

HETEROGENEIDADE ACADÊMICA DOS ALUNOS ADMITIDOS NA UFMG E PUC-MG

JOSÉ FRANCISCO SOARES¹
MARIA INÊS MARTINS²
CIBELE NORONHA BEHRENS ASSUNÇÃO³

1. Introdução

O Congresso Nacional estabeleceu através da lei 9.193 de 25/11/93 o Exame Nacional de Cursos. A forma como os primeiros resultados tem sido divulgados pela imprensa, e mesmo pelos relatórios oficiais do governo, sugere que, para cada curso, a média das notas dos alunos de uma instituição de ensino superior (IES) pode ser tomada como indicador de sua posição relativa no conjunto de todas as instituições.

¹ Ph.D. em Estatística pela University of Wisconsin-Madison, Professor Titular do Departamento de Estatística da UFMG. O autor registra e agradece o apoio da FAPEMIG, processo 71/93.

² Pro-Reitoria de Graduação da PUC-MG.

³ Bacharel em Economia, Mestranda em Estatística do Departamento de Estatística da UFMG.

As IES brasileiras tem, no entanto, perfis muito diferentes; são gratuitas ou pagas, confessionais ou laicas, situam-se nas regiões metropolitanas ou no interior dos Estados. Assim sendo, admitem alunos com perfis muito diferentes. Os alunos que se matriculam nas diferentes IFES não são homogêneos nem relação às suas habilidades acadêmicas, nem em relação à sua condição socio-econômica. Consequentemente, as notas médias das IES obtidas pelos respectivos alunos no Exame Nacional de Cursos não são imediatamente comparáveis e a classificação sugerida não se justifica.

Obter dados que permitam constatar, objetivamente, a heterogeneidade dos alunos à época de sua entrada na universidade é tarefa difícil. Não existe no Brasil um exame nacional ao fim do ensino médio e os resultados dos diferentes vestibulares não são imediatamente comparáveis. Neste artigo mostramos que os alunos admitidos no curso de Direito na Universidade Federal de Minas Gerais (UFMG) e na Pontifícia Universidade Católica (PUC-MG) em 1995 são diferentes tanto do ponto de vista de habilidade acadêmica quanto do ponto de vista sócio-econômico. Isso foi possível graças à flexibilidade da TRI - Teoria de Resposta ao Item -, do fato de que um grande número de alunos prestou vestibular nas duas universidades e que as respostas dos candidatos a todas as questões de todas as provas estavam preservadas em ambas universidades.

Este artigo pretende ainda contribuir para a importante atividade de criação e disseminação dos conhecimentos estatísticos necessários a uma avaliação dos cursos de Graduação por meio da medição das habilidades e conhecimentos adquiridos pelos seus formandos. Está organizado da seguinte forma: a seção 2 apresenta uma breve síntese da Teoria de Resposta ao Item. Na seção 3 calculam-se as habilidades dos alunos em cada uma das provas e apresentam-se gráficos e tabelas que comprovam a adequação do modelo subjacente à TRI em cada uma das provas. A seguir calculam-se e comparam-se as habilidades globais dos alunos que prestaram vestibular em ambas as instituições. Na seção 4 é apresentada uma pequena discussão sobre os impactos desta heterogeneidade no uso dos resultados do Exame Nacional de Cursos.

2. Habilidade Acadêmica dos Alunos

2.1 Teoria Clássica

Os testes educacionais, compostos de itens de múltipla escolha ou de resposta construída, medem a habilidade (ou aptidão) do aluno em uma determinada área de conhecimento. Durante longos anos a qualidade dos itens e as estimativas das habilidades dos alunos foram estudadas através do que se convencionou chamar de Teoria Clássica de análise de itens. Essa teoria utiliza, basicamente, estatísticas descritivas, coeficientes de correlação e proporções, para medir a qualidade dos itens. Usa os escores totais dos alunos nos testes como medida de aptidão. O texto clássico para esta abordagem é a segunda edição do *Educational Measurement*, uma coletânea de artigos de especialistas na área (Thorndike, 1971).

Embora ainda muito útil, a Teoria Clássica apresenta várias limitações. Conforme apresentado por Hambleton (1985) todas as suas medidas são dependentes das características dos alunos que fazem o teste. Os índices de dificuldade e discriminação usados para caracterizar a qualidade dos itens de um teste tem, portanto, uso restringido se os alunos usados no pré-teste não são representativos da população. Além disso, os escores dos alunos, usados para medir a aptidão, dependem do teste que está sendo aplicado. Ou seja, testes diferentes, com dificuldade e discriminação diferentes, produzem estimativas diferentes das habilidades. Essas e outras restrições foram resolvidas por uma nova teoria de medidas educacionais, a Teoria de Resposta ao Item (TRI).

2.2 Teoria da Resposta ao Item

A Teoria da Resposta ao Item trata a habilidade ("aptidão") dos alunos como uma *variável latente*, não diretamente observável e estabelece uma relação entre essa variável e a resposta do aluno. Na sua forma mais simples a TRI trabalha como se a resposta do aluno a cada item do teste fosse ou falsa ou verdadeira, e estabelece, através de uma equação matemática, relação entre a habilidade do aluno e sua

probabilidade de acertar o item. Vamos usar neste trabalho a seguinte especificação:

$$P_j(\theta_i) = \frac{e^{Da_j(\theta_i - b_j)}}{1 + e^{Da_j(\theta_i - b_j)}} \quad (2.1)$$

$P_j(\theta_i)$: probabilidade de o aluno i com habilidade θ_i acertar o item j ,

θ_i : habilidade do aluno i ,

b_j : índice de dificuldade do item j , $-\infty \leq b_j \leq \infty$,

a_j : índice de discriminação do item j , $a_j > 0$,

D : constante igual a 1,7 (usado para equiparar os valores da função logística com os da normal).

A função (2.1) é crescente, ou seja, com sua escolha assume-se que alunos com níveis mais altos de habilidade tem maior probabilidade de acertar o item. A despeito da existência de muitos outros modelos, considerou-se nesse trabalho um modelo no qual a probabilidade de se acertar o item depende apenas da sua *dificuldade* e do seu *poder de discriminação*, representados pelos parâmetros b_j e a_j , respectivamente.

Uma observação cuidadosa da expressão (2.1) mostra que o modelo utiliza um único θ_i , para descrever a aptidão de cada aluno. Esta característica é chamada de hipótese de unidimensionalidade e deve ser verificada quando da aplicação do modelo.

O parâmetro θ_i é característica do aluno e permanece o mesmo quando este aluno é submetido a diferentes testes. Naturalmente, existem outros fatores cognitivos ou de personalidade que afetam, de algum modo, a performance do aluno em um teste específico. Porém, para se usar a TRI, a habilidade que o teste está medindo, deve ser o fator dominante. Esta hipótese é equivalente ao que se convencionou chamar de independência local. Ou seja as respostas dos alunos para os itens em um teste são estatisticamente independentes dado o conhecimento da habilidade desses alunos.

Existem muitos métodos para a estimação dos parâmetros do modelo, discutidos por vários autores como Mislavy (1986), Swaminathan e Gifford (1985, 1986, 1990), Patz (1997). Nesse trabalho usamos o método de máxima verossimilhança marginal para

estimar os parâmetros dos itens e o método bayesiano para estimar as habilidades (Bock e Aitkin, 1981), como implementados no programa computacional BILOG (Bock e Mislevy, 1984). Apenas as opções padrão do programa foram utilizadas para obter as estimativas.

3. Comparação PUC – UFMG

3.1 Dados

Para comparar estas duas universidades, usamos os dados obtidos dos candidatos ao vestibular de 1995 para o curso de Direito da PUC-MG e da UFMG. Excluídos os candidatos que não fizeram todas as provas restaram 5991 candidatos na UFMG, 2908 na PUC-MG. Dentre esses 1774 fizeram as provas em ambas as universidades.

O vestibular nas duas universidades consiste, na primeira etapa, de itens de múltipla escolha e abrangem basicamente o mesmo programa. Neste trabalho usamos os resultados das provas de: Português, Matemática, História, Geografia, Física, Química, Biologia e Língua Estrangeira (Francês ou Inglês).

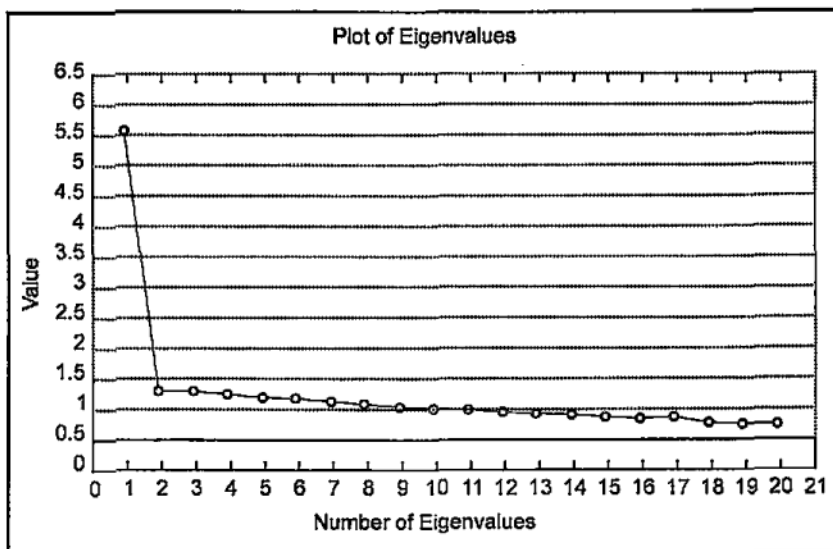
3.2 Estimativa das Habilidades

Como dito anteriormente, para usar modelos da TRI na estimação das habilidades dos alunos, é preciso verificar se a hipótese da unidimensionalidade é satisfeita. Usamos para isto a metodologia proposta por Reckase(1979), segundo a qual se o primeiro autovalor da matriz de correlação tetracórica é dominante, apenas uma habilidade é medida pelo teste.

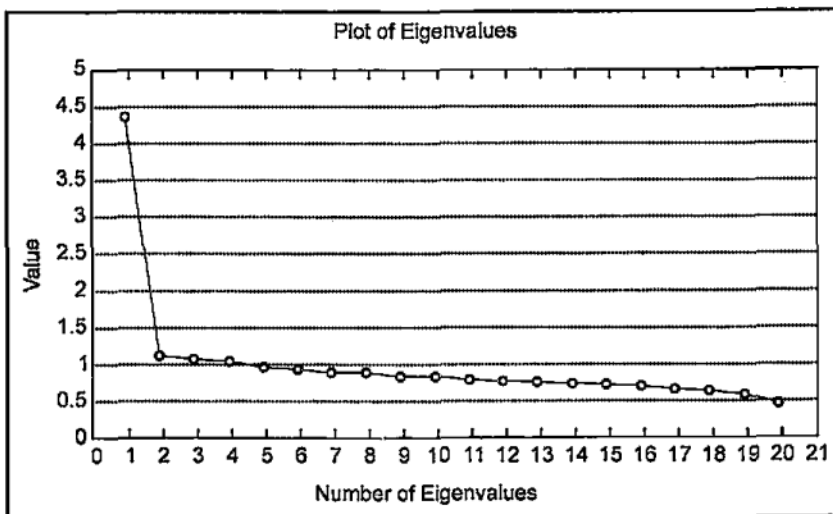
A Figura 1 mostra, para as provas de História e Português de cada uma das universidades, que há uma clara dominância do primeiro fator. Esta é a indicação empírica de que as provas dos dois vestibulares medem cada uma apenas uma habilidade. Esse mesmo comportamento foi verificado para as demais provas.

FIGURA 1
Autovalores da Matriz de Correlação Tetracórica entre os Itens

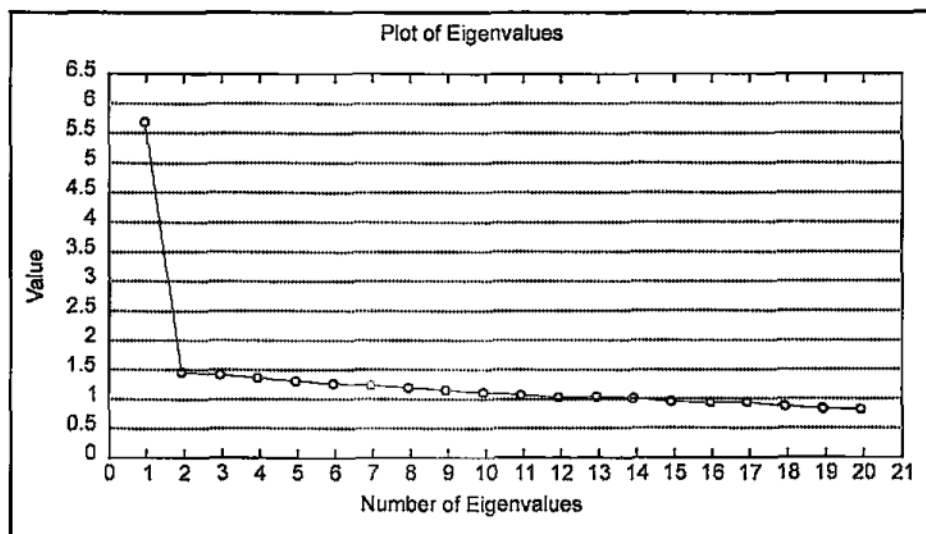
História PUC-MG (1995)



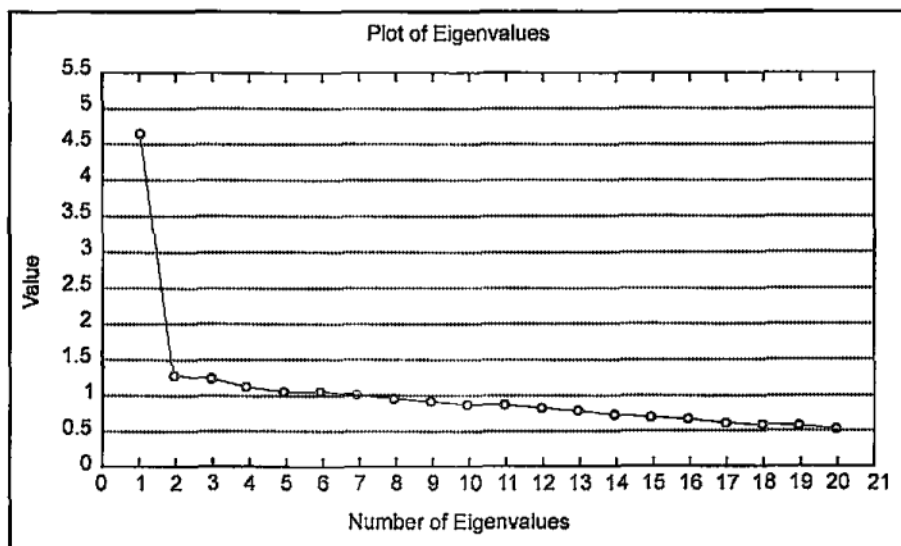
História UFMG (1995)



Português PUC-MG (1995)



Português UFMG (1995)



Assumindo-se a unidimensionalidade do traço latente em cada prova, as respectivas aptidões θ_i dos alunos de cada instituição foram

calculadas, usando-se o programa BILOG. Em todas as estimações a média das habilidades de todos os alunos envolvidos foi fixada em 0 e desvio padrão em 1. Para cada aluno foram estimadas oito habilidades, uma para cada prova à qual foi submetido.

Os 1774 alunos que fizeram o vestibular em ambas as universidades produzem informações importantes que permitem colocar as habilidades de todos os alunos em uma mesma escala. Para isso, durante o processo de estimação das habilidades, considerou-se uma configuração da matriz de resposta como se todos os alunos tivessem se submetido a um único teste, constituído das 20 questões da UFMG e das 30 da PUC. Assim, para os alunos que se submeteram apenas a um dos vestibulares foi usada a opção "Not-Presented" do programa, para retirar as questões não respondidas da verossimilhança do modelo. Isso permitiu estimar todas as habilidades de todos os candidatos que se submeteram ao vestibular da PUC-MG ou da UFMG de uma só vez e usando-se todos os dados.

Como o objetivo desse trabalho é comparar as instituições através da habilidade de seus alunos, calculou-se uma média ponderada das habilidades em cada uma das prova de modo a obter uma única habilidade que será chamada de *habilidade acadêmica* do aluno. Os pesos utilizados para o cálculo dessas habilidades dos alunos em cada universidade são os usados pela PUC-MG no curso de Direito, isto é, Português 4; Matemática 2; História 4; Geografia 2; Física 1; Química 1; Biologia 1 e Língua Estrangeira 1.

3.3 Comparação das Habilidades

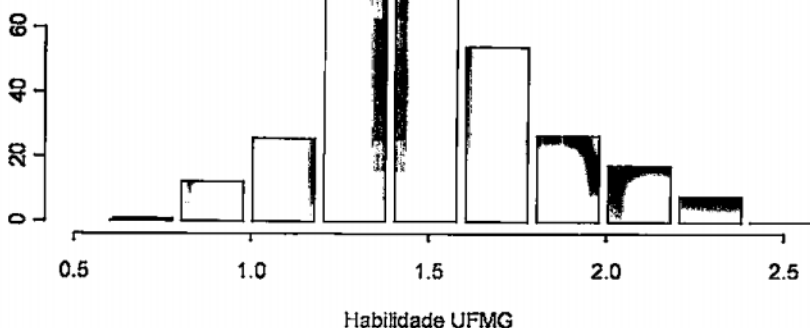
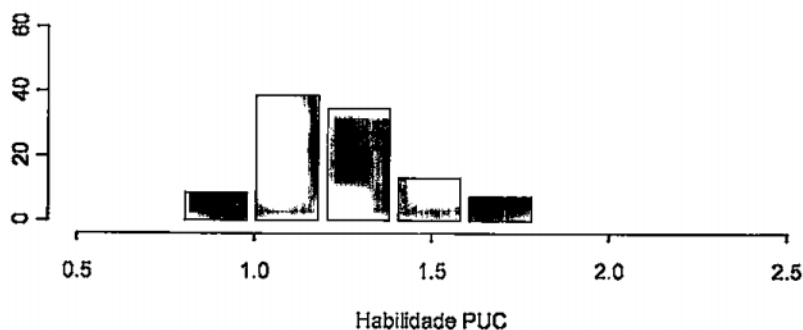
A comparação entre as duas instituições PUC-MG e UFMG foi feita usando-se as habilidades acadêmicas dos alunos que efetivamente se matricularam no curso de Direito em 1995 em cada uma das Universidades. Na UFMG matricularam-se, na primeira chamada, 306 alunos e na PUC-MG, 111.

A distribuição das *habilidade acadêmica* desses alunos é mostrada na Figura 2. Observa-se que a média das habilidades dos alunos matriculados na UFMG é um maior que a média das habilidades dos alunos matriculados na PUC-MG ($\mu_{UFMG} = 1.50$ e $\mu_{PUC} = 1.23$). Deve-se observar que as habilidades dos alunos da

UFMG estão, claramente, concentradas em valores maiores que os referentes à PUC. Essa diferença constatada é estatisticamente significativa, ao nível de 95% de confiança, usando-se o teste *t* e não-paramétrico de diferenças de médias.

Figura 2

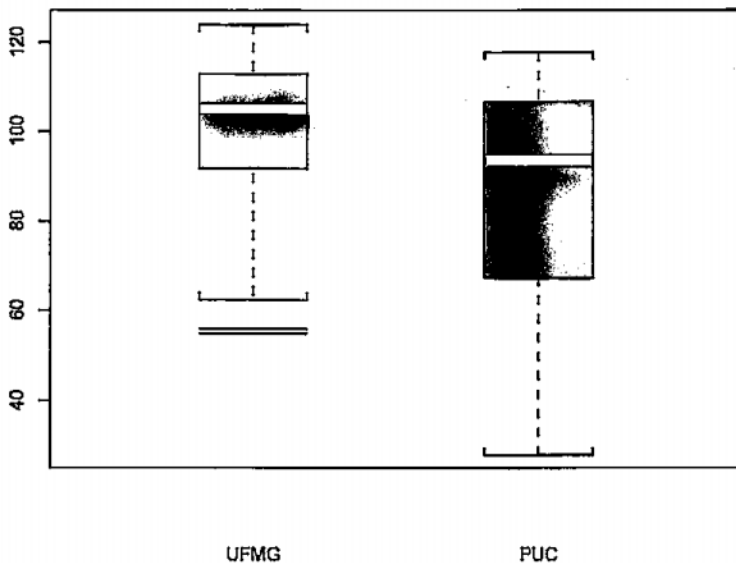
Habilidade dos Alunos Matriculados no Curso de Direito



3.4 Comparação Sócio-Econômica

Dentre os 1774 alunos que fizeram o vestibular de ambas as Universidades, 121 matricularam-se na UFMG e 74 na PUC. Estes alunos permitem uma comparação da situação socio-econômica dos alunos admitidos nas duas Universidades. Isto porque os alunos que fizeram o vestibular na UFMG preenchem ao fazer sua inscrição, um questionário onde estão incluídas várias perguntas sobre sua situação socio-econômica. Estes dados foram transformados em um índice de posição social à semelhança do índice da ABA e ABIPEME através de metodologia descrita em Soares e Fonseca (1997). O resultado está apresentado na Figura 3. Observa-se que os alunos matriculados na UFMG tem índice sócio-econômico superior aos matriculados na PUC.

Figura 3
Índice Sócio-Econômico dos Alunos Comuns



Para estes alunos a habilidade média dos matriculados na PUC é de 1.20 e para a UFMG de 1.65, ou seja, também para este grupo mais restrito a diferença na habilidade acadêmica fica constatada.

Conclusão

A comprovação empírica da heterogeneidade acadêmica e econômica entre os alunos que se matriculam na UFMG e PUC é o ponto central desse trabalho. Diante desse fato a comparação destas instituições só pode ser feita considerando-se as diferenças constatadas. é razoável imaginar que, caso existam dados, a mesma heterogeneidade aqui constatada seria verificada em muitas outras instituições. Assim sendo, a avaliação das IES realmente relevante é aquela que mede o quanto cada instituição acrescenta ao conhecimento de seu aluno e não o conhecimento final do aluno. Existe uma crescente literatura utilizando essa abordagem, seja na direção metodológica, Goldstein(1997), seja na direção de estudos de casos (Thomas,1997).

Existem dados para implementação prática dessa idéia. São os dados do Exame Nacional de Cursos (Provão), que, infelizmente, ainda não estão disponíveis nem para pesquisas. Esses dados junto com os dados do Vestibular das Universidades Públicas, que por serem gratuitas tem um grande número de candidatos, permitiriam comparar as várias instituições de ensino superior do país de forma mais confiável.

No presente trabalho utilizou-se somente o modelo logístico de dois parâmetros. O modelo de três parâmetros, escolha usual para análise de dados de testes na literatura, pressupõe a existência de testes de melhor qualidade, nem sempre possível de se obter nos exames vestibulares onde os itens utilizados não são pré-testados.

Referências Bibliográficas

- BOCK, R. Darrell; AITKIN, Murray. Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, 46, 443-45, 1981.

- BOCK, R. Darrell; MISLEVY, Robert J. BILOG. **Item Analysis and Test Scoring with Binary Logistic Models**. Mooresville (IN): Scientific Software, 1989.
- GIFFORD, J. A.; Swaminathan, H. Bias and the effect of priors in Bayesian estimations of parameters of the item response models. **Applied Psychological Measurement**, 15, 375-389, 1990.
- GOLDSTEIN, H. **Multilevel Statistical Models** (2nd. Edition) London: Arnold, 1995.
- GUILFORD, J. P. **Fundamental Statistics en Psychology and Education**. 4 ed. New York: McGraw-Hill, 1965.
- HAMBLETON, R. K. Principles and Selected Applications of Item Response Theory. In: Linn, R. L. (Ed.) **Educational Measurement**. 3 ed. New York: McMillan, p.147-200, 1989.
- MISLEVY, R. J. Bayes modal estimation in item response models. **Psychometrika**, 51, 177-195, 1986.
- PATZ, R. J.; JUNKER, B. W. **A straightforward approach to Marcov Chain Monte Carlo methods for item response models**. Technical Reports of Statistics Department. Carnegie Mellon University, 1997.
- RECKASE, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. **Journal of Educational Statistics**, 4, 207-230, 1979.
- SOARES, José F.; Fonseca, José A. **Fatores Sócio-Econômicos e Desempenho no Vestibular UFMG-97**. Relatório Técnico, 1997.
- SWAMINATHAN, H; GIFFORD, J. A. Bayesian estimation in the three-parameter logistic model. **Psychometrika**, 51, 589-601, 1986.
- _____. Bayesian estimation in the two-parameter logistic model. **Psychometrika**, 50, 349-364, 1985.
- THOMAS, S.; Mortimore, P. Comparison of value-added models for secondary school effectiveness. **Research Papers in Education**, 11(1) 5-33, 1996.