학력년수	17.14	1.41	17.05	1.14	17.02	1.22	16.94	1.04
담임의 석사이상 학위더미	0.37	0.48	0.34	0.48	0.31	0.46	0.34	0.47
담임의 전교조 소속여부	0.26	0.44	0.21	0.40	0.18	0.38	0.28	0.45
비부모 보호자더미	0.04	0.20	0.08	0.28	0.06	0.23	0.07	0.25
아버지 교육년수	13.73	3.06	12.17	2.87	13.34	2.88	12.37	2.90
어머니 교육년수	12.42	2.66	11.19	2.46	12.12	2.46	11.33	2.42
부모 중 하나 전문직 더미	0.42	0.49	0.22	0.42	0.37	0.48	0.24	0.43
아버지 소득	317.05	164.06	254.29	137.45	311.68	180.90	251.50	147.57
어머니 소득	162.17	135.03	143.06	121.23	162.59	144.22	128.71	107.59
금융 및 재산액	6.53	2.33	5.63	2.11	6.45	2.39	5.37	2.14
같은반학생 모(母)의 평균학력년수	5.21	0.63	4.83	0.51	5.17	0.59	4.83	0.53
수학 과외시간	3.28	3.07	1.75	3.35	3.11	3.18	2.16	3.15
국어 과외시간	1.78	2.50	0.90	2.03	1.82	2.82	0.94	2.22
영어 과외시간	1.94	2.38	1.74	3.00	2.13	2.59	1.49	2.53
과외시간 합계	7.00	5.94	4.39	6.41	7.06	6.39	4.58	6.12
혼자 공부시간	15.40	10.49	9.17	8.77	13.37	10.26	11.34	9.60
자신감 평균	3.60	0.56	3.48	0.54	3.53	0.53	3.45	0.54
서울더미	0.21	0.41	0.18	0.39	0.46	0.50	0.00	0.00
7대 도시더미	0.60	0.49	0.41	0.49	1	0.00	0	0
표본 수	321		312		755		788	

자료: KEEP

## 2. 2 추정식

교육생산함수 추정을 위한 회귀식은 통상적으로 Levin(1995)에 따라 아래와 같이 선형식(linear equation)으로 구성한다<sup>4)</sup>. 설명변수는 학교특성, 교사(담임)특성, 가족배경, 학급동

료, 학생의 학습노력과 계열변수, 지역으로 구성한다. 백일우(2000)는 교사의 봉급, 학생 1인당 교육비지출액, 개선된 행정과 시설 등과 같은 설명변수로 포함하고 있으며, 다른 연구에서는 목적에 따라 ‘교육과정의 직절 측면’ 변수를 설명변수로 포함하여 분석하기도 한다.

교육생산함수는 경제학의 다른 함수에 비해 매우 다양한 설명변수를 설정하고 있고 또 설명변수들이 종속변수에 영향을 주는 경로가 복잡하기 때문에 모형설정에 매우 주의가 필요하다고 한다. 앞에서 말한 모형설정의 어려움이다. 그래서 회귀추정식은 다음 몇 가지에 주의하여 아래 식 (1)과 같이 구성 한다<sup>5)</sup>.

첫째, 가족배경에 대한 변수들 중에서 ‘부모의 학력년수나 소득수준’은 ‘사교육비 지출’과 매우 밀접한 상관관계에 있다(표 4 참조). 따라서 다중공선성 문제를 방지하기 위해 두 변수 중 하나만 사용하여 각각 추정한다(추정식 1에서 A와 B로 구분).

<표 4> 사교육비 지출액과 가족배경 변수간의 상관관계

	사교육비	부모		부모중		부(父)소 득	자산액	같은반 학생 모(母)학력평균
		부(父) 교육년수	모(母) 교육년수	하나 전문직 더미	모(母) 전문직더미			
사교육비	1							
부(父) 교육년수	0.287	1						
모(母) 교육년수	0.291	0.672	1					
부모중하나 전문직	0.229	0.434	0.386	1				
모(母) 전문직	0.154	0.268	0.332	0.234	1			
부(父) 소득	0.408	0.350	0.368	0.356	0.109	1		
자산액	0.430	0.338	0.353	0.324	0.182	0.544	1	
같은반 학생 모(母)의 학력평균	0.272	0.437	0.543	0.289	0.230	0.301	0.299	1

자료: KEEP

둘째, 또래효과(peer effects)는 학생의 학업과정과 성적에 중요한 영향을 미치는 요소인데, 이를 반영하기 위한 지표로 ‘같은 반 학생 어머니의 평균학력년수’를 사용한다<sup>6)</sup>(추정식 2). KEEP 자료는 ‘친한 친구의 성적이나 특성’을 보여주는 정보를 갖고 있지 않다. 중학생 자료의 경우 이를 보완하면 활용도가 높아질 것으로 본다.

셋째, 학생의 학습노력을 나타내는 변수로는 ‘이과계열 더미’ 외에 ‘수학과외시간’, ‘혼자 공부하는 시간’, ‘학생 자신이 평가한 자기인식(자신감 평균)’을 사용한다(추정식 3).

넷째, 지역 차이를 고려하기 위해 7대도시 여부를 포함한다. 서울지역더미를 사용해 보았으나 모든 식에서 통계적으로 유의하지 않아서 사용하지 않는다.

4) Hanushek(1986), 반상진(2003)

5) 설명변수의 구성과 관련한 백성준박사님(직능원)의 코멘트에 감사드립니다.

6) 여기서 같은 반 학생의 어머니라 함은 같은 반 학생 전체가 아니라 표본으로 뽑힌 학생의 어머니만 임.



회귀추정식:

$$\ln A_t = \alpha + \beta_1 D_{t-1} + \beta_2 SR_{t-1} + \beta_3 T_{t-1} + \beta_4 F_{t-1} + \gamma P_{t-1} + \delta Z_{t-1} + \zeta R_{t-1} + e_t \quad (1)$$

여기서  $A_t$ : t기에 학생의 수능성적 (t=2005말 시점)

$D_{t-1}$ : 학생의 인적속성; 성(남학생=1)

$SR_{t-1}$ : t-1기의 학교관련 특성; 공립더미, 학교연혁, 학생-교사비율, 교사 수, 학급규모, 수준별 이동학습실시 여부, 30-45세인 교사의 비율 등

$T_{t-1}$ : t-1기의 교사관련 특성; 담임의 성(남성=1), 교사경력년수, 석사이상 학위소지더미, 전교조소속여부

$F_{t-1}$ : t-1기의 학생의 가족적 배경; 부모가 아닌 보호자와 동거여부, 부의 학력, 부모 중 어느 한쪽의 전문직 종사여부, 가계의 금융 및 부동산 자산액, 지난 6개월간 월평균사교육비 지출액

$P_{t-1}$ : t-1기에 학생의 친구 관련 요소; 같은 반 학생 어머니의 평균학력년수

$Z_{t-1}$ : 학생의 학습노력: 이과여부, 수학과외시간, 혼자 공부하는 시간, 스스로 평가한 자신감 점수

$R_{t-1}$ : 7대도시 거주여부

$e_t$ : 잔차항

### 3. 실증결과와 논의

교육생산함수의 회귀추정 결과가 <표 5>와 같다. 추정식 1에 비해 ‘또래효과’와 ‘학생의 학업노력’ 그리고 지역특성을 설명변수로 포함할 때 조정된 R<sup>2</sup>도 높고 통계적으로 유의미한 변수가 많아 추정식 3이 가장 적합한 모형으로 나타난다.

추정결과를 전체적으로 보면 ‘학생의 학습노력’, ‘또래효과’, ‘대도시 더미’ 변수가 통계적으로 유의하게 학생의 학업성취도에 영향을 주는 것으로 나타나며 ‘학교특성’ 변수는 일부만 통계적으로 유의하다. 서론에서 제기했던 문제, 즉 우리나라 교육생산함수에서 교사특성변수가 대부분 통계적으로 유의하지 않았다. 그 이유는 역시 앞에서 지적한 투입요소의 측정 문제 때문으로 보인다. 교육과정에서 교사의 노력과 헌신도는 수량화 할 수 없을 뿐 아니라 수량화 하더라도 그 기능과 역할이 성과에 영향을 주는 경로 등이 복잡하다는 것이다<sup>7)</sup>. 다음으로 ‘가족배경’ 변수는 추정식 1과 2에서 통계적으로 유의하였으나 ‘학생의 학습노력

7) 교육학에서는 이 문제에 대하여 사례연구를 통해 ‘카톨릭계’ 학교나 ‘행정이 효율적이고 투명한 학교’일수록 학생의 학업성적이 우수함을 밝히는 연구도 있다(반상진 2003). 경제학에서는 패널자료를 이용하여 교사와 학교 특성, 학교 운영시스템 등 비교적 고정적 요소의 영향을 통제(고정효과 모형)한 상태에서 분석하기도 한다.

‘을 설명변수로 포함함에 따라 통계적으로 유의하지 않게 바뀌었다.

<표 5> 회귀식 추정결과

변 수	식 1		식 2		식 3			
	A		B					
	계수	(S.E)	계수	(S.E)	계수	(S.E)		
상수	-22.345	(10.03) *	-5.941	(11.10)	-41.291	(12.59) ***	-44.207	(13.16) ***
남성더미	-0.214	(1.50)	0.223	(1.45)				
공립더미	0.867	(1.67)	0.951	(1.62)	1.448	(1.66)	2.840	(1.62) *
학교연혁	0.103	(0.04) ***	0.096	(0.03) ***	0.093	(0.04) ***	0.074	(0.03) **
학생-교사비율	1.624	(0.74) **	1.399	(0.71) **	1.264	(0.74) *	1.210	(0.54) *
교사수	0.803	(0.21) ***	0.729	(0.20) ***	0.602	(0.21) ***	0.409	(0.20) **
학생수	-0.037	(0.01) ***	-0.031	(0.01) **	-0.028	(0.01) **	-0.021	(0.01)
수준별 이동수업 더미	-2.913	(1.56) *	-2.527	(1.52) *	-2.083	(1.56)	-2.391	(1.55)
30-45세 교사비율	11.765	(7.03) *	9.807	(6.78)	13.152	(6.95) *	17.417	(6.91) **
담임 남성더미	4.888	(1.91) ***	4.799	(1.85) ***	4.254	(1.86) **	2.940	(1.85)
교사의 경력년수	-0.031	(0.11)	0.041	(0.11)	-0.054	(0.11)	-0.098	(0.11)
담임의 석사이상	-0.457	(1.53)	0.406	(1.50)	-0.597	(1.50)	-1.214	(1.49)
학위더미								
담임의 전교조	1.674	(1.81)	2.574	(1.75)	1.642	(1.79)	1.706	(1.73)
소속여부								
비부모 보호자더미	-2.243	(4.54)	-0.695	(2.83)	-2.053	(4.57)	-2.250	(4.41)
아버지 교육년수	0.934	(0.28) ***			0.561	(0.29) **	0.401	(0.28)
사교육비			0.075	(0.02) ***				
부모중 하나 전문직	3.008	(1.73) *			2.294	(1.73)	1.727	(1.70)
더미								
금융 및 재산액	0.801	(0.35) **			0.623	(0.35) *	0.339	(0.35)
같은반 학생 모(母)의					6.764	(1.45) ***	5.975	(1.43) ***
평균학력년수								
이과더미							2.919	(1.47) **
수과외시간							0.596	(0.23) ***
혼자공부시간							0.484	(0.07) ***
자신감 평균							1.956	(1.31)
7대도시 더미							3.744	(1.69) **
표본수	1,439		1,540		1,437		1,437	
조정된 R <sup>2</sup>	0.081		0.061		0.095		0.139	

자료: KEEP. 주: \*\*\* 유의수준 1%, \*\* 유의수준 5%, \* 유의수준 10%

변수 항목별로 살펴보면 첫째, 학교특성에서는 '30-45세 교사비율', '교사수'와 '학생-교사비율', '학교연혁'이 성적에 양(+)의 영향을 미치는 것으로 나타난다. 특히 '30-45세 교사비율'은 학교별 차이를 보여주는 지표로서 경험있고 핵심연령층에 있는 교사진의 구성이 학업성취에도 영향을 미칠 수 있음을 보여주는 것으로 해석된다. 그러나 학교규모(학생수) 변수와 '학생-교사비율'의 경우 상식과 다르거나 통계적으로 유의하지 않은것으로 나타나

앞으로 이에 대한 정교한 분석이 요청된다.

둘째, 가족배경으로는 '아버지 학력년수'와 '사교육비 지출액', 그리고 '금융 및 자산액'등이 자녀의 성적에 영향을 미칠 수 있으나 '학생 자신의 학습노력'과 '또래효과' 등을 고려한다면 실제로 별 영향을 주지 못하는 것으로 나타난다. 이러한 결과는 부모의 경제적 또는 학력 배경이 '자녀의 과외시간이나 대도시 거주에 따른 좋은 학습환경의 제공'으로 전환되어 성적에 영향을 미치는 것으로 볼 수 있다. 셋째, 학생의 계열이 '이과계'일수록 그리고 '혼자공부하는 시간'이 많을수록, 그리고 '수학과외시간'이 많을수록 수학점수가 높은 것으로 나타난다. 이중 '수학과외'와 '혼자공부'는 학생개인의 노력뿐 아니라 부모 및 가족환경과도 상관관계가 있기 때문에 독립적인 설명변수로 보기 어렵다. 넷째, 대도시에 거주할수록 수학성적에 통계적으로 유의하게 양(+)의 영향을 미치는 것으로 나타난다. 다섯째, 또래효과 변수인 '같은 반 학생 어머니의 학력의 평균치' 높을 수록 학생의 성적이 높은 것으로 나타난다.

회귀식을 성적상위 20%와 하위 20%집단, 그리고 7대도시 거주와 비(非) 도시지역 거주 집단으로 구분하여 추정하여 보았다(표 6). 이 두 집단간에 학업성적이 형성되는 과정이 다를 것으로 볼 수 있기 때문이다. 그 결과 흥미있는 몇 가지 결과가 보인다. 첫째, '수학과외 시간'이 대도시에서는 통계적으로 유의하게 성적을 높이는 것으로 나타나는데 반해 비 도시집단에서는 통계적으로 유의하지 않은 것으로 나타난다. 이는 똑 같은 시간 수학과외를 하더라도 대도시지역에서 하는 것이 성적에 유효하다는 것이다. 이것은 대도시 지역에서 이과계열의 계수가 큰 값으로 수학성적과 양의 관계를 갖는 것과도 일관성이 있다. 대도시 지역에 위치한 과외학원이 비 대도시 지역에 비해 학생의 수학성적을 높이는데 유효하게 영향을 미치고 있는 것으로 해석할 수 있다.

둘째는 대도시 지역변수의 경우 성적 상위집단과 하위집단 모두에서 통계적으로 유의하게 영향을 미치는 것으로 나타나며, 상위집단의 계수값이 하위집단의 값보다 더 크다. 요약하면 수학성적 상위집단일수록 7대도시에 거주하면서 수학과외를 받는 것이 통계적으로 유의하게 수학성적을 높이는데 영향을 주고 있다고 볼 수 있다.

셋째, '30-45세 교사비율'이 비 대도시지역에서 매우 큰 양의 계수로 나타난다. 지방도시 및 농촌지역의 경우 대도시지역에 비해 학교여건자체가 큰 차이가 없는 상황에서 젊고 현장경험이 충분한 핵심연령대의 교사비율이 높을 수록 학업성적에 긍정적인 영향을 주고 있는 것으로 해석할 수 있다. 이 결과는 젊고 경험있는 핵심연령층 교사의 이직을 방지하고 근속을 유도하기 위한 방안이 필요함을 시사한다. 예를 들면 대도시와 비 도시지역 간 교사의 보수에 차등을 둔다거나 비 도시지역 교사의 근속을 유인할 체계를 개발할 필요가 있겠다.

넷째, 전체집단에 대해서와 마찬가지로 '혼자 공부하는 시간'과 '같은 반 학생 어머니의 학력평균'은 지역과 무관하게 중요한 요소로 영향을 미치고 있다.

&lt;표 6&gt; 회귀식 추정결과: 성적 집단별, 대도시 여부별; 추정식 3 사용

변 수	성적 상위 20% 집단		성적 하위 20%집단		비 대도시		7대도시	
	계수	(S.E)	계수	(S.E)	계수	(S.E)	계수	(S.E)
상수	63.806	(12.28) ***	12.53	(6.25) **	-25.973	(16.31)	-48.750	(44.80)
공립더미	0.957	(0.98)	2.07	(0.92) **	2.695	(2.15)	3.276	(2.88)
학교연혁	-0.015	(0.02)	0.04	(0.02)	0.087	(0.06)	0.063	(0.05)
학생-교사비율	0.447	(0.71)	-0.02	(0.34)	0.226	(0.83)	2.536	(2.61)
교사수	0.183	(0.17)	-0.25	(0.11) **	0.301	(0.32)	0.472	(0.65)
학생수	-0.010	(0.01)	0.01	(0.01) **	-0.008	(0.02)	-0.035	(0.04)
수준별 이동수업더미	-0.963	(0.95)	-0.89	(0.88)	-5.096	(2.23) **	0.170	(2.41)
30-45세 교사비율	10.604	(4.93) **	2.54	(3.83)	20.539	(9.57) **	15.992	(12.24)
남성더미	1.479	(1.17)	-0.43	(1.13)	3.405	(2.56)	1.161	(3.01) *
교사의 경력년수	-0.040	(0.07)	0.00	(0.06)	-0.199	(0.16)	-0.090	(0.17)
담임의 석사이상 학위더미	-1.711	(0.88) *	-1.39	(0.79) *	0.680	(2.01)	-2.818	(2.54)
담임의전교조 소속여부	0.638	(1.02)	0.09	(0.97)	5.480	(2.23) **	-3.075	(3.06)
비부모 보호자더미	7.297	(3.27) **	-0.73	(1.94)	-4.267	(5.56)	0.919	(7.33)
아버지 교육년수	0.228	(0.17)	0.14	(0.15)	0.490	(0.38)	0.314	(0.42)
부모중하나 전문직 더미	0.705	(0.96)	-0.68	(1.01)	1.508	(2.39)	1.431	(2.42)
금융 및 재산액	-0.145	(0.21)	0.08	(0.21)	-0.233	(0.50)	0.822	(0.50)
같은반 학생 모(母)의 평균학력년수	2.145	(0.75) ***	-0.40	(0.92)	4.809	(2.18) **	5.244	(2.02) ***
이과더미	-0.615	(0.83)	0.13	(0.82)	0.798	(1.96)	4.066	(2.23) *
수과외시간	-0.121	(0.14)	0.17	(0.11)	0.423	(0.30)	0.730	(0.34) **
혼자공부시간	0.126	(0.04) ***	0.09	(0.04) **	0.657	(0.10) ***	0.352	(0.10) ***
자신감 평균	-2.667	(0.78) ***	-1.66	(0.70) **	0.798	(1.73)	3.140	(1.98)
7대도시 더미	2.465	(1.04) **	1.90	(0.99) *				
표본수	307		285		738		699	
조정된 R <sup>2</sup>	0.123		0.060		0.178		0.063	

자료: KEEP. 주: \*\*\* 유의수준 1%, \*\* 유의수준 5%, \* 유의수준 10%

### 3. 1 분위회귀(quantile regression) 추정결과

본 연구자료의 수학점수는 표준화된 것이지만, 모집단의 분포형태가 어떠한지에 대해서는 어떤 정보도 없다. ‘자료설명’에서 지적한 바와 같이 앞에서 추정한 회귀분석의 가정인 잔차(residual)의 정규성(normality)에 대해 검정할 필요가 있다. 통상적으로 정규성 검정(normality test)<sup>8)</sup>은 모집단의 분포를 알지 못하는 경우, 비모수적 방법인 Shapiro-Wilk의 정규성 검정을 적용하여 가능하다. 검정 결과 ‘잔차의 분포가 정규분포’라는 가설이 1%이하의 유의수준에서 기각된다(표 7).

8) 류근관(2003) 구자홍 외(1992)

<표 7> OLS회귀식 잔차의 분포의 정규성 검정

Shapiro-Wilk의 정규성 검정. Ho: 잔차항은 정규분포이다				
관찰치	Shapiro-Wilk W	test for normal V	z	p value(Pr < W )
1,437	0.98612	12.183	6.281	<0.0001

검정결과: 1%이하의 유의수준에서 Ho가설 기각됨

위 교육생산함수 회귀식의 잔차항은 정규성 가정을 충족하지 못한다. 이러한 문제를 해결하기 위한 방법으로 가중치( $\alpha$ )를 이용해 잔차의 패턴문제를 수정하는 분위회귀식을 추정하였다<sup>9)</sup>. 결과가 <표 8>과 같다. 백분위(percentile:  $\alpha$ )가 20, 40, 60, 80으로 변화함에 따라 통계적으로 유의한 설명변수가 달라진다. 60백분위에서의 추정결과가 가장 적합도가 높은 모형으로 보인다. 20, 40백분위에 비해 60백분위에서 ‘학교특성변수’와 ‘이과계열’ ‘학생의 학습노력’, 그리고 ‘대도시 더미’변수의 계수가 통계적으로 유의해 진다. 이러한 결과는 회귀분석의 추정식 3과 유사한 결과이다. 그러나 분위회귀추정결과는 교육생산함수의 추정방법과 관련하여 잔차항이 패턴을 보이는 문제를 개선한 것이고 또 OLS추정법에 비해 통계적으로 유의한 변수와 그렇지 않은 변수를 더 쉽게 가려내는 장점이 있다.

<표 8> 분위회귀 추정결과

변 수	Quantile(20)		Quantile(40)		Quantile(60)		Quantile(80)	
	계수	(S.E)	계수	(S.E)	계수	(S.E)	계수	(S.E)
상수	-34.633	(21.19)	-64.006	(18.99) ***	-69.349	(19.46) ***	-30.543	(16.80) *
공립더미	6.138	(2.54) **	3.453	(2.38)	2.836	(2.38)	1.900	(2.10)
학교연혁	0.127	(0.06) **	0.093	(0.05) *	0.077	(0.05)	0.006	(0.04)
학생-교사비율	-0.531	(1.21)	0.668	(1.06)	2.132	(1.09) **	2.208	(0.93) **
교사수	-0.434	(0.34)	0.640	(0.30) **	0.826	(0.31) ***	0.726	(0.27) ***
학생수	0.032	(0.02)	-0.029	(0.02)	-0.048	(0.02) **	-0.045	(0.02) ***
수준별 이동수업더미	-2.647	(2.47)	-3.855	(2.27) *	-1.391	(2.26)	-1.757	(1.96)
30-45세 교사비율	15.871	(11.26)	21.388	(10.44) **	13.751	(10.48)	9.278	(9.05)
담임 남성 더미	2.403	(2.89)	4.260	(2.65)	3.592	(2.68)	2.123	(2.30)
교사의 경력년수	0.154	(0.18)	-0.150	(0.16)	-0.051	(0.16)	-0.416	(0.14) ***
담임의 석사이상학위더미	-0.604	(2.35)	1.241	(2.18)	-1.942	(2.18)	-1.999	(1.86)
담임의전교조 소속여부	3.053	(2.78)	1.569	(2.56)	2.890	(2.54)	0.458	(2.21)
비부모 보호자더미	-5.863	(6.81)	-1.690	(6.40)	-0.057	(6.41)	-6.477	(5.72)
아버지 교육년수	0.682	(0.44)	0.604	(0.41)	0.580	(0.41)	0.247	(0.36)
부모 중 하나 전문직 더미	1.853	(2.68)	1.827	(2.49)	-2.136	(2.49)	3.016	(2.15)
금융 및 재산액	-0.053	(0.51)	0.599	(0.51)	0.874	(0.50) *	0.203	(0.42)
같은반 학생 모(母)의 평균학력년수	4.613	(2.29) **	6.013	(2.11) ***	5.793	(2.11) ***	8.925	(1.81) ***
이과더미	1.340	(2.32)	1.812	(2.11)	3.664	(2.13) *	5.103	(1.87) ***
수과외시간	0.344	(0.41)	0.316	(0.34)	0.764	(0.32) **	0.729	(0.25) ***
혼자공부시간	0.540	(0.11) ***	0.561	(0.10) ***	0.590	(0.10) ***	0.503	(0.09) ***
자신감 평균	-0.124	(2.07)	3.841	(1.89) **	5.656	(1.91) ***	1.068	(1.68)
7대도시 더미	-0.283	(2.73)	2.833	(2.50)	5.003	(2.51) **	5.820	(2.13) ***
표본 수	1,437		1,437		1,437		1,437	
Pseudo R <sup>2</sup>	0.079		0.095		0.107		0.104	

자료: KEEP. 주: \*\*\* 유의수준 1%, \*\* 유의수준 5%, \* 유의수준 10%

9) R. Koenker & K. F. Hallock(2001)

#### 4. 결론과 의미

추정결과 전체적으로 볼 때 학생의 학업성취도는 ‘학생의 학습노력’과 ‘가족배경’에 주로 영향을 받으며, ‘대도시 거주’와 ‘같은 반 학생 어머니의 학력수준’과 같은 변수도 중요한 요소임을 확인할 수 있다. 이외에 학교 특성 중에서 ‘30-45세 교사비율’이 통계적으로 유의하게 영향을 미치는 변수로 확인되었다. 전체적으로 학교나 교사특성변수의 통계적 유의성이 확실하게 나타나지 않은 것이 외국의 연구결과와 비슷하다. 그러나 학업성취도가 다른 변수들을 통제된 상태에서 학생의 학습노력, 과외시간이나 대도시 거주, 그리고 같은 반 학급의 분위기 등에 주로 영향을 받는다는 것을 알 수 있다. 아울러 학교별 차이를 보여주는 지표로서 ‘경험있고 핵심연령층에 있는 교사진’의 존재도 학업성취에도 영향을 미칠 수 있음을 발견하였다.

또한 분석결과 자료와 추정방법상의 한계도 파악되었다. 학교와 교사특성변수의 영향에 대한 좀 더 정확한 분석을 위한 자료와 모형이 필요하다는 것이다. 개인단위의 패널 자료 등을 이용하여 학교와 교사특성, 평준화 여부 등을 적절히 통제하고 또 위계적 선형모형을 활용하여 학교선택과 고정적 요소들의 영향을 제거한 분석이 필요함을 알 수 있다. 최근 외국의 경우 학생의 재학 중 성적에 대한 패널자료를 사용하여 교육생산함수를 추정하고 있다. 패널자료를 이용할 경우, 학생의 가족적 배경이나 학교특성과 같이 개인별로 고정적인 변수가 성적에 미치는 영향을 제거한 상태에서 교사의 급여나 부모의 소득변화, 그 외 학습여건의 변화가 성적에 미치는 영향을 분석할 수 있다. 이렇게 정교한 분석을 위해서는 KEEP의 조사항목에 학교의 재정과 지출, 교사의 보수, 학생의 재학 중 성적과 같은 변수가 새롭게 조사되어야 할 것으로 본다.

그 외 통계적으로 유의한 결과를 보인 변수들을 보면 첫째, ‘혼자 공부시간’이 많을수록 수학 성적 성취도가 높은 것으로 나타난다. 둘째로 ‘수학과외시간’도 수학성적에 통계적으로 유의한 영향을 주는데, 흥미있는 것은 대도시 여부별로 약간 다르다는 점이다. 대도시에 거주하면서 수학과외를 받는 경우, 비 도시지역에서 같은 시간 과외를 받는 것보다 수학 학업성취도에 통계적으로 유의하게 양(+)의 효과를 내는 것으로 나타난다. 다시 말하면 순수히 학생의 학업노력으로 볼 수 있는 ‘혼자 공부시간’은 지역조건이나 부모의 소득, 가족배경의 영향을 통제하더라도 그 효과가 안정적으로 성적을 높이지만, ‘과외’의 경우 대도시 지역에서만 성적에 효과가 있다는 것이다. 이것은 대도시의 특정지역에 과외학원이 집중되어 있는 상황과 무관하지 않다고 본다. 이는 사교육시장에 있어 대도시와 지역도시 간에 차이가 있는 것이며, 그 결과 자녀의 사교육을 위해 과외학원이 잘 발달한 대도시 지역으로의 이동이 앞으로 계속될 가능성이 있음을 간접적으로 보여준다. 셋째, 부모의 경제적 배경과 같은 ‘직접적 지표(proxy)’보다는 교육여건의 다른 요소들과 ‘상호작용(interaction)’하는 변수가 통계적으로 유의하게 나타난다. ‘같은 반 학생 어머니의 학력 평균’이나 ‘과외시간’ ‘대도시거주’와 같은 변수는 부모의 경제적 여건에 명문학원이 있는 지역, 그리고 어머니의 교육열이 결합하여 것이다. 이것은 학생의 성적을 둘러싼 여러 변수들 중 어느 한 요소가 아니라 매우 다양한 변수들의 상호작용에 의해 이뤄진다는 것으로 학업성취와 관련된 정책의 수립이나 기획에 있어서도 이러한 상호작용을 충분히 고려해야 함을 시사한다.

## 참고문헌

- 김진영, 부가 교육생산함수와 학교평가: 국제비교 평가의 활용, 재정논집, 19(1), 2004.
- 김태일, 고교 평준화 정책의 학업 성취효과 분석, 한국정책학회보 제7권 제3호, 1998
- 김태종 외, 고교 평준화 정책이 학업 성취도에 미치는 효과에 관한 실증분석, KDI 국제정책대학원 교육개혁연구소, 2004
- 류근관, 통계학, 법문사, 2003.
- 반상진, '교육의 생산성', 곽영우 외, 『교육경제학』, 한국교육행정학회, 2002년.
- 백일우, 교육경제학, 학지사, 2000.
- 성기선·강태중, 평준화 정책가 지적 수월성 교육의 관계에 대한 실증적 검토, KEDI정책포럼, 2001.
- 성기선, 고등학생들의 학업성취도에 미치는 학교효과 탐색, 교육사회학연구 7(4), 1997
- 양정호, 평준화와 학업성취, 교육사회학연구, 14(2), 2004.
- 양정호·김경근, 학업성취에 대한 학교조직의 효과: TIMSS-R의 위계적 선형모형연구, 교육사회학연구 13(2), 2003
- 오영수·윤정식, 일반계 고등학생의 성적결정 요인 분석, 교육재정경제연구 12(1), 2003
- 윤종혁 외, 고교 평준화 정책의 적합성 연구(I), 한국교육개발원, 2003.
- 한국직업능력개발원, '한국교육고용패널' 2005.
- Alan B. Krueger, Experimental Estimates of Education Production Functions, *Quarterly Journal of Economics*, Vol.114:1:497, May 1999.
- Changhui Kang, Classroom peer effects and academic achievement: quasi-randomization evidence from South Korea, KRIVET Fellow Forum, 2005
- E. A. Hanushek, Education productin functions,in M Carnoy(ed.) *International encyclopedia of economics of education*, Pergamon 1995.
- H. M. Levin, Raising educational productivity, in M Carnoy(ed.) *International encyclopedia of economics of education*, Pergamon 1995.
- R. Koenker & K. F. Hallock, Quantile Regression, *Journal of Economic Perspectives* vol.15:4:143, 2001