

A transmissão da tecnologia da avaliação em larga escala no Brasil

Apresentação

Ao longo de sua implementação, no ano de 1995, a avaliação em larga escala no Brasil nos moldes da Teoria da Resposta ao Item – TRI vem sofrendo alterações de modo a se ajustar às novas realidades do sistema educacional. Trata-se de um processo contínuo de retroalimentação. Ao mesmo tempo em que os resultados das avaliações fornecem diagnósticos, estes mesmos resultados estabelecem novos desafios com o objetivo de atender aos anseios da sociedade.

Podemos verificar ao longo destes vinte anos, conforme observado por Bonamino (2012), três gerações de avaliações educacionais, a primeira geração se caracterizando pela preocupação do Ministério da Educação em mensurar a qualidade da educação e as segunda e terceira gerações por políticas de responsabilização atreladas aos resultados das avaliações e estipuladas por Secretarias Estaduais de Educação.

Apresentaremos neste trabalho este percurso histórico da avaliação através de seus modelos matemáticos e suas possibilidades na tentativa de modelar a complexa realidade dos sistemas educacionais, e como estes métodos quantitativos, em uma abordagem sociológica, mais estritamente relacionada às teorias de Weber, podem ajudar a entender esta evolução ou adequação histórica da avaliação em nosso país.

Finalizando esta reflexão, apresentamos as novas perspectivas oriundas dos trabalhos de eficácia escolar e as tendências de estudos envolvendo análises quantitativas e qualitativas no intuito de melhor compreender e dar subsídios para a melhoria da qualidade e equidade da educação. Para tanto apresentamos um estudo de caso de análise de eficácia escolar desenvolvido em um estado brasileiro.

Palavras-chave: avaliação educacional; eficácia escolar; Teoria da Resposta ao Item; modelagem multinível

A avaliação educacional em larga escala no contexto brasileiro

O processo de avaliação da educação brasileira, com a utilização de modelos matemáticos da Teoria da Resposta ao Item – TRI, implantado em 1995, veio sofrendo alterações estruturais e conceituais ao longo dos anos. Inicialmente concebido para fazer um diagnóstico do sistema educacional do país e das unidades da federação, utilizando avaliações amostrais e bianuais, sua primeira grande mudança estrutural ocorreu no ano de 2005 com a implantação de avaliações censitárias para os 5º e 9º anos do Ensino Fundamental e a criação do Índice de Desenvolvimento da Educação Básica – IDEB.

Ao se construir um índice de qualidade e estipular metas de desempenho para todas as escolas públicas com base nesse índice, utilizando como variáveis o fluxo escolar e as proficiências dos alunos, observou-se uma grande movimentação ou mesmo inquietação por parte dos sistemas de ensino, no sentido de definir estratégias e políticas educacionais objetivando a melhoria dos resultados do IDEB de suas escolas. Ao que parece, ao se criar o IDEB, a consequência, mesmo que não consciente, foi criar um problema, tendo o governo deixado para os sistemas educacionais a responsabilidade de acharem suas próprias soluções.

Uma dessas soluções foi a implantação de sistemas próprios de avaliação por parte dos estados e municípios, para, com isso, se apropriarem melhor dessa tecnologia. Tem-se atualmente a quase totalidade dos estados brasileiros com seus sistemas próprios de avaliação e o início, ainda tímido, de alguns municípios nesse processo, a saber: Belo Horizonte, Salvador, Fortaleza, Rio de Janeiro, São Paulo, Curitiba, para citar os maiores, assim como pequenos municípios como Ipojuca, em Pernambuco, e Ouro Branco, em Minas Gerais. Nesse cenário, observa-se que muitas escolas chegam a realizar, geralmente no final do ano, até quatro avaliações externas: uma avaliação nacional, uma estadual, uma municipal e uma avaliação internacional, ou seja, há um excesso de avaliações externas, o que parece demonstrar uma certa desorientação de todo o processo avaliativo no país.

Este cenário traz à tona a alusão feita por Guiddens, em seu livro, *As consequências da modernidade*, em que, para definir a complexa dinâmica do mundo moderno, ele descreve uma passagem da mitologia hindu em que Jaganatha, um dos avatares de Vishnu, o deus responsável pela manutenção do universo, anda com seu carro de forma desgovernada pelo universo. Ninguém sabe para onde seu carro vai, mas todos querem entrar, afinal, deus está dentro do carro. Nessas tentativas, muitos não conseguem entrar no carro, sendo atropelados e mortos.

Ao longo desses vinte anos de avaliação educacional no Brasil, o carro de Jaganatha, ou seja, a avaliação em larga escala, não está totalmente desgovernado, ainda que com algumas inconsistências e excessos, a meta é clara: promover a qualidade e a equidade do ensino.

A forma como nós estamos nos apropriando desta tecnologia é ímpar. Temos bases de dados e exemplos de políticas públicas educacionais baseadas nas avaliações que nos

permitem traçar nossos próprios rumos e olhar com cautela estudos internacionais. Cada vez mais, a avaliação vem incorporando mais atores e se consolidando como uma ferramenta para a promoção da qualidade de ensino.

Neste sentido, valendo-nos das características matemáticas dos modelos da TRI: comparabilidade de resultados e interpretação pedagógica, procuraremos dar uma interpretação sociológica às avaliações educacionais em larga escala. Dentre as inúmeras possibilidades possíveis de análises, focaremos em apenas duas, por considerarmos que estão mais relacionadas ao propósito da melhoria da qualidade do ensino:

1. Modelos simplistas da TRI: investigando como atores específicos, nas escolas e nas redes de ensino, usam informações qualificadas para a tomada de decisões (teoria da ação).
2. Modelos complexos da TRI: como grupos sociais (raça, classe, sexo) obtêm chances distintas, em termos de desempenho escolar (e as consequências que isso gera em termos de mercado de trabalho, por exemplo), levando em consideração, justamente, suas características sociais (como grupos) e aspectos cognitivos.

Modelos simplistas da TRI.

Podemos observar que, desde a sua origem até os dias atuais, no Brasil, os modelos da TRI, conforme descrito por Goldstein (2001), limitam-se, quase que na sua totalidade, a uma tentativa simplista de modelar uma realidade complexa, que é o sistema educacional. Evidentemente, o que o autor se refere como simplista não é a formulação matemática do modelo da TRI, mas sim, utilizar apenas medidas de desempenho, definidas como proficiência (o conhecimento do aluno) para a tomada de decisões e mesmo para conhecer e explicar as diferenças entre entes avaliados: alunos, escolas, municípios e etc. Por realidade complexa do sistema educacional o autor se refere às características socioeconômicas dos alunos, o tipo de gestão da escola, sua infraestrutura, localização – se está, por exemplo, em uma região violenta. Esses fatores dentre tantos outros, deveriam ser considerados ao se proceder a comparabilidade de resultados, o que será melhor abordado na próxima seção.

Não discordando de Goldstein, mas apenas direcionando a avaliação para um contexto brasileiro, podemos destacar que, principalmente a partir de 2005, com a criação do IDEB, a utilização de informação qualificada para a tomada de decisões. O que se enquadra nas formulações de Weber (2009) e dos teóricos da ação na tentativa de obter evidências para apoiar o racional.

A ideia de se modelar uma determinada realidade é exatamente torná-la simples, isto não é uma falha e sim uma característica, é claro que nesse processo, muitas informações são perdidas. Para tornar mais clara essa ideia, vamos considerar o

desempenho de duas escolas, através das análises de suas proficiências: a escola X com 100 alunos e uma média de proficiência de 200 e uma escola Y com 180 alunos e uma média de proficiência de 150, por exemplo. Não temos como determinar o motivo pelo qual elas se diferem nem justificar nas diferenças de proficiências, todas as possíveis características intra e extraescolares dessas escolas. Mas um fato não muda: a escola X teve um desempenho superior. Dizendo de outra forma: levando-se em consideração que essas medidas estão baseadas em testes concebidos a partir de uma matriz de avaliação que, normalmente, abrange em torno de 30 descritores para uma determinada disciplina de uma determinada série e também que as medidas de proficiências por aluno possuem um erro de aproximadamente 10%, podemos concluir que a probabilidade da escola X ser melhor que a escola Y é muito alta.

A TRI é uma metodologia muito robusta para fornecer informações sobre sistemas educacionais. Quanto maior o número de indivíduos nas análises, mais precisas são as comparações. Comparar os resultados desagregados para estado, municípios e escolas fornece boas análises, possibilitando gestores educacionais a terem um bom mapeamento sobre a qualidade de suas redes de ensino, poderem tomar decisões e definirem políticas educacionais baseados em dados concretos, em sintonia com os preceitos da teoria da ação de Weber.

Um outro importante fator na comparabilidade dos resultados é o fato das avaliações estaduais e municipais estarem, praticamente todas em uma mesma escala que também é a utilizada pelas avaliações nacionais realizadas pelo INEP. Essa é sem dúvida a grande diferença das avaliações brasileiras em relação às avaliações internacionais. Mesmo nos EUA, que possuem uma cultura de avaliação bem mais antiga do que a nossa, cada estado americano tem seu próprio sistema de avaliação e uma escala que não se comunica com a escala nacional. Embora já tenham ocorrido esforços no sentido de se criar uma escala única, como nos moldes brasileiros, estas tentativas não foram bem sucedidas.

Temos um quadro nacional sobre a avaliação, nesta perspectiva de se atingir metas geradas e projetadas através de uma mesma metodologia, que se encaixa numa abordagem funcionalista com a supervalorização do papel da tecnologia, subestimando a importância do conflito e da ideologia, concebendo o sistema educacional a partir de uma perspectiva de consenso e equilíbrio. O que também se encaixa no pensamento de Weber ao acreditar que as práticas do capitalismo eram um sistema social mais apropriado para promover a liberdade humana e a produtividade econômica.

Mas temos também posicionamentos que defendem que tais políticas acentuariam ainda mais as discrepâncias entre os mais favorecidos em detrimento aos menos favorecidos. Nesse sentido, a avaliação é algo que fortalece e dá acalento aos envolvidos no sistema educacional, pois gestores e governos, submetendo os alunos a uma prática que além de servir aos seus propósitos, imputa a estudantes estes mesmos sentimentos, conforme Bourdieu (1975), é o sistema reproduzindo seus hábitos.

Poder comparar é a principal característica das avaliações pela TRI possibilitando a aplicação de políticas de *accountability* a partir da criação de índices e metas de desempenho. Temos diversos exemplos desta prática nos sistemas educacionais, dentre eles destacamos: o Ceará, onde a distribuição de ICMS para os municípios está atrelada aos índices de proficiências e fluxos de seus alunos; o Rio de Janeiro, onde já houve premiação de professores e alunos. Além desses dois exemplos, outras diversas modalidades de bonificação podem ser encontradas nos diversos estados. No entanto, normalmente essas políticas envolvem apenas bonificações e nunca sanções como demissão de professores e fechamento de escolas com baixos índices, como acontece nos EUA.

Temos, nessa utilização da avaliação, talvez escondido sob o manto da qualidade, uma visão mercadológica da avaliação remete-nos a Halsey(1977) que afirma que nas sociedades capitalistas industriais é inevitável que o sistema educacional tenha uma relação direta com a economia. Nessa perspectiva, podemos considerar que o IDEB e as políticas apoiadas em seus resultados se enquadram totalmente neste pensamento.

Uma visão bem diferente dos usos da avaliação é observada na França, onde avaliações, geralmente amostrais, fornecem um diagnóstico para que o ministério da educação desse país possa verificar a qualidade da educação e definir políticas e práticas pedagógicas com o intuito de melhorar a qualidade do ensino. Práticas como esta são o grande diferencial das avaliações realizadas pelo INEP com relação às avaliações realizadas pelos estados e municípios com seus sistemas próprios de avaliação, visto que, nestes últimos, é dada uma ênfase na interpretação da escala e no aspecto pedagógico da avaliação.

Através dessa prática, a intenção é tentar penetrar no ponto de vista subjetivo do professor, torná-lo objetivo, ou seja, objetivar a subjetividade, para perceber a forma como ele ou ela vê o mundo, de modo a compreender suas motivações, como proposto por Weber (*apud* Collins, 2009) e a busca pelo *verstehen* (compreender).

Como tornar objetivo e mensurável a relação professor-aluno através da modelagem estatística da TRI? A resposta a esta pergunta é sem dúvida a principal contribuição da psicometria para tentar entender como se dá a interação professor/aluno. A perspectiva, portanto, é a utilização da TRI como instrumento sociológico para a compreensão das diversas nuances dos sistemas de ensino. Portanto, inicialmente temos que entender como a TRI tenta modelar a relação professor-aluno para em seguida fazermos uma abordagem dos diferentes perfis de professores com relação à aplicação desta metodologia em suas práticas docentes.

Imaginemos, nessa tentativa de explicar a modelagem, três salas de aulas em que professores diferentes lecionam, para alunos diferentes a mesma disciplina em um mesmo período de escolaridade. É muito provável que estes professores tenham sentimentos bem diferentes em relação a seus alunos. Suponhamos que nestas três salas de aulas os alunos possuem o mesmo nível de conhecimento em relação à disciplina.

Um professor pode ter um sentimento mais severo e achar que a turma é fraca, um outro professor, mais leniente, pode achar a turma boa e um outro professor pode achar que a turma é mediana, ou seja, nem boa nem ruim. Uma outra consideração importante é que embora estes professores lecionem a mesma disciplina para uma mesma série, eles não conseguem dialogar entre si para comparar os desempenhos de seus alunos, eles não possuem uma medida comum. Para complicar ainda mais esta relação professor/aluno, muito provavelmente as turmas possuem desempenhos diferentes o que inviabiliza qualquer comparabilidade, tornando a prática docente algo impossível de mensuração no sentido de definir padrões de qualidade, deixando os professores isolados em seus sentimentos e percepções.

Ao se produzir medidas pela TRI, é possível colocar em uma mesma métrica (escala), o desempenho de diferentes turmas e alunos. Assim, é possível diferentes professores, com percepções e sentimentos diferentes, comparem diferentes alunos em diferentes contextos educacionais.

Nesse ponto, ocorre a quebra de um grande paradigma do sistema educacional: o sentimento dos professores em relação a seus alunos pode ser mensurável, os professores não estão mais isolados com seus alunos em uma sala de aula. É possível o diálogo com outros professores e estabelecer padrões de comparações.

No entanto, devemos recordar que a TRI é uma modelagem e, como tal, não consegue representar toda uma realidade, mas seus resultados, dentro de uma perspectiva um pouco limitada, ou reducionista (termo normalmente usado pelas pessoas contrárias às avaliações) de tudo o que se é ensinado na escola, ganha força pelo seu poder de comparabilidade, permitindo assim novas possibilidades de análises para as principais preocupações abordadas pela Nova sociologia da educação, quais sejam: A interação professor aluno, as categorias ou conceitos usados pelos educadores e o currículo.

Fazendo um paralelo com as ideias ou ‘conceitos’ teóricos de Bernstein (Collins, 2009) podemos ver na avaliação pela TRI as funções de classificação e enquadramento: onde a classificação está preocupada com a organização social do conhecimento; seu foco é a matriz de referência e não o currículo como sugerido por Bernstein (Collins, 2009); e o enquadramento é sobre a relação professor-aluno e sua ênfase na interação com todas as salas de aulas (já que estão todas na mesma métrica) e não apenas dentro da sala de aula conforme sugerido por este mesmo autor.

Temos, portanto, através da modelagem, numa perspectiva da Nova Sociologia do Ensino, um reducionismo em relação ao currículo e uma grande mudança de paradigma ao pensarmos na interação do professor, não apenas com aluno, mas com todo o sistema de ensino.

A interação professor-aluno, através da TRI, traduz-se em interpretar pedagogicamente determinadas categorias da escala de proficiência. Normalmente essas escalas são divididas em quatro padrões de desempenho, permitindo o enquadramento dos alunos em função de seus conhecimentos que podem ser categorizados como: fraco, moderado,

bom e ótimo. Cabe ao professor identificar quais habilidades, traduzidas por descritores, seus alunos dominam ou têm dificuldades. Nessa perspectiva, a avaliação torna-se uma poderosa ferramenta na promoção da qualidade e equidade nas escolas.

No entanto, isso é algo muito difícil de se implementar, pois para sua efetivação, temos diferentes sujeitos envolvidos na avaliação: alunos, professores, gestores, sindicatos, instituições avaliadoras, secretarias de ensino em uma constante situação de conflito segundo a ótica weberiana, tendo “o conflito como a expressão do caráter multidimensional das coisas, da pluralidade dos diferentes grupos, interesses e perspectivas que constituem o mundo” (COLLINS, 2009, p. 80).

Podemos observar que, atualmente no Brasil, as resistências por parte dos sindicatos em relação às avaliações, que era muito grande no início das avaliações na década de 1990, praticamente não existem mais. Temos ainda um pouco de resistência em alguns estados no Sul do país, mas, podemos considerar que o país já está totalmente imerso em uma cultura de avaliação externa e sua forma capitalista de tratar os sistemas de ensino.

Outros grandes desafios em todo este processo não estão centrados na produção das medidas e no domínio das técnicas estatísticas da TRI, mas sim na definição de um matriz de referência que faça sentido para o sistema educacional e fazer com que os professores dominem e entendam os fundamentos desta metodologia tornando-a uma aliada em suas práticas pedagógicas. Com relação à definição das matrizes de referências, temos novamente uma situação de conflito entre pedagogos e gestores educacionais em definirem o que se deve avaliar e o que deve ser considerado como prioritário no aprendizado do estudante em determinado período de escolaridade. Como tomar estas decisões de forma participativa é o grande desafio a ser vencido. Estes mesmos questionamentos podem ser observados com relação à implantação de um currículo único nacional, e a sua solução passa naturalmente por negociações e lutas de poder inerentes aos processos sociais no sistema capitalista.

O segundo desafio apresentado não é menos complexo que o primeiro, fazer com que os professores utilizem os resultados das avaliações em suas atividades diárias, como uma ferramenta para a promoção da qualidade e equidade em sala de aula, envolve mais que dominar os princípios metodológicos da TRI e algumas técnicas da estatística descritiva para a interpretação pedagógica da escala, mas sim uma mudança comportamental e, mudar comportamento é algo muito difícil. Cada professor tem a sua concepção de como ensinar aos estudantes, e, naturalmente, usar a TRI, que é algo novo no sistema educacional, gera muito atrito. Em um primeiro contato com esta metodologia, observamos uma grande reação quando são apresentados aos professores técnicas estatísticas para lidarem com a aprendizagem: Gráficos, tabelas e demais conceitos matemáticos que estes professores aprenderam no 2º grau são verdadeiros empecilhos para que eles se apropriem desta tecnologia.

Utilizar uma avaliação de larga escala para fins pedagógicos é realmente o grande diferencial do Brasil em relação aos demais países. Podemos observar diversas estratégias por parte dos sistemas de ensino e instituições avaliadoras para fazer com

que esta metodologia seja utilizada pelos professores, como por exemplo: oficinas de divulgação e apropriação dos resultados com a participação de todos os professores e diretores de escolas, nas quais são passados os conceitos da TRI, como analisar os dados estatisticamente através de tabelas e gráficos e como os resultados devem ser interpretados pedagogicamente; oficinas de elaboração de itens, com a finalidade de fazer com que os professores dominem os conteúdos das matrizes de avaliação, o alinhamento dessas matrizes com o currículo e, conseqüentemente, com o ensino, além da forma como os itens são elaborados para que possam medir as habilidades dos estudantes; revistas e sites para facilitar o acesso às informações e cursos de transferência de tecnologia, habilitando os técnicos das secretarias a lidarem com bases de dados e *softwares* estatísticos para a produção de medidas.

As reações dos professores a esse processo são as mais variadas possíveis, podendo-se identificar diferentes perfis de professores em função de suas afinidades com os usos da avaliação ou numa visão sociológica, segundo a ótica de Weber (Collins, 2009) diferentes categorias de tipos ideais interagindo ao mesmo tempo de forma construtiva, neutra e até mesmo destrutiva.

Novamente, nos referindo a Weber (Collins, 2009) podemos nos valer do seu conceito de vocação, presente no trabalho "*A ética protestante e o espírito do capitalismo*" para uma forma de interpretação de como os professores podem ter diferentes apropriações dos resultados. Normalmente, a forma construtiva é encontrada naqueles grupos que possuem uma certa afinidade com a matemática e facilidade de entender conceitos estatísticos e as formas neutra e destrutiva para o grupo oposto e/ou que estão desmotivados com suas profissões e desencantados com o sistema de ensino.

Não é possível utilizar os resultados da avaliação, através de uma análise simplista da TRI, sem que os professores saibam estatística descritiva. No mundo empresarial esta estratégia para se atingir a qualidade se remonta ao período pós-guerra com a utilização por parte do Japão das metodologias e ferramentas estatísticas de Deming e Juran para se atingir a qualidade total. No Brasil, estas técnicas, nas indústrias e empresas prestadoras de serviços, se incentivaram a partir da década de 1990 com o surgimento das normas ISO série 9000.

Esta mobilização em busca da qualidade, embora conste na nossa Constituição desde 1988, que é dever do Estado garantir uma educação de qualidade a todo cidadão brasileiro é ainda uma realidade distante e, levando-se em consideração que não se tem como pensar em qualidade sem que a mesma seja medida, no universo escolar, o principal, senão o único método quantitativo é a TRI, daí a importância da sua apropriação por parte de todo o sistema educacional.

Modelos complexos da TRI.

Os modelos matemáticos da TRI são o estado da arte da psicometria para a construção de um determinado traço latente, que, no caso da educação, a medida mais utilizada é a proficiência, ou conhecimento, do aluno. Em todos os países, é utilizada esta mesma metodologia. Porém os modelos matemáticos variam em função dos tipos de itens utilizados (dicotômicos ou politômicos), das dimensões do teste (unidimensional ou multidimensional) e do número de parâmetros na modelagem do item (de 1, 2 ou 3 parâmetros). Não existe até a presente data uma outra metodologia para modelar a realidade educacional. Mas, não devemos esquecer a firme asserção de Weber segundo a qual, “em ciência, cada um de nós sabe que suas realizações estarão ultrapassadas em dez, vinte ou cinquenta anos” e, também e Weber, a insistência em que “ toda contribuição científica exige ser ‘ultrapassada’ ou superada (MERTON, 1967a, p.28-29)

Embora os modelos da TRI ainda não tenham sido ultrapassados ou substituídos, observamos que a partir da década de 1980 com o surgimento dos modelos multiníveis, tornou-se possível incorporar aos modelos já existentes da TRI, modelos simplistas na visão de Goldstein (2001), análises mais elaboradas e conseqüentemente mais complexas na tentativa de melhor diagnosticar a realidade educacional. Conforme ressaltado por Goldstein (2001) é a tentativa da ciência em modelar melhor realidades complexas, através de formulações também complexas. Esta abordagem multinível proporcionada pela modelagem encontra paralelo na forma multidimensional que Weber via o mundo, sendo que este não é um todo coeso como uma grande unidade social. “O todo é uma mistura de partes em conflito.”(COLLINS, 2009, p.80).

Para podermos ter essa visão multinível, é necessário o cumprimento das seguintes etapas na elaboração do modelo:

- i) Inclusão de variáveis oriundas de diferentes hierarquias do sistema educacional, iniciando esta hierarquia com os alunos, estes por sua vez são agrupados em turmas e as turmas são agrupadas em escolas. Normalmente o número máximo de hierarquias nestes modelos não ultrapassa a três, tornando possível verificar as variações nos desempenhos dos alunos pelo fato de como eles são agrupados em diferentes turmas de diferentes escolas;
- ii) Construção de índices intra e extra escolares oriundos de questionários destinados a alunos, pais, professores e diretores;
- iii) Identificação do aluno nos diferentes anos de escolaridades, através, por exemplo, de um código que acompanhe ao longo de toda a sua trajetória escolar.

Este tipo de modelagem ainda está começando no Brasil. Podemos destacar o estudo pioneiro no campo acadêmico com o projeto GERES 2005, que foi uma avaliação longitudinal em larga escala aplicada em cinco municípios brasileiros e a publicação do livro pesquisa em eficácia escolar: Origens e trajetórias dos pesquisadores Nigel Brooke e Francisco Soares no ano de 2008, onde são apresentados os principais estudos nacionais e internacionais sobre esta metodologia.

Um dos principais focos desta metodologia é identificar quais fatores tornam as escolas eficazes, no sentido de proporcionar um maior ganho de conhecimento aos alunos, e como entender os motivos que, por exemplo, levam duas escolas que atendem a alunos com um mesmo perfil socioeconômico a apresentar desempenhos por vezes diferentes. Neste sentido, torna-se essencial, conforme apontado por Franco (2008) a utilização de pesquisas quantitativas e qualitativas de forma conjunta, utilizando os pontos fortes destas metodologias para um melhor entendimento da problemática educacional.

Também nestas abordagens complexas podemos mencionar uma tendência à elaboração de testes cognitivos no sentido de fornecer informações mais precisas sobre as dificuldades dos alunos e as estratégias para superá-las. Testes de desempenho como normalmente utilizados, servem um pouco a este propósito, através de técnicas de *retrofit* é possível utilizá-los com a finalidade cognitiva, estes estudos ainda estão em fase embrionária, mas já demonstram a suas grandes possibilidades pedagógicas.

Apresentamos, a seguir, um estudo longitudinal realizado em um determinado estado do nordeste brasileiro para a identificação de escolas eficazes, utilizando a metodologia multinível, numa tentativa de melhor entender a complexa realidade do Ensino Médio das escolas desse sistema de ensino.

Estudo de caso de um estado brasileiro

Para esse estudo, utilizamos como método de análise o modelo multinível, por meio do *software* HLN, tendo a seguinte coorte: Alunos do 1º ano do Ensino Médio, em 2012 e estes mesmos alunos no 3º ano do Ensino Médio, em 2014.

Assim, através do cálculo dos valores agregados das escolas, esse estudo teve como objetivo selecionar grupos de escolas eficazes e não eficazes para, através de pesquisa qualitativa nestas escolas, detectar que características estruturais ou de gestão estariam relacionadas com estes desempenhos. Servindo assim como fonte de informações para os funcionários da Secretaria de Educação do referido estado para orientar suas ações para a melhoria da qualidade das escolas como um todo.

Para realização desse estudo, foram adotadas cinco etapas para a análise:

1. Simulações para diferentes modelos utilizando o HLM
2. Definição do modelo para análise multinível
3. Cálculo dos resíduos
4. Classificação das escolas em quartis de proficiências e resíduos
5. Seleção de escolas para análise qualitativa

A seguir, detalhamos os procedimentos adotados em cada uma das etapas do estudo.

1. Simulações para diferentes modelos utilizando o HLM

Realizamos 10 simulações no HLM utilizando como variável independente a proficiência do ENEM- 2014 da disciplina em linguagens e códigos e suas tecnologias e como variáveis dependentes a proficiência no 1º ano do ensino médio em 2012 obtida através de avaliação realizada pelo estado.

As diferentes simulações foram geradas alterando as posições das variáveis independentes nos modelos multiníveis.

A principal características destes modelos é explicar a proficiência no ENEM, tendo como variáveis explicativas a proficiência no 1º ano do ensino médio e o turno dos alunos no nível 1 e os percentuais dos alunos no turno no nível 2. Os turnos utilizados forma: diurno, noturno e integral.

2. Definição do modelo para análise multinível

Apresentamos a seguir o modelo selecionado para análise. Este modelo, de dois níveis, sendo nível 1 aluno e nível 2 escola, apresentou a menor medida de *deviance* dentre os 10 modelos estudados:

Modelo adotado:

Level-1 Model

$$Y = B0 + B1*(PRF_LPT0) + B2*(NOTURNO) + B3*(INTEGRAL) + R$$

Level-2 Model

$$B0 = G00 + G01*(P_DIURNO) + G02*(P_NOTURN) + U0$$

$$B1 = G10 + U1$$

$$B2 = G20$$

$$B3 = G30$$

Random level-1 coefficient Reliability estimate

INTRCPT1, B0 0.892

PRF_LPT0, B1 0.275

Deviance = 606220.638048

3. Cálculo dos resíduos

Os valores dos resíduos encontrados nos modelos multiníveis se referem à eficácia escolar controlada pela proficiência de entrada do aluno no primeiro ano, o seu turno e o percentual nos turnos dentro de cada escola.

Para a elaboração dos arquivos com os valores de cada escola do estudo, um total de 593 escolas, num total de 56 666 alunos, os valores dos resíduos para Língua Portuguesa foram mergeados para a base de escolas.

Em seguida as proficiências e os resíduos foram classificados em quartis:

Enen_lc – Quartis para as proficiências na escala ENEM para Língua Portuguesa em 2014

EBINTRCP_LP– Quartis para os resíduos em Língua Portuguesa em 2014

4. Classificação das escolas em quartis de proficiências e resíduos

Realizamos análises cruzadas utilizando o SPSS.

Nossas escolas de interesse foram:

- 1 - Escolas com proficiência final baixa (quartil 1) e resíduo alto (quartis 3 ou 4)
- 2 - Escolas com proficiência final alta (quartil 1) e resíduo baixo (quartis 1 ou 2)
- 3 - Escolas com proficiência final alta (quartil 4) e resíduo alto (quartil 4)

Análise cruzada

Percentile Group of enem_ic* Percentile Group of EBINTRCP_LP Crosstabulation

			Percentile Group of EBINTRCP_LP				Total
			1	2	3	4	
Percentile Group of enem_ic	1	Count	116	30	2	0	148
		% within Percentile Group of enem_ic	78,4%	20,3%	1,4%	,0%	100,0%
	2	Count	6	103	36	3	148
		% within Percentile Group of enem_ic	4,1%	69,6%	24,3%	2,0%	100,0%
	3	Count	9	1	82	57	149
		% within Percentile Group of enem_ic	6,0%	,7%	55,0%	38,3%	100,0%
	4	Count	17	14	29	88	148
		% within Percentile Group of enem_ic	11,5%	9,5%	19,6%	59,5%	100,0%
Total		Count	148	148	149	148	593
		% within Percentile Group of enem_ic	25,0%	25,0%	25,1%	25,0%	100,0%

Fonte: Análise realizada pelos autores a partir de dados dos alunos da rede de ensino tomada como referência para esse estudo.

Ao analisarmos tabela de análise cruzada, deparamo-nos com escolas em três diferentes situações:

- Situação 1: 2 escolas que tiveram proficiência final baixa um resíduo alto, ou seja, conseguiram agregar proficiência no 2º e 3º ano do Ensino Médio, embora tenham proficiência no ENEM (Exame Nacional do Ensino Médio) baixa
- Situação 2: 31(17+14) escolas que tiveram proficiência final alta e resíduo baixo, ou seja, não conseguiram agregar proficiência no 2º e 3º ano, o que significa um esforço negativo. O valor esperado para estas escolas não foi atingido.
- Situação 3: 88 escolas tiveram proficiência final alta e resíduo alto e uma, o que indica um esforço positivo

Diante desse quadro, chegamos às seguintes conclusões:

- contatamos que escolas com proficiências altas tendem a terem valores agregados altos e escolas com proficiências baixas tendem a terem valores agregados baixos.
- há forte correlação da proficiência de entrada com a proficiência de saída, em torno de 60% no nível do aluno e 85% no nível da escola nos indica que a escola reproduz no 3º ano o resultado alcançado no final do 1º ano.
- tivemos poucas escolas com desempenho baixo em língua portuguesa e resíduo alto
- as escolas com proficiência alta e resíduo baixo estão todas relacionadas às escolas de tempo integral

Em face dos resultados observados no estudo de caso, podemos concluir que os princípios de reprodução apresentados por Bourdieu (1975) encontram similaridades nas escolas analisadas. Visto que alunos que, ao ingressarem no 1º ano do Ensino Médio apresentam um bom desempenho, ao sair do Ensino Médio, permanecem com bom desempenho. Ao passo que estudantes que apresentam baixo desempenho ao ingressarem no Ensino Médio, saem do Ensino Médio sem alteração no seu desempenho, permitindo-nos inferir que houve poucas alterações em suas aprendizagens.

Considerações finais:

Com estas modelagens complexas, as concepções metodológicas de macrossociologia e microsociologia se fundem. Podemos tentar conhecer os problemas educacionais penetrando nas bases de dados de formas desagrupadas por municípios, escolas, turmas e chegando mesmo ao nível dos estudantes e suas características de desempenho e características pessoais, no intuito de entender os sucessos e fracassos de escolas e alunos, que fatores intra e extra escolares podem provocar alterações positivas e/ou negativas nos desempenhos de escolas e alunos.

A perspectiva, portanto, é a utilização de modelagens complexas como instrumento sociológico para a compreensão das diversas nuances dos sistemas de ensino. Nota-se aqui que os objetivos das modelagens multiníveis se contrapõem com a visão marxista da teoria do conflito, que apesar de toda sua preocupação com a reprodução através da educação, não tem nada a dizer sobre os padrões de sucesso educacionais relacionados à prática na sala de aula.

Referências Bibliográficas

BONAMINO, Alicia e SOUSA, Sandra Zákia. “**Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola**”. In: Educação e Pesquisa, São Paulo, v. 38, n. 2, p. 373-388, abr./jun. 2012.

BOURDIEU, Pierre; PASSERON, Claude. **A reprodução**: elementos para uma teoria do sistema de ensino. Rio de Janeiro: Francisco Alves, 1975.

CAEd-UFJF. Microdados de avaliação educacional de 2012 e 2014.

COLLINS, Randall . **Quatro tradições Sociológicas**. Petrópolis: Editora Vozes, 2009. 277p.

FRANCO, Creso. e ALVES, MTG. “ A pesquisa em Eficácia Escolar no Brasil: Evidências sobre o efeito das escolas e fatores associados à eficácia escolar”. In: BROOKE, N.; SOARES, J. F. (Orgs). **Pesquisa em eficácia escolar: origem e trajetória**. Belo Horizonte: Editora UFMG, 2008. p. 482-500.

GIDDENS, Anthony. **As consequências da Modernidade**. São Paulo: Unesp, 1995.
GOLDSTEIN, Harvey. “Modelos da Realidade: Novas Abordagens para a
Compreensão de Processos Educacionais”. In: FRANCO, C (org). **Avaliação, Ciclos e
Promoção na Educação**. Porto Alegre: Editora Artmed. 2001, p. 85-99.
Hierarchical Linear and Nonlinear Modeling (HLM).

KARABEL, Jerome e HALSEY, A,H. **Power and Ideology in Education**. New York:
Oxford Press, 1977.

IBM - Statistical analysis software package - SPSS Statistics.