International Education Journal, 2005, 6(3), 308-315.
ISSN 1443-1475
http://iej.cjb.net

Analysing education production in Malaysia using canonical correlation analysis

Noor Azina Ismail
University of Malaya *nazina@um.edu.my*

Ang Guat Cheng
University of Malaya

Data from the Third International Mathematics and Science Study carried out in 1999 and canonical correlation analysis were used to investigate the effects of school inputs, environmental inputs and gender influence in the production of a joint educational production function in mathematics and science subjects for eighth grade students in Malaysia. School inputs include per pupil non-teaching expenditure, pupil teacher ratio, teaching experience and instructional hours. On the other hand, environmental inputs consist of home educational resources index and an out-of-school study time index. Gender influence is represented by the percentage of female students in a class. From the study, teaching experience can be dropped from the model because it does not give any additional explanatory power. Marginal products and marginal rates of substitution are calculated and it was found that schools with low-level out-of-school study time can compensate for these deficiencies by having extra instructional hours.

Education production, canonical correlation analysis, marginal rates of substitution, marginal products, school achievement

INTRODUCTION

Education has always been regarded as a vital factor in achieving the general aims of society. In order to achieve this aim, the Malaysian Government (Government of Malaysia, 1996) has placed special priority on improving the educational quality especially in science programs, and efforts are made to increase the number of students in this area of study. Additional emphasis is placed on continuing to improve the quality of and access to schools in underserved areas. This includes expanding the hostel program for students from rural areas, amalgamating small schools, providing incentives for teacher training, and encouraging private sector activities in education.

Mathematics and science education in primary and secondary schools are the most important factors in the promotion of science capacity building of any country. It enables countries to build an indigenous science based on solid foundation. Consequently, an investigation on how school and environment inputs into the educational production process affects student performance in Mathematics and Science education in Malaysian secondary schools become a very interesting and important study. Furthermore, Hanushek (1979) noted that science professors found that students' performance in mathematics is correlated with their performance in science. Hence education production could be treated as a joint production between performances in these two subjects.

As outlined by O'Sullivan (2000), school achievement depends on five inputs: the school curriculum, educational equipment, the classroom teacher, the home environment, and the

achievement level of the child's classmate. In general, these five inputs to the production function can be divided into three groups: school resources, environmental inputs and peer group effects. In this study, only the effects of school resources, environmental inputs, and gender influence on students' achievement in mathematics and science subjects are investigated. This is due to unavailable data on peer group effects. School resources or inputs include non-teaching recurrent expenditure, teachers with more than five years experience and yearly school hours spent on instruction. On the other hand, environmental inputs include students with at least medium level in home educational index and students with at least medium level in out-of school study index.

There are long debates on the effect of school expenditure on student performance. Hanushek (1986, 1989) posed a major conclusion that there is no relationship between school expenditure and student performance. However, many researchers refuse to accept Hanushek's conclusion and claimed that these findings are based on poor data and inappropriate use of statistical methods. A more refined set of studies carried out by Greenwald et al. (1996), Jacques and Brorsen (2002) and Summer and Wolfe (1977) found that expenditure is one of the factors in influencing students performance.

According to a study carried out by Monk (1994), teachers' experience has a positive impact on performance of students in lower level studies. Many studies (Murnane and Phillips, 1998; Rivkin et al., 1998) noted that students learn more from experienced teachers (those with at least five years of experience).

Carroll (1963) suggests that the time-spent learning and the time needed by a particular student to learn, are crucially important factors influencing achievement. However, both analytic and empirical results suggest extreme caution in viewing increased instructional time as an efficient method for increasing student achievement (Levin and Tsang, 1987).

One of the major ways that students can consolidate and extend classroom learning is to spend time out-of-school studying or doing homework in school subjects. Lewis and Seidman (1994) estimated that if the United States lengthened its school year by three weeks and assigned required summer mathematics homework, it would raise the academic achievement of their students.

One of the variables used to proxy family background is the number of books in the home (Cooper and Cohn, 1997). The existence of books, magazines, encyclopaedias and newspapers is often a sign of a dedication to learning in the household. Researchers have reported that these reading materials are important aspects of the overall home environment.

Brown (1991) found that boys learned significantly more in mathematics while girls learned more in reading. Hence, gender should also be included as one of the factors influencing students' performance in science and mathematics.

This paper uses data from national sources and the Third International Mathematics and Science Study – Repeated (TIMSS-R)[1] carried out in 1999 and employs Vinod's adaptation of Hotelling's Canonical Correlation Analysis (Vinod, 1968) to measure whether school resources and home environment contribute significantly to the production of educational outputs. Canonical correlation analysis is employed because several studies published in the 1980s and 1990s in the United States have argued that in the presence of joint production, ordinary least squares regression (OLS), or even a simultaneous equations system, gives inconsistent estimates. The problem with estimating a regression equation when there are two or more dependent variables is substantially solved by Vinod (1968) by using canonical correlation analysis. Through this

1 Data are available at http://timss.bc.edu. TIMSS is an educational research project conducted by the International Association for the Evaluation of Educational Achievement (IEA), to investigate student achievement in Mathematics and Science in about 40 countries around the world.

technique it is possible to estimate directly a multiple production function, as an implicit function of all the products and inputs. Chizmar and Zak (1984) and Gyimah-Brempong and Gyapong (1991) have also used canonical regression to estimate the joint education production process in the United States.

DATA

In Malaysia, the TIMSS-R sample contains data for 150 schools with a population of 5,713 students. The school samples were selected using a simple random sampling method from all secondary schools in Malaysia including public and private Islamic secondary schools. A single classroom of Form 2[2] lower secondary students is chosen randomly from a number of Form 2 classes in each selected school. After excluding missing values and making necessary corrections, data from 131 schools and 4,854 students are used for this study.

Educational outputs

Educational outputs are measured by school average standardised test scores in mathematics (MATH) and science (SCIENCE) in the TIMSS 1999 for Form 2 students. The school average standardised test scores are calculated by the mean score of standardised test score in mathematics and science from all participants in each sampled school. Because the test scores used in this analysis are given only to Form 2 students, it is necessary to provide an outcome for lower secondary education. The estimation is conducted in level form because the prior achievement of students for value-added estimation is not available.

School inputs

The four variables that were used to represent school inputs are per pupil non-teaching expenditures (PPNTE), pupil teacher ratio (PTR), teaching experience (TE), and instructional hours (INSHRS). PPNTE is obtained by averaging the total expenditures of library facilities, counselling services, other recurrent and miscellaneous expenses on the total school enrolment. According to Jacques and Brorsen (2002), test scores were negatively related to expenditure on student support. Thus negative relationships between PPNTE and student achievements in Mathematics and Science are expected.

PTR is the pupil teacher ratio in a school. The total number of teachers is derived by assigning 1 to a full-timer and 0.5 for a part-timer. For each school, PTR is calculated as the total school enrolment divided by the total number of teachers. Assuming that students learn less in a bigger class, a negative relationship is expected between PTR and the educational outputs.

The percentage of teachers with more than five years experience is represented by TE. A positive relationship is anticipated since students are expected to learn more from experienced teachers.

The influence of instructional hours on students’ performance in these two subjects is considered by INSHRS, the percentage of yearly school hours spent on instruction. It also implies the effectiveness of schools in optimising the school hours on instructional work. It is obtained by dividing the total of yearly instructional hours excluding lunch breaks, study hall time and after school activities by total school hours in a school year. Brown and Saks (1987) found that more instructional hours in classrooms do increase learning and hence a positive relationship is expected.

[2] Form 2 is equivalent to eighth grade.

Environmental inputs

In our study, home educational resources index (HER) and an out-of-school study time index (OST) are used to represent environmental inputs. HER was derived from students' reports on the availability of books in the home; educational aids in the home (computer, study, desk/table for own use, dictionary); and their parents' education. It is defined as the percentage of students or participants with at least medium level in home educational resources index. A positive relationship is expected indicating that academic support in the home environment encourages learning.

OST is the percentage of students with at least a medium level out-of-school study time index and it takes into consideration the relationship between time spent doing homework and school average performance in mathematics and science. A positive relationship indicates that as the amount of time a student spends on doing his or her homework increases, academic achievement improves.

Gender influence

The gender influence is considered by including in the analysis the FEMALE variable as the percentage of female students in class.

The analyses were carried out using the Statistical Package for Social Science (SPSS) version 10.0 and StatistiXL version 1.1.

THE MODEL

The joint production function of a generalised Cobb-Douglas form, can be written as:

$$\mathrm{MATH}_1^{\alpha_1}\ \mathrm{SCIENCE}_2^{\alpha_2} = \beta_0 \mathrm{PPNTE}_1^{\beta_1}\ \mathrm{PTR}_2^{\beta_2}\ \mathrm{TE}_3^{\beta_3}\ \mathrm{INSHRS}_4^{\beta_4}\ \mathrm{HER}_5^{\beta_5}\ \mathrm{OST}_6^{\beta_6}\ \mathrm{FEMALE}_7^{\beta_7}\mu \quad (1)$$

where

MATH	=	School average standardised mathematics test scores,
SCIENCE	=	School average standardised science test scores,
PPNTE	=	Per pupil non-teaching recurrent expenditure (RM),
PTR	=	Pupil teacher ratio,
TE	=	Percentage of teachers with more than five years experience,
INSHRS	=	Percentage of yearly school hours spent on instruction,
HER	=	Percentage of students with at least medium level in home educational resources index,
OST	=	Percentage of students with at least medium level in out-of-school study index, and
FEMALE	=	Percentage of female students in class.

μ is a stochastic error term, γ and α_p, β_q are coefficients to be estimated.

Rewriting equation (1) by taking the natural logs gives

$$\alpha_1 \ln\mathrm{MATH} + \alpha_2 \ln\mathrm{SCIENCE} = \beta_1 \ln\mathrm{PPNTE} + \beta_2 \ln\mathrm{PTR} + \beta_3 \ln\mathrm{TE} + \beta_4 \ln\mathrm{INSHRS} + \beta_5 \ln\mathrm{HER} + \beta_6 \ln\mathrm{OST} + \beta_7 \ln\mathrm{FEMALE} + \ln \mu$$

since β_0 is an efficiency index ($0< \beta_0 \leq 1$) where $\beta_0 =1$ if and only if production is technically efficient.

The marginal elasticity (ME) between MATH and PPNTE is

$$\mathrm{ME(MATH,\ PPNTE)} = \frac{\partial \ln \mathrm{MATH}}{\partial \ln \mathrm{PPNTE}} = \frac{\beta_1}{\alpha_1}$$

and the corresponding marginal productivity (MP) is

$$\mathrm{MP(MATH,\ PPNTE)} = \frac{\mathrm{MATH}}{\mathrm{PPNTE}}\ \mathrm{ME(MATH,\ PPNTE)}$$

From MP, the marginal rates of technical substitution (MRTS) between inputs, for example between PPNTE and PTR, can be calculated as follows

$$\text{MRTS(PPNTE, PTR)} = \frac{MP_{PTR}}{MP_{PPNTE}}.$$

The MEs and corresponding MPs and MRTS with respect to the other inputs are similarly defined. The marginal rate of transformation (MRT) between the two outputs is

$$\text{MRT(MATH, SCIENCE)} = \frac{\partial \ln \text{MATH}}{\partial \ln \text{SCIENCE}} = -\frac{\alpha_2 / \text{SCIENCE}}{\alpha_1 / \text{MATH}}.$$

RESULTS

Table 1 shows the mean values and standard deviation as well as the skewness of each variable considered in the analysis. It is not surprising that the mean scores for mathematics and science are around 50 since the scores used in this study are standardised test scores. Data on instructional hours, home educational resources and out-of-school study time are slightly negatively skewed. All of the distributions of the other variables are approximately normal. It should be noted that the mean percentages for INSHRS, OST and HER are on the high side, suggesting most of the students in schools under this study had very high instructional hours, out of school study time and home educational resources.

Table 1. Description of school outputs, school and environmental inputs and gender influence

Variables	Mean	Standard Deviation	Skewness
School Outputs			
Mathematics Score (MATH)	50.14	7.25	0.07
Science Score (SCIENCE)	50.01	6.54	0.18
School Inputs			
Per Pupil Non-teaching Expenditure (PPNTE)	66.74	1.05	0.33
Pupil Teacher Ratio (PTR)	19.12	3.01	-0.20
Teacher Experience (TE) (%)	44.71	23.17	0.30
Instructional Hours (INSHRS) (%)	92.78	5.87	-0.57
Environmental Inputs			
Home Educational Resources (HER) (%)	72.22	18.51	-0.83
Out-of-School Study Time (OST)	96.73	3.90	-1.13
Gender Influence			
Female Students (FEMALE) (%)	53.97	18.05	1.68

N = 131

The estimation results of the generalised production function using Vinod's procedure are given in Table 2 for four different models. Model 1 contains all of the variables discussed. On the other hand, Model 2 takes into account all of the variables except TE. Model 3 is the results after deletion of INSHRS and Model 4 is the model without TE and INSHRS. For all of these models, the canonical estimates are significant overall, which support the hypothesis of jointness in the education production. The canonical redundancy analysis (R_d) did not differ significantly between these four models which are just over 50 per cent, indicating that more than 50 per cent of the total variance for mathematics and science are explained by these variables.

All of the parameter estimates have the expected signs and the magnitude of the coefficients express the importance of an input from independent canonical variates with regard to the dependent canonical variates in obtaining a maximum correlation between sets. Both TE and INSHRS show a relatively small value in their parameter estimates. In order to investigate whether TE and INSHRS variables are significant in influencing school performance at the national level, we re-estimated equation (1) without TE, INSHRS and both TE and INSHRS. The

parameter estimates in Model 1 and Model 2 are remarkably stable. The overall canonical correlations also remain stable. On the other hand, dropping the INSHRS variable did reduce the redundancy index slightly and change the parameter estimates. Hence, adding TE to the national education production function does not give additional explanatory power and can be ignored. On the other hand, INSHRS should not be dropped although it has the minor importance on determining school performance.

Table 2. Canonical fit estimates of the Cobb-Douglas production function

Variables	Model 1	Model 2	Model 3	Model 4
School Outputs				
MATH	0.825	0.817	0.916	0.897
SCIENCE	0.186	0.194	0.089	0.110
School Inputs				
PPNTE	-0.389	-0.390	-0.368	-0.371
PTR	-0.215	-0.215	-0.206	-0.205
TE	0.008	Omitted	0.022	Omitted
INSHRS	0.098	0.099	Omitted	Omitted
Environmental Inputs				
HER	0.491	0.493	0.485	0.605
OST	0.215	0.217	0.195	0.286
Gender Influence				
FEMALE	0.117	0.116	0.124	0.220
N	131	131	131	131
Rao's F	10.025	11.206	11.016	12.731
Wilk's Lambda	0.403	0.418	0.423	0.437
Bartlett's Chi-square	113.60	109.458	107.951	104.407
P value	0.000	0.000	0.000	0.000
R_d (%)	51.30	51.40	50.08	50.16

In short, Model 2 is chosen for further analysis. Based on the absolute value of parameter in Table 2, the two educational inputs with highest contribution to the canonical variate are home educational resources and per pupil non-teaching expenditure. Not surprisingly, this reveals that the home educational resources' effect is a strong influence on academic performance, and this conclusion is in line with the Coleman Report and most previous studies.

Using the parameter estimates in Model 2, the marginal rate of output transformation showing the relationship among outputs and the marginal elasticity linking the input and output can be calculated, and these results are shown in Table 3.

Table 3. Canonical fit estimates of the Cobb-Douglas production function

Variables	Parameter estimates	ME (MATHS)	ME (SCIENCE)	MP (MATHS)	MP (SCIENCE)
School Outputs					
MATH	0.817				
SCIENCE	0.194				
School Inputs					
PPNTE	-0.390	-0.478	-2.012	-0.359	-1.508
PTR	-0.215	-0.264	-1.110	-0.691	-2.903
INSHRS	0.099	0.122	0.516	0.066	0.276
Environmental Inputs					
HER	0.493	0.603	2.538	0.402	1.688
OST	0.217	0.265	1.118	0.138	0.578
Gender Influence					
FEMALE	0.116	0.142	0.599	0.132	0.555

The marginal products of INSHR, OST and FEMALE are quite small. On the other hand, a one per cent increase in the percentage of students with at least medium level of home educational resources index would lead to an increase of around 1.7 in science score and 0.4 increase in mathematics score. It is also found that larger pupil teacher ratio leads to a deterioration of

educational quality and output. Similarly, a RM1 increase of per pupil non-teaching expenditure will decrease the science score by 1.5 points and mathematics score by 0.4 points. This result indicates that non-teaching expenditure does not contribute directly to learning. In general, taking these educational inputs into consideration, the marginal effect on science achievement outperformed the mathematics achievement.

From the marginal products, marginal rates of substitution are calculated and presented in Table 4. In this analysis, it is unlikely that we can compensate for the number of female students, the parent's educational level or the home educational resources with other variables. However, an increase in time spent in instruction can compensate for low level of out-of-school study time index. The result in Table 4 indicates that the MRTS of INSHRS for OST (which equals MP_{INSHRS}/MP_{OST}) has a value of 2.096. This result indicates that schools with majority low level out-of-school study time can compensate for deficiencies in this environment factors with extra instructional hours of 2.1 per cent is necessary to offset a one per cent in percentage of students with at least medium level in out-of-school study time index (OST) and keep output constant.

Table 4. Marginal rates of substitution derived from the Cobb-Douglas production function estimates of Table 3

	PPNTE	PTR	INSHRS	HER	OST	FEMALE
PPNTE	1.000	1.926	-0.183	-1.120	-0.383	-0.368
PTR	0.519	1.000	-0.095	-0.581	-0.199	-0.191
INSHRS	-5.468	-10.530	1.000	6.122	2.096	2.012
HER	-0.893	-1.720	0.163	1.000	0.342	0.329
OST	-2.609	-5.024	0.477	2.921	1.000	0.960
FEMALE	-2.718	-5.233	0.497	3.043	1.042	1.000

CONCLUSIONS

This study uses data from national sources and TIMSS-R carried out in 1999, together with canonical regression to investigate the importance of the school and environmental inputs and gender influence in the production of mathematics and science education. These analyses demonstrate the significant effects of home educational resources on the Malaysian school's mathematics and science achievement. Furthermore, pupil teacher ratio appears to be the most productive input among the educational inputs considered. Last but not least, it finds that instructional hours can be used to offset the low level of out-of-school study time.

ACKNOWLEDGEMENT

We would like to thank the TIMSS group for conducting such a comprehensive study and sharing the data with the public. However, we bare full responsibility for any errors in this paper.

REFERENCES

Akerhielm, K. (1995). Does class size matter? *Economics of Education Review,* 14 (2), 229-241.

Brown, B. W. (1991). How gender and socialeconomic status affect reading and mathematics achievement. *Economics and Education Review*, 10 (4), 343-357.

Brown, B.W. and Saks, D.H. (1980). Production technologies and resource allocations within classroom and schools: theory and measurement. In R. Dreeben and J.A. Thomas (eds), *Analysis of Educational Productivity.* Cambridge, Mass: Ballinger.

Carroll, J.B. (1963). A model for school learning. *Teachers College Record,* 64, 723-733.

Chizmar, J.F. and Zak, T.A. (1984). Canonical estimation of joint educational production functions. *Economics of Education Review,* 3 (1), 37-43.

Cooper, S.T. and Cohn, E (1997). Estimation of a frontier production function for the South Carolina educational process. *Economics of Education Review*, 16 (3), 313-327.

Greenwald, R., Hedges, L.V. and Laine, R.D. (1996). The effect of school resources on student achievement. *Review of Educational Research,* 66 (3), 361-396.

Gyimah-Brempong, K. and Gyapong, A.O. (1991). Characteristics of education production functions: An application of canonical regression analysis. *Economics of Education Review,* 10 (1), 7-17.

Government of Malaysia (1996). *Seventh Malaysia Plan.* Percetakan Nasional Malaysia Berhad, Kuala Lumpur, Malaysia.

Hanushek, E.A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14(3), 351-388.

Hanushek, E.A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature,* 26, 1141-1177.

Hanushek, E.A. (1989). The impact of differential expenditures on school performance. *Educational Researcher,* 18 (4), 45-51.

Jacques, C. and Brorsen, B.W. (2002). Relationship between types of school district expenditures and student performance. *Applied Economics Letters*, 9, 997-1002.

Levin, H.M. and Tsang, M.C. (1987). The economics of student time. *Economics of Education Review,* 6 (4), 357-364.

Lewis, K.A. and Seidman, L.S. (1994). Math-time capital matters: A cross-country analysis. *Economics of Education Review,* 13 (3), 215-226.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13 (2), 125-145.

Murnane, R.J. and Phillips, B.P. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review,* 1 (4), 453-465.

Rivkin, S.G. Hanushek, E.A. and Kain, J.F. (1998). *Teacher, Schools and Academic Achievement.* Working Papers, Cambridge: National Bureau of Economic Research.

Summers, A. and Wolfe, B. (1977). Do schools make a difference? *American Economic Review,* 67, 639-652.

O'Sullivan, A. (2000). *Urban Economics*, 4th ed. Boston: Irwin McGraw-Hill, 622-654.

Vinod, H.D. (1968). Econometrics of joint production. *Econometrica,* 36 (2), 322-336.

❖IEJ