DOCUMENT RESUME

ED 370 990 TM 021 584

AUTHOR	Shen, Linjun
TITLE	The Growth Patterns of General Medical Achievement.
PUB DATE	Apr 94
NOTE	21p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Achievement Gains; Clinical Teaching (Health Professions); High Schools; *Institutional Characteristics; *Knowledge Level; Medical Education; Medical Schools; *Medical Students; *Osteopathy; Premedical Students
IDENTIFIERS	*National Board of Osteopathic Medical Examiners

ABSTRACT

This longitudinal study investigates the growth of medical achievement as a multilevel process and emphasizes the structure of the growth. Subjects were students in all 15 U.S. osteopathic medical schools, a total of 1,060 (78 percent of the 1987 osteopathic cohort). Students took appropriate portions of the National Board of Osteopathic Medical Examiners' 3-part series of examinations in 1989, 1991, and 1992. Results indicate a substantial overall gain in medical achievement, with an achievement increase of 18 percent over the 3 years. Results also provide empirical evidence of discontinuity among the preclinical, clinical, and residency phases of medical education. Some of the schools had statistically zero, or negative growth, but there were substantial positive gains during clinical education for students in other schools. Institutional differences have sustained effects on student achievement growth both during and after medical school. An appendix discusses hierarchical linear modeling of the growth in general medical knowledge. One table and four figures present study data. (Contains 7 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ERIC

ED 370 990

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LINJUN SHEN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Growth Patterns of General Medical Achievement

Linjun Shen

National Board of Osteopathic Medical Examiners

Paper presented at the Annual Meeting of the
American Educational Research Association
New Orleans, April 1994.

TM021584

BEST COPY AVAILABLE

The Growth Patterns of General Medical Achievement

Growth is thc fundamental objective of education. Assessing growth evaluates the effectiveness of school education most directly and efficiently. The purpose of this study is to investigate the growth of medical students' general medical knowledge along the whole period of medical cducation.

Many studies find that the growth of student medical knowledge is linear and positive (Donovan, Salzman, & Allen, 1969; Willoughby & Hutcheson, 1978; Albers, Does, and et al, 1989; Verwijnen, van der Vleuten & Imbos, 1990). Common methodological deficiencies of these studies are: first, they did not address the multilevel nature of the growth, second, the scope of the studies were relatively small, especially, number of schools examined were small. Consequently, the growth of medical achievement was simplified.

This longitudinal study investigates the growth of medical achievement as a multilevel process and emphasizes the structure of the growth. An assumption of this study is that medical knowledge is an entity and it is operationable. Basic science knowledge and clinical science knowledge, which are traditionally measured independently, are conceptualized as two interrelated components of the general medical knowledge.

Method

Subjects

The subjects of this longitudinal study were students in all 15 U.S. osteopathic medical schools who started osteopathic medical education in 1987. This cohort was the latest cohort available for this study and reflects the most recent changes of osteopathic medical education in the United States.

All subjects completed basic science education and took the NBOME Part I examination for the first time in June 1989, completed clinical science education and took Part II for the first time in March 1991, and completed at least six months of the first year of residency and took Part III for the first time in February 1992. Students in the 1987 cohort who had not completed these three examinations were excluded from this study. Because of policy differences, students in one school were excluded. In spite of these exclusions, 1060 subjects, or 78% of the 1987 osteopathic student cohort, were included in this analysis.

Instruments

The NBOME Board's three-part series of examinations was developed for the sole purpose of licensing osteopathic physicians. The NBOME Part examinations are primary

care–oriented and are intended to test candidates' medical knowledge and their ability to apply the knowledge, concepts, and principles of osteopathic medicine in solving problems related to maintaining health and combating disease.

Part I includes a total of about 850 multiple–choice questions in the basic sciences unequally divided among anatomy, physiology, biochemistry, pharmacology, pathology, microbiology, and osteopathic principles. Part II contains about 940 multiple–choice questions in the clinical sciences unequally divided among surgery, psychiatry, obstetrics/gynecology, community medicine, pediatrics, internal medicine, and osteopathic principles. Part III has about 600 multiple–choice questions covering the same clinical disciplines as Part II. The majority of Part III items are written clinical problems written by clinicians.

Measurement Scale

To study the growth of medical knowledge, students need to be measured by a single scale over the entire period of medical education. The measurement tool must have a high psychometric comparability so that measures taken at different time points during the medical program will have the same qualitative and quantitative explanation. The comparability criterion requires that measures be taken at each stage of the medical program on the same scale, and, that the measurement scale must have equal intervals assuring that equal score differences at different ability levels would have identical quantitative meanings.

Subjects in this study were measured three times during their medical program by the Part I in June 1989 (A891), Part II in March 1991 (B911), and Part III in February 1992 (C921) separately. Three exams were constructed independently from separate blueprints. Qualitatively each test assesses different components of medical knowledge. Quantitatively, three exams were on three independent measurement scales. Those three measures did not have the essential comparability described above. Therefore, they were not valid measures for a longitudinal analysis of medical achievement.

A measurement scale of general medical knowledge (GMK) was constructed by equating A891, B911, and C921 via 6 other NBOME licensing examinations. The equating adopted the Rasch Measurement one–step equating approach. This procedure hypothesizes a single "super" exam comprised of all nine exams to be equated according to the overlapping structure among the nine exams. Figure 1 represents the design of this one–step equating. Under this design, the 1987 cohort was measured three times by the same "super" exam at three different time points along their medical education.

Calibrating this "super" exam accomplishes the equating. The global calibration yields a single measurement scale defined by items from all participating exams. Since the NBOME Part I, II, and III examinations together cover all the major concepts and principles of the entire medical sciences, GMK practically defined a holistic concept of medical knowledge.

A careful analysis of dimensionality, scale equity, and sample indifference of the GMK scale indicated that the GMK scale had all psychometric properties required for growth analysis (Shen, 1993).

HLM Modelling

Individual growth is a three-level phenomenon. As Bryk and Raudenbush (1992) conceptualize, this type of research problem has three foci: the individual growth of students over the course of the academic years (or segment of a year), the effects of personal characteristics and individual educational experiences on student learning, and how these relations are in turn influenced by schools and the specific features of schools.

> Correspondingly, the data have a three-level hierarchical structure. The Level-1 units are the repeated observations over time, which are nested within the Level-2 units of persons, who in turn are nested within the Level-3 units of classrooms or schools" (Bryk & Raudenbush, 1992, p. 2).

The Hierarchical Linear Model (HLM) addresses academic growth most appropriately (Bryk & Raudenbush, 1987). This study applies HLM to the growth of medical achievement.

The Level-1 model fitted three gain parameters. The Level-2 and 3 models were unconditional models. Since the purpose of this study is to analyze the general growth patterns of medical achievement, no student and school variables were fitted. Appendix describes the models.

Results

The HLM3 version 2.0 developed by Bryk, Raudenbush, and Congdon (Bryk, Raudenbush & Congdon, 199) executed the unconditional three-level HLM analysis defined by Equations 8, 9, and 10. 1060 level-2 units and 14 level-3 units participated in this analysis. The program stopped after 450 iterations due to small change in likelihood function. The analysis was well executed.

Table 1 summarizes this unconditional HLM analysis. After a brief assessment of the model adequacy, the following presentation of the results focuses on growth in general and growth variation as captured by this unconditional models.

Model Adequacy

Significance test of parameter variance

Bryk and Raudenbush point out that if the χ^2 test is rejected for the null hypothesis that the parameter variance is zero, the investigator may conclude that there is random variation in the parameter (Bryk & Raudenbush, 1992). As panel 2 of Table 1 shows, χ^2 tests for homogeneity of variance for r_{tij} , Level–1 parameters, and for u_{t0j} , Level–2 parameters, were all significant at .001 level. This suggests that there was substantial growth variation among students within schools, and substantial variation of mean growth across schools. Clearly, hierarchical modelling was needed to explain the large amount of growth differences among students and schools.

Reliability of parameter estimates

The third panel of Table 1 provides high reliability estimates for each of the model parameters. Except the reliability of π_{3ij} , the gain parameter during the first year of residency, all other parameter reliabilities are in the range of .78 to .94. The relatively lower reliability, .53, of π_{3ij} , suggests that the multiple–choice question examinations are less sensitive to individual achievement gained from practice during residency training.

Reliability reflects the degree to which the true underlying parameters varied from student to student or school to school and the precision with which each individual's growth trajectory and each school's regression were estimated (Bryk & Raudenbush, 1992). High reliabilities of π_{tij} and β_{t0j} were essential for this hierarchical linear model analysis in two respects. First they suggest the adequacy of the three–level modelling. Second, high reliabilities of model parameters provide evidence of high psychometric qualities of the one–step equating which provide the measurements for growth parameters.

Growth in General

The growth trajectory

Figure 2 depicts the mean growth trajectory for the total group. Clearly, the overall growth trajectory is not linear. The growth between 1989 and 1991 was flat. It picked up after the end of clinical education.

$\gamma_{300} + \gamma_{100}$ is the amount of the overall GMK growth for the average student between the end of preclinical education and the end of the first year of residency training. According to the first panel of Table 1, the overall growth was .103 logits, or an 18.6% gain from the status at the end of preclinical education. A one–tailed dependent t–test of $\gamma_{300} + \gamma_{100}$ indicated that a

growth of .103 logits was significantly greater than zero ($p<.01$). γ_{100} , the average individual GMK growth between the end of preclinical education and the end of clinical education, was −.015 logits, not significant at .05 level ($p>.340$). Therefore, statistically, there was no GMK growth during the clinical medical education period. γ_{300} , the growth taking place during the first year of residency training, was .570, significant at .05 level ($p<.01$).

Variation of growth

The variability of GMK growth shrank. More variation of gain was observed during the early period between 1989 and 1991 than during the period between 1991 and 1992. The total variance of GMK gain at the first stage, $r_{1ij} + u_{10j}$, was .057, while the total variance of GMK at the second stage, $r_{3ij} + u_{30j}$, was only .021. These results confirm the findings of early study (Shen, 1993) that the observed achievement variances during the early stages were larger than those at the later stages.

The ratio of gain over its variance for gain between 1989 and 1991 was .26, whereas for gain between 1991 and 1992 was 5.6. In other words, during the first stage, there was little average GMK gain but larger variation. During the second stage, the gain was 7.8 times larger but the variation was 2.7 times smaller. This comparison suggests that, in order to explain the small negative average gain, more efforts are needed to study the large gain variation at the first stage.

Decomposition of parameter variance

This analysis decomposed the growth variance into the variance caused by the differences among individual students and the variance caused by the school dissimilarities. As Table 1 suggests, the growth of medical knowledge substantially varied both within and between schools. A large amount of the variance for the growth of medical knowledge were due to differences at the person level. For the 1989–1991 gain , 80.7% of the variance, or $r_{1ij} / (r_{1ij} + u_{10j})$ was due to differences among students, and 19% of the variance was caused by school differences. For the 1991–1992 gain, 91.5% of the variance, or $r_{3ij} / (r_{3ij} + u_{30j})$, was due to the person level variables, only while 8.5% of the variance came from school effects.

Correlations between gains

The panel 4 of Table 1 indicates the correlation between π_{1ij} ar. π_{3ij} was positive, but for β_{10j} and β_{30j} the correlation was negative. This suggests a tracking effect within schools where initial differences were somehow predictive of subsequent learning. Compared with others within the same schools, students with larger gains in the first period were more likely to have higher gains during the second period. Nevertheless, this relationship was not very strong.

By contrast, the correlation between β_{10j} and β_{30j} was −.28, stronger than the same relationship within schools. This implies that school as a unit which gained more in the first period tended to gain less in the second period compared with other schools which gained less in the first period. This further suggests that the institutional differences among schools were beneficial to students in schools which gained less in the first period but disadvantageous to students in schools which gained more in the same period. Since only schools gaining less in early years could gain more later, a student's GMK growth was limited by the potential ceiling set by the school attended.

Correlations between gains and the status at the end of clinical education

Again, the relationship between the GMK gain at the first stage and the 1991 status was different at the person level from that at the school level. The correlation between π_{1ij} and π_{2ij} was −.11, while the correlation between β_{10j} and β_{20j} was .52. In other words, at the school level, the higher the school mean achievement in 1991, the more the school mean gain from 1989 to 1991. Within a school, students with lower status in 1991 tended to gain more from 1989 and 1991 compared with other students with higher status in 1991. Therefore, students in schools with a high quality of clinical education had a better chance to grow in the first stage.

Interestingly, the correlations between 1991 status and the gain 2 were negative at both student and school levels with −.37 for the correlation between π_{2ij} and π_{3ij} , and −.81 for the correlation between β_{20j} and β_{30j}. The negative relationship at the school level was much stronger than at student level. This implies schools had a large impact on the GMK growth at the second stage.

Growth by schools

Decomposition of variances at student and school levels and comparisons of correlations at two levels all suggest that differences among school substantially influenced students' GMK growth. Figure 3 demonstrates the variation of mean growth among schools.

By reviewing the shape of the growth trajectories, two types of growth patterns occur for schools. As Figure 4 shows, half of the 14 schools had a continuous growth pattern, while other schools had a V−shaped growth pattern with a considerable decline at the end of clinical education. The distinction between two types of schools suggests that the "no gain" phenomenon at the first stage for the overall growth need more careful analysis.

Discussion

General Effectiveness of the Current Medical Education

Consistent with other longitudinal studies, results of this study indicate a substantial overall gain of medical achievement. Between the end of preclinical education and the end of the first year of residency, or in three years alone, the achievement increased by 18%.

Discuntinuity of Three Phases of Medical Education

Many authors believe that structural trichotomization exists in the current medical education. The finding that the mean GMK growth of the total group from the end of preclinical education to the end of clinical education is statistically zero provides empirical evidence of discontinuity among the preclinical, clinical, and residency phases of medical education.

This study, on the other hand, also suggests that zero growth is not an inevitable reality for medical education. Though some of the 14 schools had statistically zero or negative growth, the analysis shows significant positive gains during clinical education for other schools.

Variations of Growth

Decomposition of growth variance, and correlations among growth parameters indicate that institutional differences had substantial effects on student achievement growth both during and after medical school. Practically, the proportion of school level variance would be higher, if student level variance is adjusted for the student background differences such as MCAT scores. Two types of school mean growth demonstrate that schools not only differed in the amount but also in the patterns of growth.

Methodological Implications

The research methodology of this study is uncommon to traditional research in medical education. The differences are paradigmatic. Four features represent the methodology of this study: first, a holistic conceptualization of medical knowledge – General Medical Knowledge, second, operationalization of General Medical Knowledge, third, longitudinal inspectation of medical education, and fourth, multi-level analysis of medical achievement. The results demonstrate that the methodology is appropriate for the research objectives.

Admittedly, this study has some limitations. First, the findings may not generalize to

ERIC

allopathic medical education. Second, this study has three time points. More mearures along the medical program will depict the growth more accurately. Finally and most importantly, by limiting itself to academic achievement, this analysis does not address the interactions between academic growth and the parallel growth of other components of clinical competence.

References

Albers, W., Does, R. J. M. M., Imbos, Tj. & Janssen, M. P. E. (1989). A stochastic growth model applied to repeated tests of academic knowledge. Psychometrika, 54, 451-466.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101, 147-158.

Bryk, A. S. & Raudenbush, S. W. (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Sage: Newbury Park.

Donovan, J. C., Salzman, L. F., & Allen, P. Z. (1969). Patterns of learning in medical school. Journal of Medical Education, 44, 589-594.

Shen, L. Constructing a measure for longitudinal medical achievement studies by the Rasch Model one-step equating. Paper presented at the American Educational Research Association annual meeting, Atlant , Ga, April 1993.

Verwijnen, M., van der Vleuten, C. & Imbos, Tj. (1990). A comparison of an innovative medical school with traditional schools: an analysis in the cognitive domain. In Z. M. Nooman, H. G. Schmidt & E. S. Ezzat (Eds.), Innovation in Medical Education: An Evaluation of Its Present Status. Springer: New York.

Willoughby, T. L. & Hutcheson, S. J. (1978). Edumetric validity of the quarterly profile examination. Educational and Psychological Measurement, 38, 1057-1061.

Appendix

Hierarchical Linear Modelling of the GMK Growth

HLM modelling in this study was different from conventional HLM analysis in two aspects. First, it took full advantage of Rasch model scaling to adjust for the measurement errors and misfit of GMK measures. By doing so, the measurement error was washed out from the models' overall random error term. Therefore, the growth would be better estimated. Second, this study used gains as the model parameters instead of observed measures at each time point. Since gain is a more direct indicator of growth, this parameterization would present growth more conveniently and effectively

A Model for Parameterizing Gains

The model for gains was based on the IRT measurement model:

$$Y_{tij} = \delta_{tij} + e_{tij} \qquad (1)$$

for $i = 1, \ldots, 1060$ subjects of school j, $j = 1, \ldots, 15$, each of whom is observed on t occasions, $t = 1, 2, 3$; where

Y_{tij} = the observed status of individual i of school j at time t ;

e_{tij} = measurement error, it is assumed independent and normally distributed with mean of zero and assumed known variance V_{tij} ,

To transform δ_{ij} to gain parameter π_{ij} , set

$$T\delta_{ij} = \pi_{ij} \qquad (2)$$

where

$$T = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \qquad (3)$$

such that π_{ij} represents the gain matrix

$$\pi_{ij} = \begin{bmatrix} Y_{2ij} - Y_{1ij} \\ Y_{2ij} \\ Y_{3ij} - Y_{2ij} \end{bmatrix} \quad (4)$$

Equivalent to the model 2,

$$\delta_{ij} = (T'T)^{-1}T'\pi_{ij} \quad (5)$$

Let $A=(T'T)^{-1}T'$, then

$$\delta_{ij} = A\pi_{ij} \quad (6)$$

where

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad (7)$$

In this gain parameterization, the base was set at the status at second time point in stead of first time point. The reason to do so was that a preliminary analysis found, at school level, the GMK status at the end of clinical education was the turning point for medical knowledge growth. By setting the base level at the second time point, it would be more convenient to investigate the relationships between the time 2 status with the gain from time 1 to time 2 or the gain from time 2 to time 3.

Level–1 Model

The Level–1 model represented each student's growth trajectory which depended on unique growth parameters. By combining the equations (1) and (4), the Level–1 model fitted three gain parameters:

$$Y^{*}_{tij} = A^{*}\pi_{ij} + e_{tij} \quad (8)$$

In this model, Y^{*}_{tij} and A^{*} were pre–weighted by $(1/V_{tij})^{1/2}$. As a result, $e_{tij} \sim N(0, 1)$.

According to this model, Y_{tij}, the observed status of general medical knowledge for subject

i of school j at time t, was a function of π_{ij}, the gain from the previous stage plus random error e_{tij}. By specifying the gains as random variables, this model reflected the reality that growth varied across students. As a result, level–2 and level–3 models were built to represent these parameter variations.

Level–2 Model

Level–2 models were unconditional models. No student background variables were fitted. Each of the gain parameters were specified as the function of the mean growth of a school and random variation of individuals within the school.

$$\pi_{tij} = \beta_{t0j} + r_{tij} \tag{9}$$

where,

β_{t0j} is the unadjusted mean gain of school j at time t;
r_{tij} is random error with mean of 0 and covariance matrix T_{π}.

At this stage, gain parameters in the level–1 model became outcomes for the level–2 model. The unconditional level–2 model estimated the variability of gains across subjects within schools.

Level–3 Model

Similarly, Level–3 models were also unconditional. No any school characteristics variables were included. Each of the school mean growth, β_{t0j}, was treated as the function of grand mean growth and the variation of school means from the grand mean. For each of the unadjusted school mean gains

$$\beta_{toj} = \gamma_{too} + u_{toj} \tag{10}$$

where

γ_{t00} is the grand mean growth at time t;
u_{t0j} is random error. It is assumed that u_{t0j} is distributed multivariate normal with mean 0 and covariance matrix T_{β}.

Gain parameters in the level–2 model became outcomes for the level–3 model. The unconditional level–3 model estimated the total variation of gain parameters across schools.

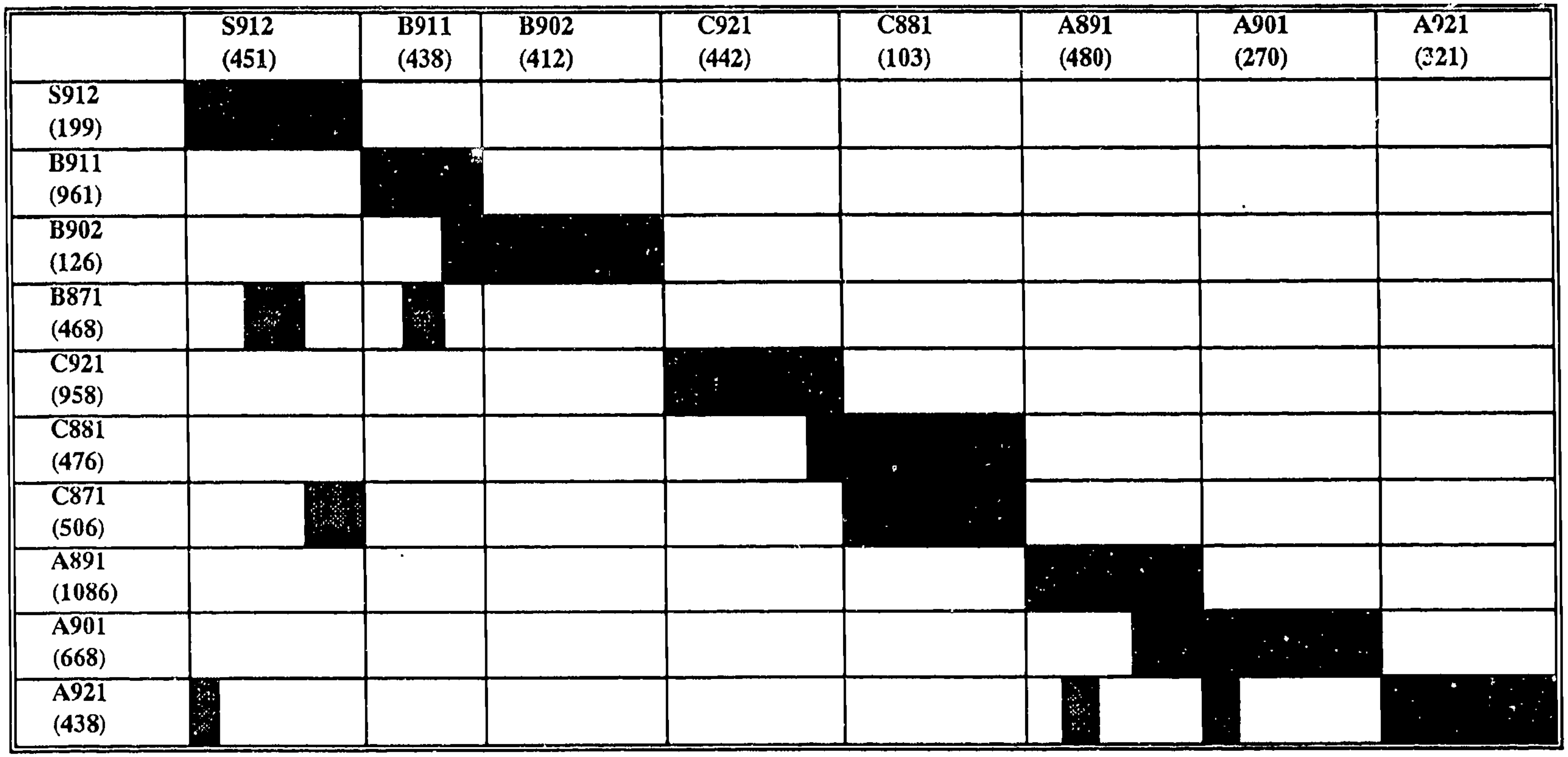

Figure 1. The data matrix for the Rasch one–step equating. Each exam is represented by four characters. The first character designates the Part with A, B, and C representing Part I, II, and III respectively. The second and third characters represent the year of the exam. The last character designates the administration. Therefore, for example, A921 represents the first administration of Part I in 1992. Numbers in parentheses in the first row identify the numbers of items of each exam selected for the equating. Numbers in parentheses in the first column identify the numbers of persons of each exam selected for the equating. In total, 2814 items and 5168 persons are included.

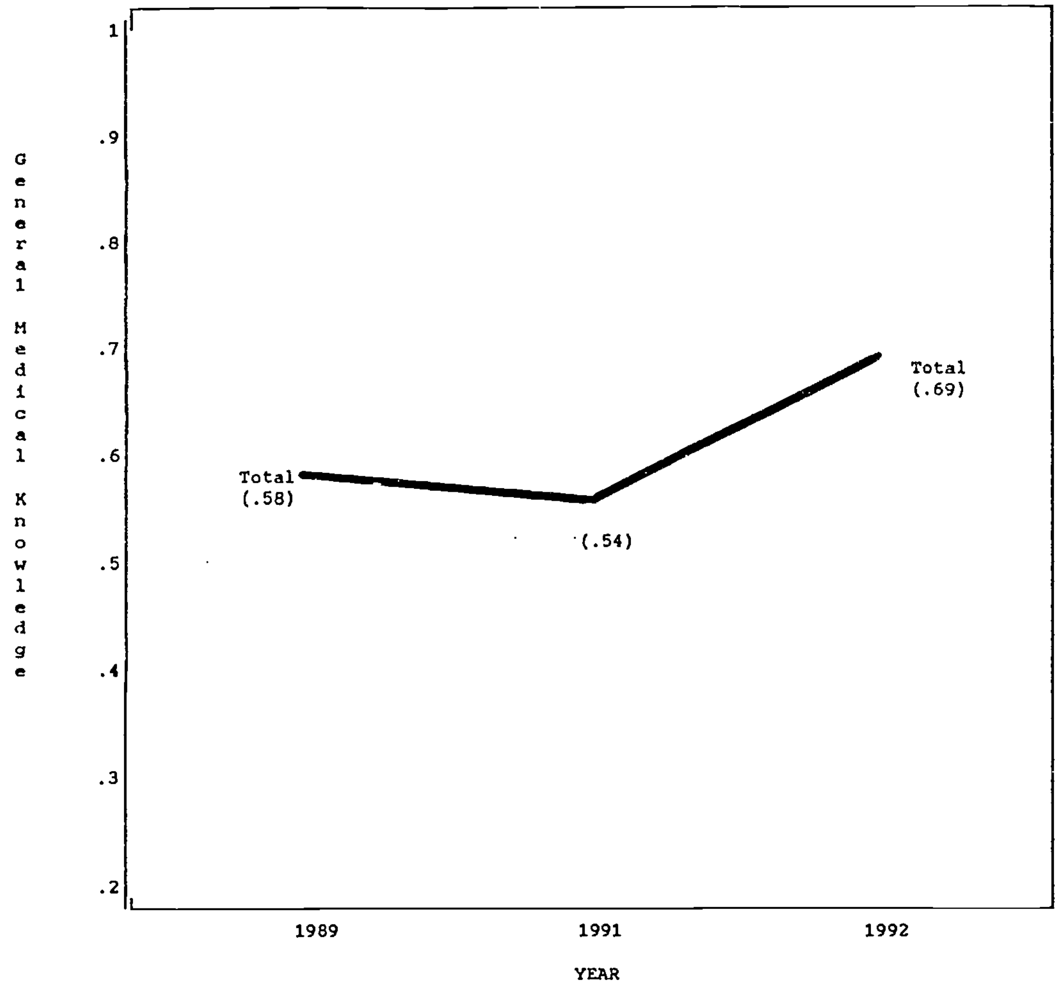

Figure 2. Mean MGK growth for the total group.

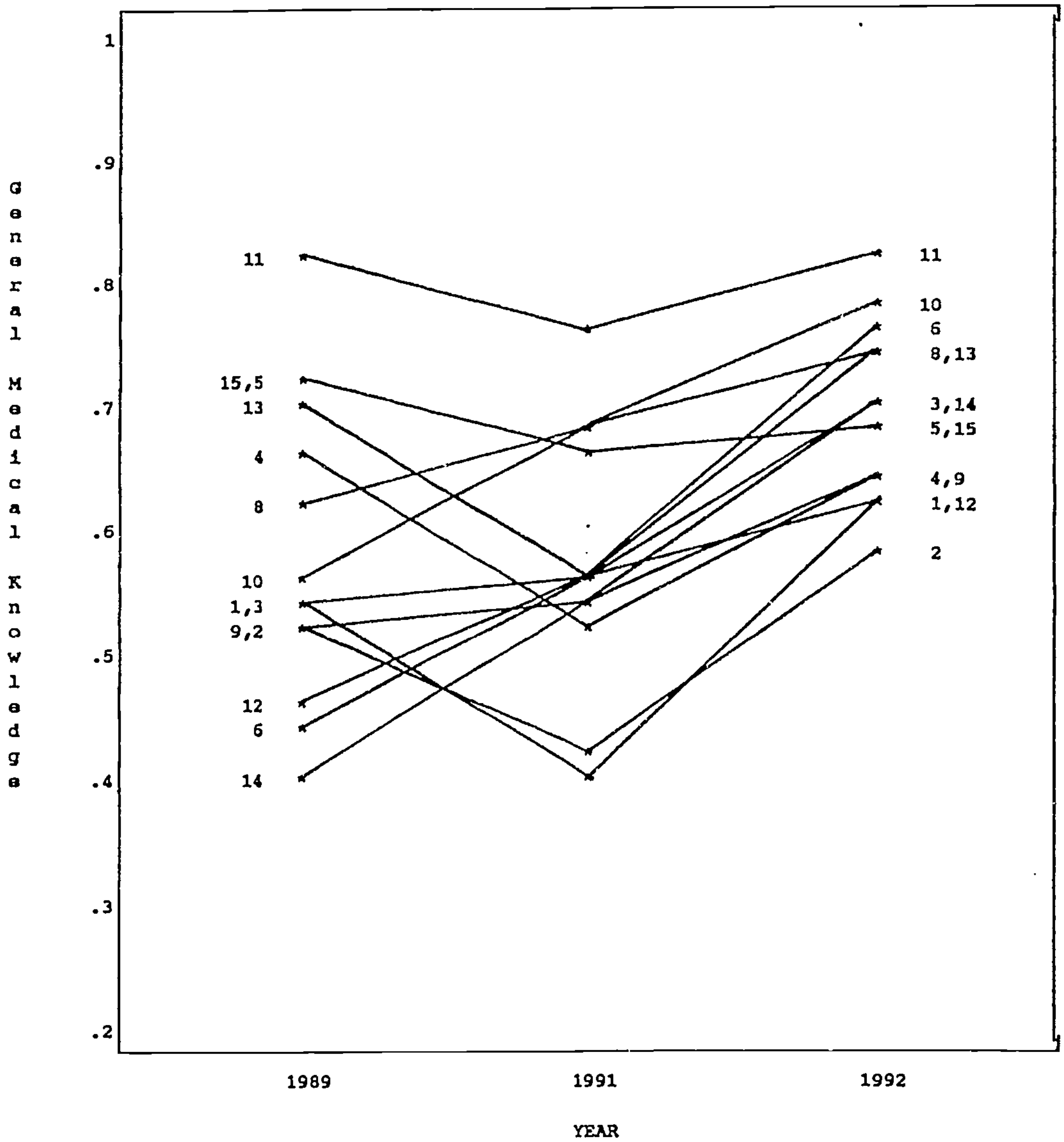

Figure 3. School mean MGK growth. Numbers at two sides of lines are school codes.

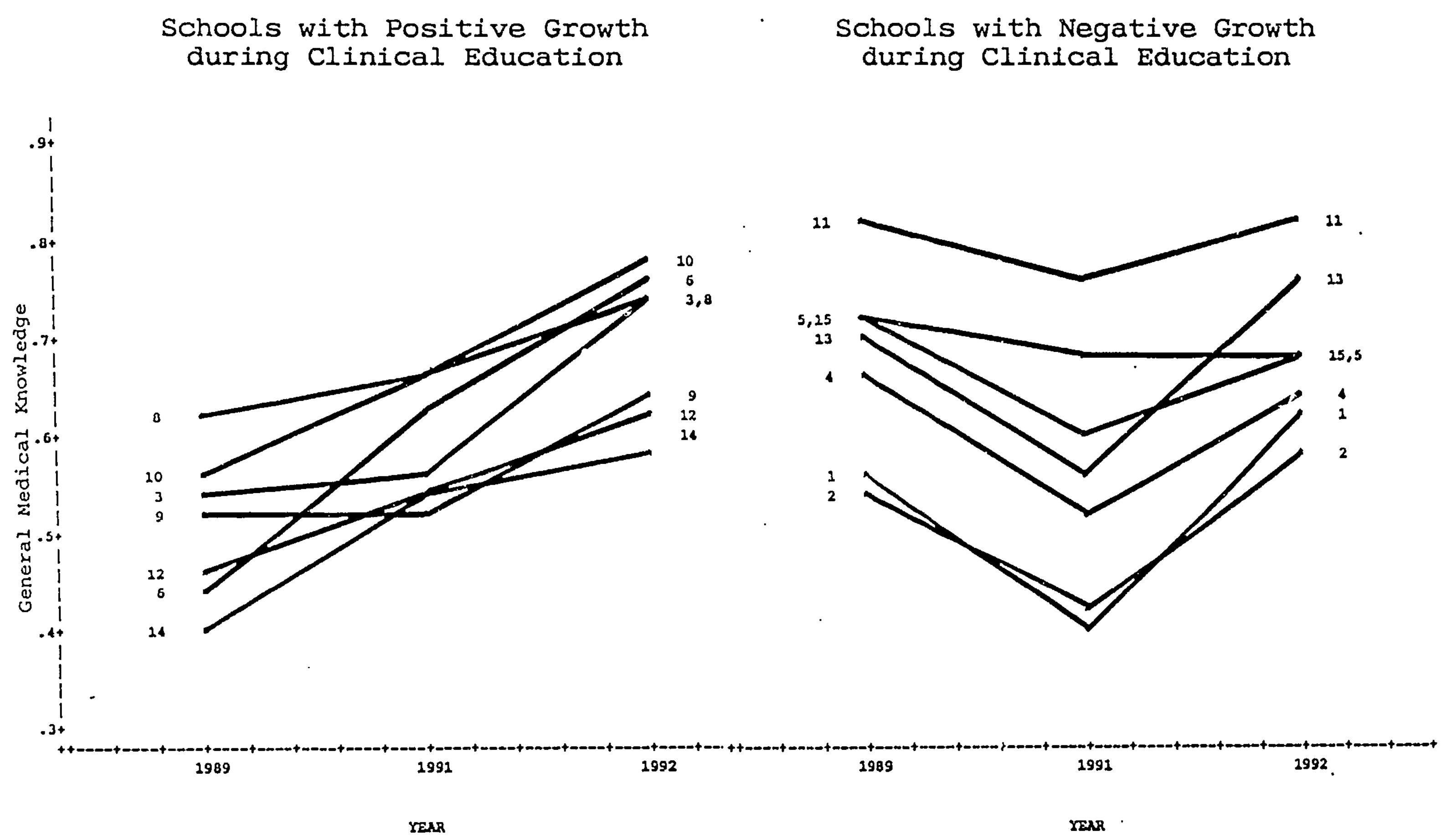

Figure 4. Two types of school mean GMK growth. Numbers at two sides of lines are school codes.

TABLE 1

SUMMARY OF UNCONDITIONAL MODEL

Fixed Effect	Coefficient	se	t Ratio	p value
γ_{100}	-.015	.030	-.499	.341
γ_{200}	.570	.025	23.261	.000
γ_{300}	.118	.014	8.667	.000

Random Effect	Variance Component	df	χ^2	p value
r_{1ij}	.046	1046	3019.470	.000
r_{2ij}	.100	1046	8824.687	.000
r_{3ij}	.019	1046	2115.095	.000
u_{10j}	.011	13	214.445	.000
u_{20j}	.007	13	81.337	.000
u_{30j}	.002	13	76.369	.000

	Percentage of Variance Between Schools
π_{1ij}	20.3
π_{2ij}	6.4
π_{3ij}	9.9

Random Coefficient	Reliability
π_{1ij}	.783
π_{2ij}	.939
π_{3ij}	.526
β_{10j}	.926
β_{20j}	.803
β_{30j}	.787

TABLE 1 (continued)

Correlations Among Random Effects

	π_{1ij}	π_{2ij}	π_{3ij}
π_{1ij}			
π_{2ij}	-.11		
π_{3ij}	.14	-.37	
	β_{10j}	β_{20j}	β_{30j}
β_{10j}			
β_{20j}	.52		
β_{30j}	-.28	-.81	

Deviance	df
15611.54	16

ERIC