



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Um Modelo de Equações Estruturais aplicado a dados de satisfação de alunos do ensino superior privado

Autor: Luiz Guilherme Giordani
Orientador: Professora Dra. Luciana Neves Nunes

Porto Alegre, 27 de Novembro de 2015.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Um Modelos de Equações Estruturais
aplicado a dados de satisfação de alunos do
ensino superior privado

Autor: Luiz Guilherme Giordani

Trabalho de Conclusão de Curso
apresentado para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professora Dr. Luciana Neves Nunes (orientador)
Professora Dr. Liane Werner

Porto Alegre, 27 de Novembro de 2015.

Agradecimentos

Agradeço a minha família, meus pais Luiz Otávio e Maria Ângela e meus irmãos Luiz Otávio Júnior e Laura pelo apoio e carinho durante toda a minha graduação.

Também agradeço a professora Luciana, minha orientadora e mentora durante a construção desse trabalho. Com certeza a atenção, interesse e dedicação dela foram impactantes na qualidade final do trabalho.

Resumo

A Modelagem de Equações Estruturais (MEE) é uma poderosa técnica estatística que permite o pesquisador estudar o comportamento simultâneo de diversas variáveis em um modelo. No entanto, devido a sua nomenclatura diferenciada e caráter confirmatório, necessita-se um bom entendimento e familiaridade com a técnica. O presente trabalho tem como objetivo apresentar a técnica de Modelagem de Equações Estruturais e demonstrar uma aplicação prática em dados reais proveniente de pesquisas de mercado. Para as análises foi utilizado o *software* R na versão 3.2.2 e sua extensão chamada *lavaan* na versão 0.5-20. Para os menos familiarizados com o R, trechos de código das análises são apresentados para facilitar o entendimento. Para a aplicação da técnica de MEE foram usados dados provenientes de pesquisas de mercado feitas com estudantes do setor educacional privado brasileiro. O modelo testado tinha o propósito de entender a associação de cinco fatores (*Custo-benefício*, *Empregabilidade*, *Localização*, *Infraestrutura*, *Qualidade de Ensino* e *Tradição*) com a variável resposta satisfação do aluno. A validação dos referidos fatores foi feita através da Análise Fatorial Confirmatória (AFC). Considerando os resultados dessa análise, para o ajuste dos modelos, somente os fatores *Custo-benefício*, *Localização*, *Infraestrutura* e *Qualidade de Ensino* mostraram-se associados com a variável resposta. As medidas de ajuste dos modelos foram obtidas e nem todas se mostraram satisfatórias, entretanto uma das possíveis justificativas se deve ao fato dos dados analisados não terem sido coletados com o propósito de ser utilizada a MEE.

Palavras-chave: Modelagem de Equações Estruturais; Pesquisa de Mercado; Educação; *Software* R; *lavaan*.

Sumário

1. Introdução	6
1.1 Comentários iniciais.....	6
1.2 Objetivos	7
2. 2. A modelagem de equações estruturais.....	8
2.1 Introdução	8
2.2 O diagrama de caminhos.....	8
2.3 Modelo de medida e modelo estrutural.....	10
2.4 A linguagem gráfica e o software R	11
2.5 Criação de um modelo de equações estruturais.....	11
2.5.1 Especificação do modelo	12
2.5.2 Identificação do modelo	14
2.5.3 Seleção de medidas e coleta de dados.....	16
2.5.4 Estimação e validação do modelo	16
2.5.5 Interpretação, relato e modelos equivalentes.....	19
3. Estudo de caso.....	21
3.1 Apresentação dos dados	21
3.2 Especificação do modelo	21
3.3 Identificação do modelo.....	24
3.4 Estimação e validação do modelo	25
3.4.1 Validação do instrumento de medida	25
3.4.2 Modelo de equações estruturais.....	37
3.5 Interpretação das estimativas	51
3.6 Relato dos resultados	52
5. Considerações Finais	54
Referências	56
Anexos	58

1. Introdução

1.1 Comentários iniciais

Modelagem de Equações Estruturais (MEE ou Structural Equation Modeling - SEM em inglês) é uma técnica estatística que permite em uma única estrutura estudar a relação entre todas as variáveis do modelo (Hair et al, 2009). De tal modo, um modelo de equações estruturais se destaca por unir técnicas multivariadas em um único método de análise e apresentar o resultado em um gráfico, conhecido como diagrama de caminhos. Isso é consequência de uma série de equações lineares que simultaneamente regridem cada variável independente na variável dependente, levando em consideração o impacto das outras variáveis na variável de interesse (Hair et al, 2009). Portanto, um modelo de equações estruturais permite avaliar a relação entre as variáveis independentes, a magnitude de sua influência na variável dependente e a relação entre variáveis externas ao modelo e as variáveis independentes. Por esse motivo essa técnica é muito utilizada em estudos onde é de interesse estudar e obter respostas a partir de conjuntos de variáveis relacionadas, um exemplo de estudos desse tipo são pesquisas de mercado (Malhotra, 2012).

No contexto de pesquisa de mercado, a literatura científica tem produzido muitos artigos onde a MEE é utilizada para entender o impacto de medidas referentes ao consumidor e sua satisfação com o produto/serviço. Uma busca na base de dados de periódicos da CAPES com a sigla “MEE” e a palavra-chave “pesquisa de mercado” mostrou 24 artigos. Por exemplo, Urdan e Rodrigues (1999) adaptam um índice de satisfação americano para o mercado brasileiro utilizando dados da indústria automobilística e Farias e Santos (2000) apresentam o método MEE e aplicam a técnica em dados sobre a satisfação de idosos que se hospedaram em hotéis na região Nordeste do Brasil. Pesquisando na mesma base, mas com termos em inglês “*Structural Equation Modeling*” e “*Market Research*” encontrou 1.009 publicações. Por fim, realizou-se uma terceira pesquisa com os mesmos termos em inglês na base *Web of Science* e obtiveram-se 1.842 publicações. (Consultas feitas em 22/11/2015)

O foco desse trabalho é estudar o método MEE e aplicá-lo em dados provenientes de pesquisas de satisfação com alunos de cursos presenciais em Instituições de Ensino Superior (IES) privadas. Escolheu-se o tópico de educação

pelos seguintes motivos: 1. Disponibilidade dos dados; 2. Tópico com pouca referência em publicações científicas nacionais utilizando métodos estatísticos; 3. O recente crescimento do mercado educacional universitário privado no Brasil e a necessidade de desenvolver métodos e técnicas para a construção de inteligência para a tomada de decisões gerenciais na área. Publicações internacionais já utilizaram a MEE para entender os determinantes no aprendizado e na satisfação de alunos no modelo a distância como os trabalhos de Eom, Wen & Ashill (2006) e Martínez-carro (2009).

O software escolhido para as análises dos dados é o R, utilizando o pacote *lavaan* (Rosseel, 2012). Esse pacote do R foi desenvolvido para realizar diversas análises de variáveis latentes, permitindo de uma forma simples e intuitiva especificar e analisar o modelo, e desenhar o diagrama de caminhos com um pacote auxiliar, o *semPlot*. No decorrer do trabalho, as funções e saídas do pacote serão discutidas.

Finalmente, esse trabalho irá contribuir ao aplicar a modelagem de equações estruturais no mercado educacional universitário presencial, tópico pouco explorado na literatura nacional.

1.2 Objetivos

O presente trabalho possui como objetivo principal a descrição da técnica de modelagem de equações estruturais. No decorrer do trabalho, serão apresentados os principais embasamentos teóricos, pressupostos e o processo de estimação e validação de um modelo de equações estruturais.

Posterior a descrição da técnica, é apresentado um estudo de caso com o objetivo de entender os principais determinantes na satisfação de alunos de três IES localizadas em três cidades brasileiras. Durante esse estudo, visa-se apresentar a relação entre a MEE e a análise fatorial confirmatória (AFC) e descrever a utilização do R e do *lavaan*.

2.2. A modelagem de equações estruturais

2.1 Introdução

A MEE é uma técnica confirmatória que permite estudar a relação de dependência, independência e interdependência entre variáveis. Um modelo de equações estruturais alcança isso através de uma série de regressões múltiplas regredidas de forma separada e simultaneamente, e interdependentes entre si (Hair et al, 2009). Essas equações têm por objetivo estudar efeitos de causalidade, apesar de que inferir a causalidade é algo muito mais complexo. Resumidamente, por Farias e Santos (2000), MEE tem por objetivo estudar a influência de X em Y e posteriormente a influência de Y em Z.

Essa é a motivação do uso de um modelo de equações estruturais. A possibilidade de estudar relações entre variáveis em conjuntos de equações e mostrar essas relações em um método gráfico próprio, o **diagrama de caminhos**.

Além disso, um modelo de equações estruturais permite a incorporação de variáveis não observadas, chamadas de **variáveis latentes** (conhecidas também como construtos latentes), obtidas através da relação entre variáveis observadas (chamadas também de variáveis manifestas, medidas ou indicadores). Dentro do diagrama de caminhos, as variáveis latentes podem ser do tipo **exógena** ou **endógena**, sendo essa nomenclatura relacionada com quais variáveis do modelo a medida se relaciona.

Por ser uma técnica confirmatória, a MEE necessita que o modelo a ser construído tenha um bom embasamento teórico, seja esse proveniente de estudos anteriores ou de estudos empíricos do pesquisador. Posterior à especificação teórica utiliza-se o **modelo estrutural** e o **modelo de medida** para representar a teoria. O modelo de medida apresenta a relação entre as variáveis observadas e as variáveis latentes enquanto que o modelo estrutural apresenta somente a relação entre variáveis latentes.

2.2 O diagrama de caminhos

O método gráfico comumente utilizado para o modelo de equações estruturais é o diagrama de caminhos. Esse, por um conjunto de formas geométricas e setas, permite observar de uma forma prática todas as relações medidas pelo modelo. O quadro 1 mostra a notação mais utilizada dentro do diagrama de caminhos.

Quadro 1 – Símbolos de um diagrama de caminhos [Adaptado de PEREIRA, 2013]

Símbolo	Significado
□	Variável observada
○	Variável latente
→	Relação de causa
↔	Relação de correlação/covariância

Definida a simbologia, a figura 1 apresenta um exemplo simples de MEE por um diagrama de caminhos.

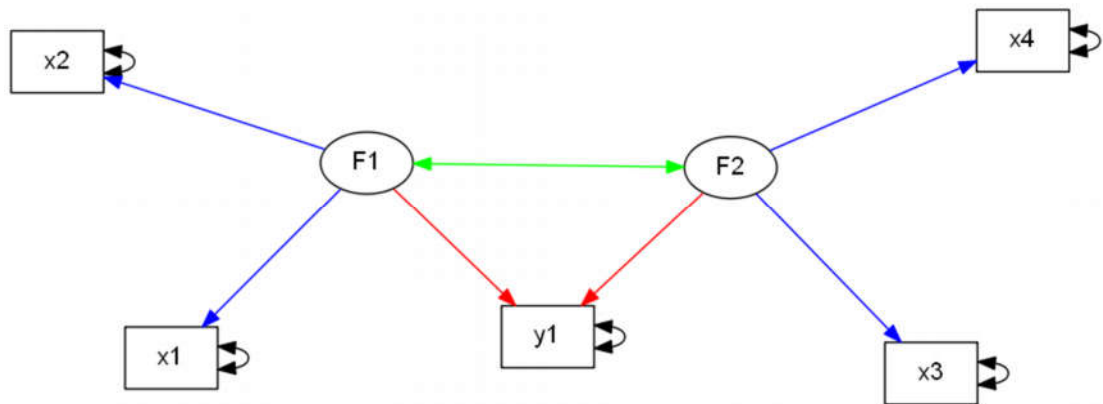


Figura 1- Diagrama de caminhos exemplificando um modelo de equações estruturais

Na figura 1 as variáveis latentes F1 e F2 são criadas a partir das variáveis observadas x1, x2, x3 e x4. Essas variáveis, por sua vez, influenciam de forma causal a variável y1. As setas pretas recíprocas apontando para as variáveis medidas são estimativas de variância ou de erro de mensuração. A reta verde entre F1 e F2 é uma medida de covariância entre as duas variáveis.

Ademais, F1 e F2 são **variáveis exógenas** do modelo, pois estão relacionadas somente por variáveis não explicadas por nenhuma outra no modelo, ou seja, variáveis independentes, enquanto que y1 é uma **variável endógena**, pois é determinada (ou dependente) de variáveis exógenas (F1 e F2).

2.3 Modelo de medida e modelo estrutural

O embasamento teórico possui um papel fundamental para a técnica de MEE, pois não é comum se utilizar métodos exploratórios dentro da MEE para entender o relacionamento entre as variáveis. Consequentemente é necessário que o pesquisador construa antecipadamente as relações esperadas entre as variáveis da sua pesquisa. Por esse motivo, é comum especificar essas relações pelo modelo de medida e pelo modelo estrutural.

O modelo de medida especifica quais são as relações entre as variáveis observadas e as variáveis latentes. É dentro desse modelo em que se definem quais construtos são manifestados por quais variáveis. O modelo de medida pode ser representado por um conjunto de equações, como é mostrado a seguir, utilizando o exemplo da figura 1.

$$x_1 = F_1 + e_1$$

$$x_2 = F_1 + e_2$$

$$x_3 = F_2 + e_3$$

$$x_4 = F_2 + e_4$$

Onde e_1, e_2, e_3, e_4 são erros de medida das variáveis x_1, x_2, x_3 e x_4 respectivamente.

O modelo estrutural possui propósito semelhante ao do modelo de medida, porém serve para especificar quais são as relações de dependência entre as variáveis latentes do modelo e a resposta. Esse modelo também pode ser representado por um conjunto de equações. Então, ainda utilizando o exemplo da figura 1, mostra-se as equações a seguir

$$F_1 = y_1 + r_1$$

$$F_2 = y_1 + r_2$$

Onde r_1 e r_2 são os resíduos representando o erro de predição do modelo.

A figura 2 mostra o modelo de medida e o modelo estrutural para o exemplo da figura 1 dentro do diagrama de caminhos.

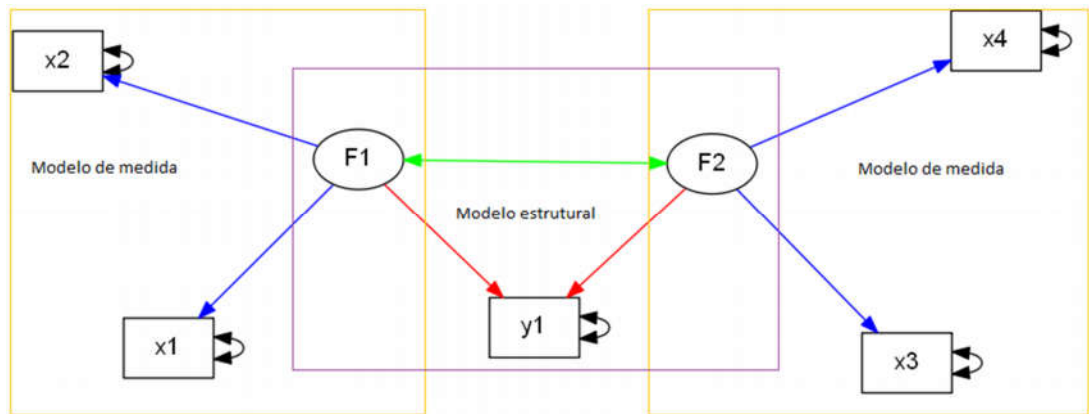


Figura 2- Modelo de medida e modelo estrutural

2.4 A linguagem gráfica e o software R

Nesse trabalho escolheu-se por utilizar o *software* R na versão 3.2.2 (R Core Team, 2015) como ferramenta para a análise dos dados devido a sua crescente popularidade e flexibilidade. Adicionalmente, o R é uma linguagem de programação aberta e gratuita, podendo então, os códigos desse trabalho servirem de base para outros estudos.

Também por ser um *software* livre, o R possui diversas extensões (chamadas pacotes) criadas por colaboradores. Nesse trabalho será utilizado o pacote *lavaan* na versão 0.5-20 (*LA*tent *VA*riables *AN*alysis) (Rosseel, 2012). O *lavaan* foi desenvolvido para aplicação de diversas análises de variáveis latentes, entre elas, a MEE. A grande vantagem do *lavaan* é que ele utiliza um método similar ao procedimento *lm* do R para especificar os modelos. O *lavaan* será brevemente especificado nesse trabalho e mais detalhes podem ser encontrados em Pereira (2013) e Rosseel (2012).

Adicionalmente, a não ser que especificado, os gráficos construídos nesse trabalho foram produzidos utilizando a linguagem gráfica *graphviz* (diminutivo de *Graph Visualization Software*) na versão 2.38. Essa ferramenta é principalmente utilizada na construção de grafos e fluxogramas.

2.5 Criação de um modelo de equações estruturais

Na figura 3 é apresentado um algoritmo simples que tem por objetivo apresentar e detalhar resumidamente os passos para a construção e análise de um modelo de equações estruturais. Cada passo da figura será trabalhado na sequência.

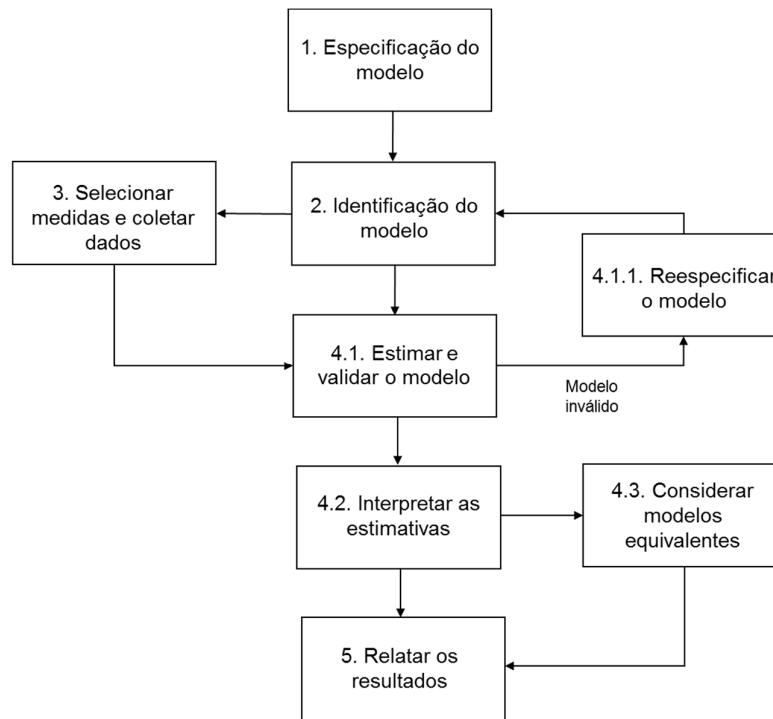


Figura 3- Fluxograma com o passo a passo da MEE [Traduzido e adaptado de KLINE, 2011]

2.5.1 Especificação do modelo

Especificar o modelo a ser construído significa fundamentar a base teórica para o modelo e argumentar sobre sua validade, pelo menos no sentido teórico. É também nesse momento da pesquisa que devem ser definidas as variáveis indicadoras, variáveis latentes e suas funções no modelo (variável exógena ou endógena).

Primeiramente, então, são definidas as variáveis de interesse para o estudo. Toda potencial variável do modelo deve possuir uma explicação plausível para sua presença. Esta pode ser devido ao seu uso em outros estudos e/ou relação empírica (Kline, 2011). Na literatura científica referenciada, é comum os autores descreverem sobre o sentido do efeito da variável na resposta e a justificativa para ela estar no modelo.

Porém, o pesquisador precisa ter cuidado com erros graves durante a fase de especificação, como, por exemplo, incluir variáveis em excesso no modelo, isso pode causar problemas na fase seguinte (identificação) ou tornar o modelo de difícil interpretação (Hair et al, 2009). Entretanto, o pior erro nessa fase é o de não incluir

uma variável importante no modelo, pois isso pode viciar as estimativas e invalidar o modelo.

Consideradas as variáveis de interesse e as relações entre variáveis medidas e indicadores, pode-se representar essas relações pelo modelo estrutural e o modelo de medida. Posteriormente o diagrama de caminhos pode ser construído. Nesse momento, o pesquisador precisa definir todas as relações de interesse do modelo (relações de causa, variância e covariância), pode também considerar um modelo **não recursivo**, onde duas variáveis afetam uma a outra simultaneamente, ou relações de **mediação**, quando o efeito de uma variável é modificado a partir de outra. A figura 4 exemplifica esses dois casos.

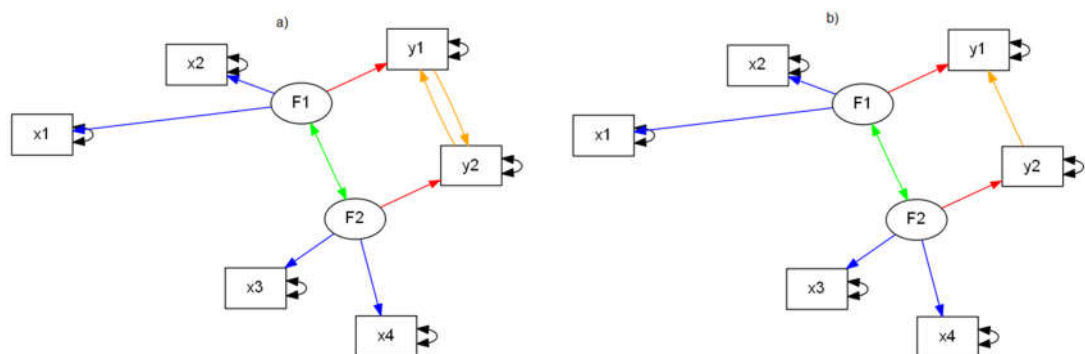


Figura 4 – (a) Exemplo de modelo não recursivo (b) Exemplo de efeito mediador

Um modelo não recursivo é aquele em que pelo menos duas variáveis afetam uma a outra simultaneamente. Por exemplo na figura 4(a), existe um efeito direto de y_1 em y_2 e ao mesmo tempo existe um efeito direto de y_2 em y_1 . No momento de interpretação é comum interpretar cada relação de forma independente.

O efeito mediador pode ser completo ou parcial. Completo é quando o único efeito de uma variável é intermediado por outra variável. Parcial é quando existe um efeito direto entre as duas variáveis e um efeito intermediado por uma variável intermediária. A figura 4 (b) exemplifica um efeito mediador completo. F_2 só afeta y_1 através do seu efeito em y_2 .

Para mais informações sobre modelos não recursivos e efeitos mediadores, recomenda-se a leitura de Kline (2011).

2.5.2 Identificação do modelo

Identificar o modelo significa concluir que o modelo especificado pode ser estimado. A fase de identificação do modelo não está relacionada com tamanho de amostra, mas sim com o número de variáveis que estão sendo consideradas para o modelo final.

Para um modelo ser identificado, ele deve cumprir duas regras: 1) Possuir graus de liberdade maior ou igual a zero; 2) Definir o escalonamento dos erros e das variáveis latentes (Kline, 2011).

Quanto aos graus de liberdade, eles são calculados conforme a seguinte equação:

$$g.l = \frac{n(n+1)}{2} - q$$

Onde,

n = Quantidade de variáveis indicadoras no modelo

q = Quantidade de relações a serem estimadas pelo modelo

Mais especificamente, **n** é a quantidade de termos na diagonal inferior (ou superior) da matriz de covariância e **q** é a quantidade de parâmetros a serem estimados pelo modelo no diagrama de caminhos.

Após calculado os graus de liberdade, a quantidade resultante de graus de liberdade pode encaixar o modelo em três estados diferentes (Lemke, 2005), são eles:

- **Underidentified:** Quando os graus de liberdade são menores que zero, o que impossibilita estimação dos parâmetros. Um modelo nesse estado precisa ser reespecificado;
- **Just-identified:** Estado em que os graus de liberdade são iguais a zero. O modelo estimará os parâmetros e se ajustará perfeitamente a amostra fornecida, porém, não será possível testar hipóteses, tornando esse modelo não interessante;
- **Overidentified:** Estado ótimo. Ocorre quando os graus de liberdade são maiores do que zero, e, conseqüentemente, é possível testar hipóteses do modelo.

Além de considerar os graus de liberdade, para o modelo poder ser estimado necessita-se que algumas relações do modelo sejam escalonadas. Escalonar as relações significa fixar em um valor arbitrário algumas relações do modelo. A escolha do valor e da variável normalmente é arbitrário e não modificam as estimativas.

As primeiras variáveis consideradas são os erros das variáveis observadas (os e's das equações de medida). Considera-se que somente os erros de medida afetam as variáveis indicadoras e atribui-se o valor 1 a essa relação, ou seja, supõe-se que os erros de medida afetam os valores observados com magnitude pressuposta 1. Essa relação está exemplificada na figura 5 onde os erros de medida das variáveis observadas x_1 , x_2 , x_3 e x_4 são respectivamente e_1 , e_2 , e_3 e e_4 e a relação destes com as variáveis foi fixado em 1. Os diagramas da figura 5 são os únicos em que se mostrará os erros de medida, nesse trabalho foi decidido ocultá-los por motivos de estética.

As outras variáveis consideradas são as variáveis latentes. Nesse caso existem duas escolhas. Fixar uma variável indicadora para cada variável latente com peso 1 ou fixar a variância dos fatores em 1. A escolha da primeira gera variáveis não padronizadas enquanto a segunda gera variáveis padronizadas. A escolha de qual variável indicadora fixar em 1 é arbitrária e não muda as estimativas. A figura 5 demonstra visualmente essa relação.

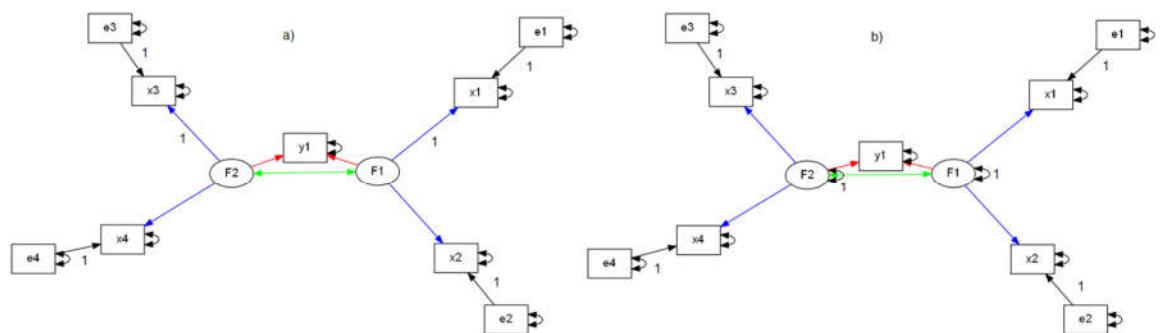


Figura 5- (a) Escalonamento que resulta em variáveis não padronizadas
 (b) Escalonamento que resulta em variáveis padronizadas

A diferença entre variáveis padronizadas e não padronizadas costuma não alterar a interpretação do modelo. A escolha baseia-se somente em como serão interpretadas as relações entre fatores. No modelo não padronizado, as setas com duas direções indicam covariâncias, enquanto no modelo padronizado elas indicam correlações (Kline, 2011). A escolha também depende se o modelo vai ser comparado com outros modelos construídos com outras amostras independentes.

2.5.3 Seleção de medidas e coleta de dados

Nessa fase se decide de onde virão os dados e quais escalas serão utilizadas para medir as variáveis. Após a coleta é recomendado realizar uma análise descritiva da amostra para a identificação de possíveis *outliers*, valores estranhos e faltantes. Como essa discussão não faz parte do objetivo do trabalho é recomendada a leitura do capítulo específico sobre esse tema em Kline (2011).

2.5.4 Estimação e validação do modelo

2.5.4.1 Métodos de estimação

Existem diversos tipos de métodos para estimar os parâmetros do modelo construído. A escolha do método depende principalmente do tamanho amostral e da distribuição dos dados. Presume-se também que os dados são contínuos, porém, como vai ser apresentado, se os dados discretos possuem algumas propriedades, alguns métodos para variáveis contínuas podem ser utilizados. Também é importante notar que como MEE estima o modelo a partir de uma matriz de covariância ou de correlação, pode acontecer de o método de estimação perder algumas propriedades dependendo da matriz de dados utilizada.

O primeiro método a ser apresentado é o da **máxima verossimilhança** (MLE, do inglês *Maximum Likelihood*). Esse assume normalidade multivariada dos dados, variáveis contínuas e não padronizadas, e um modelo corretamente especificado (Kline, 2011). Estimar o modelo com variáveis padronizadas com o MLE pode resultar em estimativas viciadas das relações e dos erros padrões.

Errar na especificação do modelo ao utilizar o MLE resulta no problema de propagação dos erros, isso significa que ao errar na especificação da relação entre pelo menos duas variáveis, são geradas estimativas tendenciosas que contaminam todo o modelo.

Sobre o tamanho amostral, o método de MLE pode gerar estimativas para amostras muito pequenas, porém os resultados são pouco confiáveis. Um tamanho ideal de amostra sugerido é entre 150 e 400 observações. Tamanhos maiores que isso tornam as estimativas do MLE muito sensíveis a novas observações (Hair et al, 2009).

Apesar do método MLE pressupor variáveis contínuas, ele pode ser utilizado também para variáveis categorizadas que possuam quatro ou mais categorias (Lemke, 2005), porém aconselha-se cuidado, pois alguns resultados de outros estudos utilizando simulação indicam que os parâmetros são superestimados ao

utilizar variáveis categóricas, e isso piora conforme a distribuição das respostas se afasta da normalidade (Kline, 2011).

Outro método que pressupõe normalidade multivariada é o dos **mínimos quadrados generalizados** (GLS, do inglês *Generalized Least Squares*). Esse possui a vantagem de ter o tempo de computação mais rápido que do MLE (Kline, 2011) e ser um pouco mais brando nas suposições de normalidade (Pereira, 2013). A primeira vantagem é irrelevante com a potência dos computadores atuais. Outra desvantagem é que esse método costuma gerar erros do tipo I com maior frequência.

Quando as variáveis não seguem distribuição normal multivariada, existem dois caminhos possíveis para estimar os dados. Um deles é a estimação do modelo por *bootstrap*. Resultados de outros estudos mostraram que para tamanhos de amostra maiores que 200 os resultados obtidos por bootstrap possuíram menores erros de estimação do que a estimação direta por MLE com dados não normais. Contudo, para tamanhos de amostra iguais a 100, os resultados não se mantiveram (Kline, 2011)

A outra opção é utilizar o método de estimação **assintótica de distribuição livre** (ADF, do inglês *Asymptotically Distribution-Free*). Esse não supõe nenhuma distribuição a priori dos dados. Porém necessita de um tempo de processamento computacional muito elevado para modelos complexos. Além disso, requer um tamanho amostral elevado. Para modelos simples, sugere-se entre 200 a 500 observações e o número sugerido cresce conforme aumenta a complexidade do modelo (Kline, 2011). Por não supor nenhuma distribuição dos dados, é esse o método de estimação sugerido para dados categóricos.

2.5.4.2 Medidas de validação do modelo

Existem diversas estatísticas utilizadas para validar o modelo construído. Nessa seção será apresentado brevemente as medidas implementadas no pacote *lavaan*. A profundidade teórica para cada método será breve, logo, se o leitor quiser obter maior conhecimento do assunto é recomendada a leitura de Kline (2011), Pereira (2013) ou Hair et al. (2009).

A primeira medida apresentada é também a mais comumente utilizada e a primeira a ser citada. É o teste χ^2 (Qui-quadrado) de ajustamento. A estatística de teste possui relação direta com o método de estimação escolhido e com o tamanho amostral. As hipóteses testadas são:

H₀: A matriz de covariância amostral é igual a matriz de covariância estimada.

H₁: As matrizes de covariância são diferentes.

Logo, para um bom modelo, valores pequenos da estatística qui-quadrado são esperados o que indicaria um p-valor alto e um modelo ideal. Porém, recomenda-se cuidado ao utilizar somente essa medida como validação do modelo pois ela é altamente sensível ao tamanho de amostra e a desvios de normalidade. Além disso, a sua conclusão é muito simples e não reflete o ajuste como um todo. Adicionalmente, Kline (2011) comenta que a não rejeição da hipótese nula do teste qui-quadrado somente leva a conclusão de que o modelo é consistente com a matriz de dados, mas não que este está correto.

Para complementar as deficiências do teste qui-quadrado, utiliza-se outras medidas, e as mais comuns são: **GFI**, **AGFI**, **RMSR**, **RMSEA**, **CFI**, **NFI**, e, **NNFI**. Os índices GFI (*Goodness-of-Fit Index*), AGFI (*Adjusted GFI*) e RMSR (*Root Mean Square Residual*) são índices de ajuste geral do modelo construídos a partir dos resíduos obtidos pela diferença entre a matriz de covariância observada e a estimada. Os índices GFI e o AGFI são muito similares as medidas R² e R²-ajustado da regressão linear pois também medem a proporção de variância explicada pelo modelo. O GFI é a proporção das covariâncias observadas explicadas pelo modelo. O índice AGFI penaliza o GFI pelos graus de liberdade. Tanto para o GFI como para o AGFI, valores próximos a 1 são considerados ideais com 0,90 e 0,95 considerados aceitáveis.

O RMSR é a raiz quadrada da matriz de resíduos assumindo que o modelo é correto (Pereira, 2013). Valores mais próximos de 0 são preferíveis com 0,05 sendo o mínimo admissível.

As medidas NFI (*Normed Fit Index*), CFI (*Comparative Fit Index*) e NNFI (*Non-Normal Fit Index*) são índices de ajuste incremental que comparam o desempenho do modelo com o modelo nulo (sem variáveis). O NFI apresenta um incremento em % do modelo ajustado relativo ao modelo nulo. O CFI aplica uma correção ao NFI pois esse possui uma tendência de subestimar o ajuste em amostras pequenas. Por fim, o NNFI é similar ao NFI, porém os seus resultados são não normalizados, tanto o NFI quanto o NNFI costumam gerar valores próximos e possuem interpretação similar (Hair et al, 2009). Para todas essas medidas, o ideal são valores próximos de 1, sendo pelo menos 0,90 um valor ideal.

O RMSEA (*Root Mean Square Error of Approximation*) é um índice de parcimônia que tenta corrigir as falhas da medida qui-quadrado. Na sua equação, o índice “penaliza” o valor da estatística χ^2 ao subtrair seu valor pelos graus de liberdade. Valores ideais são os mais próximos de 0, sendo 0,10 aceitável e 0,05 ideal.

O quadro 2 apresenta todas as medidas geradas pelo modelo e os seus respectivos valores de referência.

Quadro 2 – Resumo das medidas de validação do modelo

Estatística	Valor de referência
Qui-quadrado	Quanto menor melhor ou valor de p maior que 0,05
GFI, AGFI	Bom ajuste em 0,9 Ajuste ótimo em 0,95
RMSR	Ajuste ótimo em 0,05
NFI, CFI e NNFI	Valor ideal 0,9
RMSEA	Aceitável quando igual a 0,10 Ótimo quando igual a 0,05

As equações para essas medidas se encontram no anexo 1.

2.5.4.3 Reespecificar o modelo

Caso o modelo não esteja bem ajustado pelos critérios de validação ou não foi identificado na fase de identificação, este precisa ser reespecificado. Esse processo consiste em considerar novas variáveis (observadas ou latentes), considerar possíveis relações não especificadas, ou até excluir algumas relações previstas anteriormente. Além disso, pode-se considerar efeitos mediadores e/ou moderadores (interações)

É importante lembrar que nessa fase de re-especificação as alterações devem fazer sentido com a teoria proposta e devem ser propriamente justificadas.

2.5.5 Interpretação, relato e modelos equivalentes

Com o modelo ajustado e aceito é este o momento de concluir o estudo e descrever os resultados.

A interpretação e o relato das estimativas dependerão do objetivo do estudo e da padronização das variáveis já que modelos com variáveis padronizadas

possuem uma interpretação diferenciada de suas estimativas. Mais detalhes serão discutidos na seção de análise de dados.

Considerar modelos equivalentes significa avaliar se um modelo com menos ou mais variáveis chegaria a mesma conclusão que o modelo proposto. Ao mesmo tempo, nessa fase o pesquisador deve considerar se um modelo com variáveis muito diferentes poderia chegar a uma conclusão similar à proposta pelo seu projeto.

3. Estudo de caso

3.1 Apresentação dos dados

Para a parte prática do trabalho, serão construídos três modelos de equações estruturais com dados provenientes de pesquisas encomendadas por três IES, sendo uma de Brasília, outra de São Paulo e outra de uma cidade da região metropolitana de São Paulo. O público alvo das pesquisas foram os alunos das instituições e essas pesquisas possuíam como objetivo principal avaliar a satisfação dos alunos com a instituição que em estavam matriculados e entender o porquê de terem escolhido o atual local de estudo. A pesquisa foi conduzida por uma empresa de pesquisa de mercado e as respostas foram obtidas por conveniência com auxílio da IES. O questionário foi aplicado de forma *online*. A tabela 1 fornece um resumo dos respondentes.

Tabela 1: Resumo das amostras coletadas

	São Paulo	Brasília	Metrop. SP
Total de Alunos*	21.025	14.139	7.114
Total de respondentes	3.792	1.785	1.141
(% do total)	(18,0%)	(12,6%)	(16,0%)
Sexo:			
Masculino	34,8%	35,7%	37,5%
Feminino	65,2%	64,3%	62,5%
Idade Média	26 anos	27 anos	28 anos
(Desvio Padrão)	(8,22)	(9,70)	(8,51)

* Conforme censo INEP 2013

3.2 Especificação do modelo

Os modelos que serão construídos nesse trabalho possuem o propósito de entender quais características das IES são importantes para os alunos. Para medir isso, o modelo de equações estruturais irá “desconstruir” a satisfação geral do aluno em variáveis latentes que representam os principais fatores considerados pelos alunos no momento de decidir em qual instituição se matricular. As variáveis latentes, perguntas do questionário que as compõem e as escalas utilizadas são mostradas no quadro 3.

As perguntas são comuns nas três pesquisas realizadas. A pergunta de satisfação geral que se encontra no começo do questionário e é uma das primeiras respondidas. Para os atributos que compõem as variáveis latentes, a pergunta realizada era: *“Utilizando a escala abaixo, aponte o quanto você concorda ou discorda com cada uma das seguintes afirmativas em relação à IES. Para isso, quanto mais você concordar com a afirmação, mais próxima de 7 deve ser a sua resposta e, quanto mais você discordar da afirmação, mais próxima de 1 deve ser a sua resposta.”* Essa pergunta encontra-se mais próximo do final do questionário. Como opção de resposta, o respondente marcava a opção enumerada que correspondia com sua opinião.

A origem das variáveis latentes é empírica e a identificação de sua existência são provenientes de análises exploratórias. Em outros estudos, para as mesmas perguntas que compõem as variáveis latentes, foram realizadas análises fatoriais exploratórias e cada fator foi nomeado em função das variáveis que possuíam a maior carga fatorial em um específico fator. Logo, nesse trabalho será a primeira vez em que se utilizará uma técnica confirmatória nessas variáveis. Assim, se o resultado for positivo, será confirmada a existência desses fatores.

Em termos de hipóteses associadas ao modelo, espera-se que cada variável latente possua uma relação significativa e positiva com a satisfação geral dos alunos. São objetivos da construção desses modelos confirmar se esses construtos estão associados com a satisfação do aluno, entender a relação entre eles e descobrir o que é mais impactante na satisfação do aluno. A figura 6 mostra o diagrama de caminhos e os caminhos que serão estimados. A construção de três modelos também permite comparar os resultados entre diferentes regiões.

Quadro 3 - Perguntas do questionário e respectivas variáveis latentes

Pergunta	Variável no modelo	Variável latente	Escala
1 - De uma forma geral, o quanto você está satisfeito com a sua IES	Satisf.	Satisfação	1 a 5*
2 - Tem as mensalidades mais acessíveis	CxB_1	Custo Benefício	1 a 7**
3 - Oferece as melhores bolsas de estudo	CxB_2		1 a 7
4 - Cabe no meu bolso	CxB_3		1 a 7
5 - Garantirá as melhores oportunidades de emprego	Emprg_1	Empregabilidade	1 a 7
6 - É um diferencial para o diploma	Emprg_2		1 a 7
7 - É reconhecida pelo mercado de trabalho	Emprg_3		1 a 7
8 - Possui o melhor campus da região	Infra_1	Infraestrutura	1 a 7
9 - Dispõe de ótimos serviços de atendimento aos alunos	Infra_2		1 a 7
10 - Tem ótimas instalações físicas para estudo (salas de aula, bibliotecas, laboratórios, etc)	Infra_3		1 a 7
11 - É fácil de chegar	Local_1	Localização	1 a 7
12 - Fica próxima de estações de transporte público	Local_2		1 a 7
13 - Possui corpo docente qualificado	Q_Ens_1	Qualidade de ensino	1 a 7
14 - Oferece a melhor qualidade de ensino	Q_Ens_2		1 a 7
15 - Exige esforço do aluno para se formar	Q_Ens_3		1 a 7
16 - É bem avaliada pelo MEC/ Guia do Estudante	Q_Ens_4		1 a 7
17 - É recomendada por pessoas que eu conheço	Trad_1	Tradição	1 a 7
18 - É uma das faculdades mais tradicionais da região	Trad_2		1 a 7
19 - É referência no curso que tenho interesse	Trad_3		1 a 7
20 - Dá prestígio/status para quem estuda lá	Trad_4		1 a 7

* Escala Likert de 5 pontos, onde 1 = Muito Insatisfeito e 5 = Muito Satisfeito

** Escala de 7 pontos, onde 1 = Discorda Totalmente e 7 = Concorde Totalmente (Válido para perguntas 2 a 20).

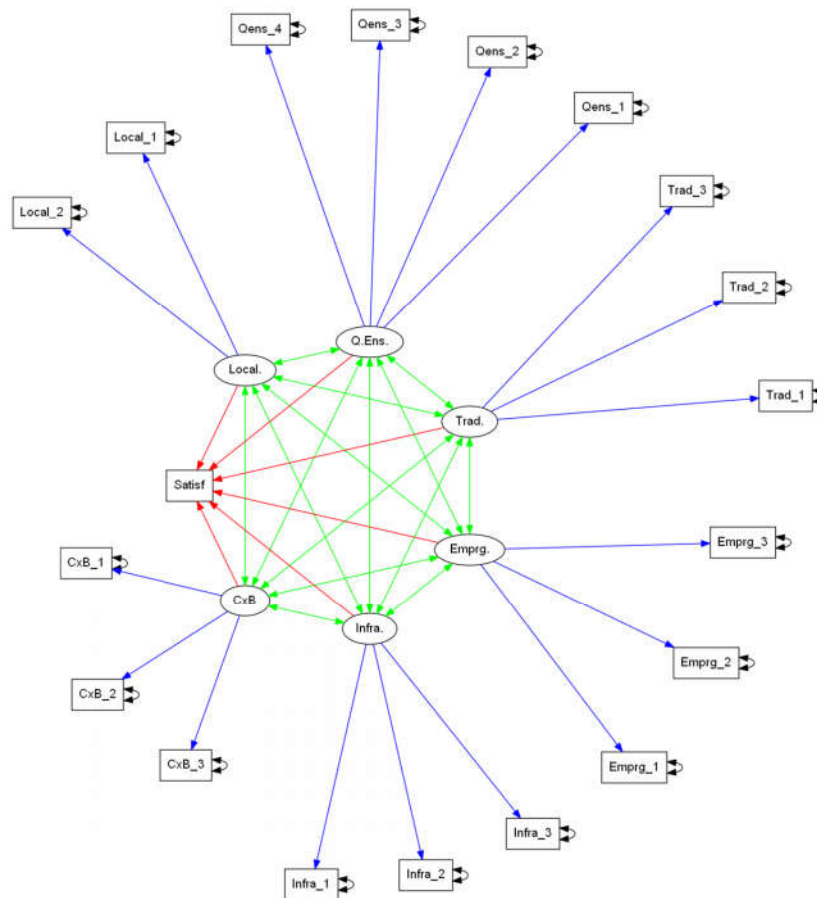


Figura 6 - Modelo de equações estruturais para explicar a satisfação geral.

3.3 Identificação do modelo

O primeiro passo para identificar o modelo é conferir a equação dos graus de liberdade. O modelo da figura 6 possui 60 relações a serem estimadas, enquanto existem 20 variáveis indicadoras. Logo, pela equação:

$$g.l = \frac{n(n + 1)}{2} - q = \frac{20(21)}{2} - 60 = 150$$

Então, com 150 graus de liberdade, o modelo é *overidentified*. Ou seja, todas as relações do modelo podem ser estimadas e testadas.

O segundo passo é definir o escalonamento das variáveis. Nesse caso, os modelos serão construídos para exibir os resultados padronizados e não padronizados por motivos de exemplificação. Na fase de interpretação das estimativas será considerado qual tipo de escalonamento é mais apropriado.

3.4 Estimação e validação do modelo

3.4.1 Validação do instrumento de medida

A primeira parte dessa seção consistirá em realizar uma análise fatorial confirmatória para ratificar que as variáveis observadas de fato podem medir as variáveis latentes. Essa prática é comum em casos onde não existem referenciais teóricos suficientes para afirmar a existência das relações apresentadas na fase de especificação.

A validação ocorrerá nos três bancos de dados, mas o processo somente será detalhado nos dados da IES localizada na região metropolitana de São Paulo. Essa escolha foi arbitrária e a construção do primeiro modelo fatorial será feita acompanhada de código do R e de explicações sobre o mesmo.

O primeiro passo no *software* é especificar o modelo e as variáveis que serão estimadas. O código que gera o modelo no *lavaan* é dado pela figura

```
require(lavaan)
modelo_cfa1='CxB =~ CxB_1 + CxB_2+ CxB_3
Emprg =~ Emprg_1 + Emprg_2 + Emprg_3
Infra =~ Infra_1 + Infra_2 + Infra_3
Local =~ Local_1 + Local_2
Q_Ens =~ Q_Ens_1 + Q_Ens_2 + Q_Ens_3 + Q_Ens_4
Trad =~ Trad_1 + Trad_2 + Trad_3 + Trad_4'
```

Figura 7- Especificação do modelo no R

O objeto *modelo_cfa1* é onde guardamos o modelo especificado. No pacote *lavaan* não existe uma função própria para especificar o modelo, somente requer que as relações estejam entre aspas. O tipo de relação muda também conforme o símbolo utilizado depois da variável especificada, utilizando como exemplo a primeira relação da figura 7

```
CxB =~ CxB_1 + CxB_2 + CxB_3
```

Essa linha traduz para o *software* que a variável latente Custo-Benefício é manifestada (\approx) pelas perguntas “*Tem as mensalidades mais acessíveis*” (CxB_1), “*Oferece as melhores bolsas de estudo*” (CxB_2) e “*Cabe no meu bolso*” (CxB_3). Outras relações que podem ser especificadas no pacote estão no quadro 4.

Quadro 4 – Outras relações possíveis de serem especificadas no *lavaan*

[Adapatado de Pereira (2013)]

Relação	Operador
Regressão	~
Covariância Residual	~~
Intercepto	~1
Parâmetro definido	:=
Restrição de igualdade	==
Restrições de desigualdade	> ou <

Com o modelo especificado, os seguintes comandos especificados na figura 8 ajustam o modelo e exibem as medidas de validação.

```
cfa_GSP=cfa(modelo_cfa1, estimator="WLS", data=GSP_ss)
summary(cfa_GSP, fit.measures=TRUE, standardized=TRUE, rsquare=TRUE)
fitMeasures(cfa_GSP, c("gfi", "agfi"))
```

Figura 8 - Realizando as análises dentro do R utilizando funções do *lavaan*

A função *cfa* realiza a análise fatorial confirmatória nos dados. Precisamos alimentar essa função com o modelo especificado, o método de estimação e os dados. Nesse caso, como vamos utilizar a estimação por ADF (No *lavaan* chamada de WLS, do inglês *Weighted Least Squares*), precisamos providenciar a base de dados inteira. Para outros métodos de estimação pode-se somente providenciar a matriz de covariância e o tamanho da amostra. Utilizamos esse método de estimação, pois as variáveis indicadoras são categóricas ordinais e não possuem distribuição de suas respostas próximas de uma distribuição Normal.

Executando a função *summary* obtemos as medidas de diagnóstico e as estimativas do modelo. As opções *fit.measures*, *standardized* e *rsquare* adicionam mais medidas de validação que não são padrão no *lavaan*, o ajuste do modelo com as variáveis padronizadas e medidas de R² para as variáveis indicadoras, respectivamente. Por fim, o comando *fitMeasures* inclui após os resultados do *summary* outras medidas de validação que o *lavaan* calcula, mas não exibe dentro do *summary*. A execução desse código gera os seguintes resultados apresentados nas figuras 9 a 14.

lavaan (0.5-20) converged normally after 83 iterations			
Number of observations		1141	
Estimator		WLS	
Minimum Function Test Statistic		603.141	
Degrees of freedom		137	
P-value (Chi-square)		0.000	
Model test baseline model:			A
Minimum Function Test Statistic		1810.051	
Degrees of freedom		171	
P-value		0.000	
User model versus baseline model:			
Comparative Fit Index (CFI)		0.716	
Tucker-Lewis Index (TLI)		0.645	B
Root Mean Square Error of Approximation:			
RMSEA		0.055	
90 Percent Confidence Interval	0.050	0.059	C
P-value RMSEA <= 0.05		0.042	
Standardized Root Mean Square Residual:			
SRMR		0.079	D
Parameter Estimates:			
Information		Expected	
Standard Errors		Standard	

Figura 9- Medidas de ajuste da primeira AFC

Latent Variables:	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB =~						
CxB_1	1.000				1.345	0.822
CxB_2	1.068	0.038	28.120	0.000	1.437	0.816
CxB_3	0.988	0.030	33.231	0.000	1.329	0.770
Emprg =~						
Emprg_1	1.000				1.414	0.882
Emprg_2	1.057	0.022	47.450	0.000	1.495	0.892
Emprg_3	0.992	0.022	44.646	0.000	1.403	0.894
Infra =~						
Infra_1	1.000				1.457	0.841
Infra_2	0.871	0.028	30.898	0.000	1.269	0.734
Infra_3	0.920	0.027	34.503	0.000	1.341	0.789
Local =~						
Local_1	1.000				1.104	0.898
Local_2	0.597	0.063	9.510	0.000	0.660	0.643
Q_Ens =~						
Q_Ens_1	1.000				1.100	0.759
Q_Ens_2	1.370	0.040	33.843	0.000	1.507	0.922
Q_Ens_3	0.906	0.038	23.690	0.000	0.997	0.679
Q_Ens_4	1.095	0.040	27.335	0.000	1.204	0.801
Trad =~						
Trad_1	1.000				1.430	0.815
Trad_2	0.970	0.026	37.140	0.000	1.387	0.821
Trad_3	1.010	0.026	38.233	0.000	1.444	0.837
Trad_4	1.026	0.025	41.478	0.000	1.467	0.857

Figura 10 - Escores fatoriais da primeira AFC

Covariances:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
CxB ~						
Emprg	1.523	0.077	19.836	0.000	0.800	0.800
Infra	1.552	0.076	20.320	0.000	0.792	0.792
Local	0.617	0.058	10.607	0.000	0.415	0.415
Q_Ens	1.146	0.067	17.066	0.000	0.774	0.774
Trad	1.567	0.077	20.393	0.000	0.815	0.815
Emprg ~						
Infra	1.854	0.080	23.062	0.000	0.900	0.900
Local	0.548	0.063	8.694	0.000	0.351	0.351
Q_Ens	1.510	0.079	19.220	0.000	0.970	0.970
Trad	1.983	0.085	23.290	0.000	0.981	0.981
Infra ~						
Local	0.545	0.061	8.950	0.000	0.339	0.339
Q_Ens	1.459	0.073	20.001	0.000	0.911	0.911
Trad	1.958	0.081	24.291	0.000	0.940	0.940
Local ~						
Q_Ens	0.444	0.056	7.857	0.000	0.365	0.365
Trad	0.545	0.061	8.887	0.000	0.345	0.345
Q_Ens ~						
Trad	1.506	0.077	19.580	0.000	0.958	0.958

Figura 11 - Medidas de covariância/correlação da primeira AFC

Variances:						
	Estimate	Std.Err	Z-value	P(> z)	Std.lv	std.all
CxB_1	0.867	0.070	12.345	0.000	0.867	0.324
CxB_2	1.040	0.081	12.855	0.000	1.040	0.335
CxB_3	1.213	0.083	14.611	0.000	1.213	0.407
Emprg_1	0.572	0.048	11.843	0.000	0.572	0.222
Emprg_2	0.574	0.052	11.125	0.000	0.574	0.204
Emprg_3	0.497	0.035	14.396	0.000	0.497	0.201
Infra_1	0.881	0.065	13.473	0.000	0.881	0.293
Infra_2	1.377	0.083	16.636	0.000	1.377	0.461
Infra_3	1.091	0.066	16.451	0.000	1.091	0.378
Local_1	0.291	0.129	2.258	0.024	0.291	0.193
Local_2	0.619	0.059	10.400	0.000	0.619	0.587
Q_Ens_1	0.890	0.051	17.351	0.000	0.890	0.424
Q_Ens_2	0.398	0.037	10.624	0.000	0.398	0.149
Q_Ens_3	1.165	0.064	18.264	0.000	1.165	0.540
Q_Ens_4	0.809	0.057	14.190	0.000	0.809	0.358
Trad_1	1.032	0.066	15.550	0.000	1.032	0.336
Trad_2	0.931	0.059	15.858	0.000	0.931	0.326
Trad_3	0.892	0.064	13.949	0.000	0.892	0.300
Trad_4	0.776	0.053	14.564	0.000	0.776	0.265
CxB	1.810	0.103	17.575	0.000	1.000	1.000
Emprg	2.001	0.097	20.686	0.000	1.000	1.000
Infra	2.122	0.102	20.777	0.000	1.000	1.000
Local	1.219	0.157	7.766	0.000	1.000	1.000
Q_Ens	1.210	0.088	13.717	0.000	1.000	1.000
Trad	2.044	0.101	20.171	0.000	1.000	1.000

Figura 12 - Medidas de variância da primeira AFC

R-Square:	Estimate
CxB_1	0.676
CxB_2	0.665
CxB_3	0.593
Emprg_1	0.778
Emprg_2	0.796
Emprg_3	0.799
Infra_1	0.707
Infra_2	0.539
Infra_3	0.622
Local_1	0.807
Local_2	0.413
Q_Ens_1	0.576
Q_Ens_2	0.851
Q_Ens_3	0.460
Q_Ens_4	0.642
Trad_1	0.664
Trad_2	0.674
Trad_3	0.700
Trad_4	0.735

Figura 13 - R² das variáveis indicadoras na primeira AFC

```
> fitMeasures(cfa_GSP, c("gfi", "agfi"))
gfi agfi
0.883 0.838
```

Figura 14 - Medidas GFI e AGFI da primeira AFC

Consideraremos agora a validação do modelo. Na parte A da figura 9 estão localizados os resultados dos testes baseados na estatística qui-quadrado. Logo, rejeitamos a hipótese nula do primeiro teste qui-quadrado que testa as matrizes de covariância, ou seja, conforme o teste, a matriz de covariância estimada pelo modelo difere da gerada pela amostra. Porém, com tamanho de amostra de 1141 casos, o teste se torna muito sensível e pode detectar diferenças significativas onde elas não existem, devido a isso, necessita-se muito cuidado ao interpretar o resultado do teste com tamanhos de amostras grandes. Também é rejeitado o segundo teste que compara o modelo ajustado com o modelo nulo, então concluímos que o atual modelo é diferente do modelo sem variáveis.

As outras medidas de ajuste não convergem a uma decisão de que o modelo é adequado. Os índices CFI e NNFI (O *lavaan* chama-o pelo seu sinônimo, *Tucker-Lewis Index*), vistos em B na figura 9, possuem valores 0,716 e 0,645 respectivamente e estão longe do valor ideal indicado pela literatura (0,90). A medida RMSEA, apresentada em C, possui valor 0,055 e está um pouco acima do valor ideal (0,05) na medida pontual. O RMSR (No *lavaan* chamado de SRMR), demarcado pela letra D, está acima do valor ótimo (0,05) com um valor igual a 0,079. Finalmente, as medidas GFI e AGFI, que se encontram na figura 14, estão

menores do que o valor considerado aceitável (0,90) estando com valores 0,883 e 0,838 respectivamente.

Avaliando os resultados da seção *Latent Variables* na figura 10, a coluna *Estimates* mostra o score fatorial de cada variável indicadora na respectiva variável latente considerando as variáveis não padronizadas. As colunas *Std.lv* e *Std.all* são frutos da opção *standardized=TRUE*. Na primeira coluna as estimativas são considerando que somente as variáveis latentes foram padronizadas e na segunda coluna considera que todas as variáveis foram padronizadas. Concluímos desses resultados que todas as variáveis indicadoras são altamente significativas ($p < 0,001$) para medir as variáveis latentes. Também concluímos que todas as variáveis indicadoras contribuem de forma similar na manifestação de seus respectivos construtos com exceção da variável “*Fica próxima de estações de transporte público*” (Local_2), pois seu coeficiente foi muito menor do que da variável Local_1.

Na parte *Covariances* localizada na figura 11, a coluna *Estimates* mostra a covariância entre as variáveis latentes, enquanto as colunas *Std.lv* e *Std.all* exibem as correlações entre elas. Notam-se valores de correlação muito altos (maior que 0,9) entre diversas variáveis latentes. Assim como outros métodos de modelagem de dados, a MEE também pode sofrer de multicolineariedade. Com essa nova evidência e considerando os resultados das medidas de validação, concluímos que podemos tentar estimar um modelo mais parcimonioso e estável ao retirar uma das variáveis altamente correlacionadas.

Por fim, as seções *Variances* e *R-Square* encontradas nas figuras 12 e 13, são complementares e demonstram o total de variância não explicada e percentual de variância explicada respectivamente. Ao contrário do esperado, os valores da coluna *Estimates* (Figura 12) não são os valores estimados de variância das variáveis indicadoras e sim o total da variância não explicada em cada uma das variáveis na escala original dos dados. Similarmente, os valores na coluna *Std.all* exibem o percentual de variância não explicada ($1-R^2$). Note que percentuais altos de variâncias não explicadas estão relacionados com valores baixos de escore fatoriais e vice-versa. Em geral, nesse caso, não existem valores preocupantes.

Com essas conclusões, um novo modelo fatorial mais parcimonioso irá ser construído com o objetivo de obter medidas de validação melhores e retirar a multicolineariedade. Decidiu-se por retirar a variável latente *Tradição*. Escolheu-se essa variável latente porque julgou-se ser a com menor relação teórica com a

satisfação. Embora *Tradição* seja um fator importante no processo decisório do aluno, ou seja, quando ele escolhe qual IES estudar, em geral possui pouca importância na satisfação do aluno, pois não é algo com forte presença no dia a dia.

Para rodar as seguintes análises, foi excluído do código inicial a linha referente a *Tradição* no modelo. Os resultados são apresentados nas figuras 15 a 20.

lavaan (0.5-20) converged normally after 62 iterations			
Number of observations		1141	
Estimator		WLS	
Minimum Function Test Statistic		378.423	
Degrees of freedom		80	
P-value (Chi-square)		0.000	
Model test baseline model:			A
Minimum Function Test Statistic		1429.897	
Degrees of freedom		105	
P-value		0.000	
User model versus baseline model:			
Comparative Fit Index (CFI)		0.775	B
Tucker-Lewis Index (TLI)		0.704	
Root Mean Square Error of Approximation:			
RMSEA		0.057	
90 Percent Confidence Interval	0.051	0.063	C
P-value RMSEA <= 0.05		0.020	
Standardized Root Mean Square Residual:			
SRMR		0.084	D
Parameter Estimates:			
Information		Expected	
Standard Errors		Standard	

Figura 15 - Medidas de validação do modelo de AFC sem *Tradição*

Latent Variables:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
CxB ~						
CxB_1	1.000				1.398	0.830
CxB_2	1.007	0.038	26.705	0.000	1.409	0.787
CxB_3	0.979	0.030	32.231	0.000	1.369	0.783
Emprg ~						
Emprg_1	1.000				1.427	0.883
Emprg_2	1.031	0.024	42.742	0.000	1.471	0.873
Emprg_3	0.965	0.024	40.366	0.000	1.377	0.881
Infra ~						
Infra_1	1.000				1.361	0.793
Infra_2	0.904	0.034	26.541	0.000	1.231	0.706
Infra_3	0.959	0.032	29.570	0.000	1.305	0.756
Local ~						
Local_1	1.000				1.111	0.867
Local_2	0.555	0.067	8.315	0.000	0.617	0.605
Q_Ens ~						
Q_Ens_1	1.000				1.115	0.753
Q_Ens_2	1.352	0.043	31.226	0.000	1.507	0.917
Q_Ens_3	0.889	0.041	21.821	0.000	0.991	0.662
Q_Ens_4	1.041	0.042	24.979	0.000	1.160	0.759

Figura 16 - Escores fatoriais no modelo AFC sem *Tradição*

Covariances:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
CxB ~						
Emprg	1.506	0.080	18.777	0.000	0.755	0.755
Infra	1.474	0.080	18.420	0.000	0.775	0.775
Local	0.631	0.063	10.039	0.000	0.406	0.406
Q_Ens	1.148	0.072	15.922	0.000	0.737	0.737
Emprg ~						
Infra	1.721	0.083	20.708	0.000	0.886	0.886
Local	0.479	0.067	7.138	0.000	0.302	0.302
Q_Ens	1.511	0.082	18.425	0.000	0.949	0.949
Infra ~						
Local	0.509	0.063	8.065	0.000	0.337	0.337
Q_Ens	1.352	0.076	17.887	0.000	0.891	0.891
Local ~						
Q_Ens	0.395	0.061	6.499	0.000	0.319	0.319

Figura 17 - Medidas de covariância/correlação no modelo AFC sem *Tradição*

Variances:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
CxB_1	0.885	0.075	11.847	0.000	0.885	0.312
CxB_2	1.219	0.091	13.422	0.000	1.219	0.381
CxB_3	1.185	0.087	13.555	0.000	1.185	0.387
Emprg_1	0.578	0.053	10.888	0.000	0.578	0.221
Emprg_2	0.674	0.058	11.671	0.000	0.674	0.237
Emprg_3	0.549	0.042	13.214	0.000	0.549	0.225
Infra_1	1.095	0.075	14.593	0.000	1.095	0.372
Infra_2	1.523	0.089	17.179	0.000	1.523	0.501
Infra_3	1.275	0.074	17.210	0.000	1.275	0.428
Local_1	0.409	0.146	2.797	0.005	0.409	0.249
Local_2	0.658	0.062	10.564	0.000	0.658	0.634
Q_Ens_1	0.948	0.054	17.432	0.000	0.948	0.433
Q_Ens_2	0.428	0.041	10.326	0.000	0.428	0.159
Q_Ens_3	1.257	0.069	18.271	0.000	1.257	0.561
Q_Ens_4	0.989	0.064	15.495	0.000	0.989	0.424
CxB	1.955	0.111	17.578	0.000	1.000	1.000
Emprg	2.037	0.102	20.060	0.000	1.000	1.000
Infra	1.852	0.106	17.460	0.000	1.000	1.000
Local	1.234	0.174	7.100	0.000	1.000	1.000
Q_Ens	1.243	0.095	13.123	0.000	1.000	1.000

Figura 18 - Medidas de variância do modelo AFC sem *Tradição*

R-Square:	
	Estimate
CxB_1	0.688
CxB_2	0.619
CxB_3	0.613
Emprg_1	0.779
Emprg_2	0.763
Emprg_3	0.775
Infra_1	0.628
Infra_2	0.499
Infra_3	0.572
Local_1	0.751
Local_2	0.366
Q_Ens_1	0.567
Q_Ens_2	0.841
Q_Ens_3	0.439
Q_Ens_4	0.576

Figura 19- R² das variáveis indicadoras no modelo AFC sem *Tradição*

```
> fitMeasures(cfa_GSP2, c("gfi", "agfi"))
gfi agfi
0.915 0.873
```

Figura 20 - Medidas GFI e AGFI do modelo AFC sem *Tradição*

As medidas de validação do novo modelo de AFC mostradas na figura 15 ainda rejeitam as hipóteses nulas do teste qui-quadrado e de comparação com o modelo nulo. No entanto, ocorreram melhorias nos índices CFI e NNFI que aumentaram de 0,716 e 0,645 para 0,775 e 0,704, respectivamente. Porém, esses valores ainda se encontram longe do valor ótimo. Similar foram as melhorias nos índices GFI e AGFI que aumentaram de 0,883 e 0,838 para 0,915 e 0,873. Finalmente aconteceram aumentos nos índices RMSEA e RMSR, sendo que aumentaram de 0,055 e 0,079 para 0,057 e 0,084. Esses aumentos enquanto esperados, não causam preocupação.

Constatamos também que ainda existem medidas de correlação altas como demonstrado na figura 16. Conseqüentemente, ainda existe multicolineariedade, porém com a construção desse modelo já verificamos evidências de uma melhora no ajuste ao se retirar uma das variáveis latentes altamente correlacionadas. Prosseguimos com a exclusão de variáveis na esperança de construir um melhor modelo.

No próximo passo será retirado o construto de Empregabilidade. A razão é similar à da retirada da variável *Tradição*. O fato de ser um fator muito mais relacionado com o processo decisório do aluno do que com sua satisfação. Nas figuras 21 a 26 os resultados do terceiro modelo fatorial.

Tavaan (0.5-20) converged normally after 52 iterations			
Number of observations		1141	
Estimator		WLS	
Minimum Function Test Statistic		281.333	
Degrees of freedom		48	
P-value (Chi-square)		0.000	
Model test baseline model:			A
Minimum Function Test Statistic		1242.454	
Degrees of freedom		66	
P-value		0.000	
User model versus baseline model:			
Comparative Fit Index (CFI)		0.802	B
Tucker-Lewis Index (TLI)		0.727	
Root Mean Square Error of Approximation:			
RMSEA		0.065	C
90 Percent Confidence Interval	0.058	0.073	
P-value RMSEA <= 0.05		0.000	
Standardized Root Mean Square Residual:			
SRMR		0.080	D
Parameter Estimates:			
Information		Expected	
Standard Errors		Standard	

Figura 21 – Medidas de validação do terceiro modelo de AFC

Latent variables:	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB =~						
CxB_1	1.000				1.431	0.834
CxB_2	0.970	0.037	26.183	0.000	1.387	0.774
CxB_3	0.977	0.031	31.950	0.000	1.398	0.784
Infra =~						
Infra_1	1.000				1.340	0.781
Infra_2	0.934	0.038	24.659	0.000	1.250	0.714
Infra_3	0.978	0.036	27.542	0.000	1.310	0.749
Local =~						
Local_1	1.000				1.041	0.810
Local_2	0.664	0.079	8.351	0.000	0.691	0.626
Q_Ens =~						
Q_Ens_1	1.000				1.142	0.754
Q_Ens_2	1.326	0.044	30.282	0.000	1.515	0.915
Q_Ens_3	0.874	0.041	21.442	0.000	0.998	0.657
Q_Ens_4	1.024	0.043	24.061	0.000	1.169	0.740

Figura 22 – Escores fatoriais do terceiro modelo de AFC

Covariances:						
	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB ~						
Infra	1.479	0.082	17.974	0.000	0.772	0.772
Local	0.577	0.066	8.771	0.000	0.388	0.388
Q_Ens	1.164	0.075	15.562	0.000	0.712	0.712
Infra ~						
Local	0.453	0.064	7.109	0.000	0.325	0.325
Q_Ens	1.338	0.077	17.373	0.000	0.875	0.875
Local ~						
Q_Ens	0.415	0.064	6.511	0.000	0.349	0.349

Figura 23 - Medidas de covariância/correlação do terceiro modelo de AFC

Variances:						
	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB_1	0.897	0.078	11.568	0.000	0.897	0.305
CxB_2	1.289	0.095	13.579	0.000	1.289	0.401
CxB_3	1.227	0.091	13.557	0.000	1.227	0.386
Infra_1	1.147	0.082	14.017	0.000	1.147	0.390
Infra_2	1.503	0.096	15.589	0.000	1.503	0.490
Infra_3	1.342	0.077	17.419	0.000	1.342	0.439
Local_1	0.567	0.134	4.240	0.000	0.567	0.343
Local_2	0.742	0.067	11.138	0.000	0.742	0.609
Q_Ens_1	0.993	0.057	17.441	0.000	0.993	0.432
Q_Ens_2	0.444	0.049	9.013	0.000	0.444	0.162
Q_Ens_3	1.315	0.072	18.357	0.000	1.315	0.569
Q_Ens_4	1.132	0.072	15.631	0.000	1.132	0.453
CxB	2.047	0.115	17.787	0.000	1.000	1.000
Infra	1.794	0.109	16.409	0.000	1.000	1.000
Local	1.083	0.164	6.622	0.000	1.000	1.000
Q_Ens	1.305	0.099	13.170	0.000	1.000	1.000

Figura 24 - Medidas de variância do terceiro modelo de AFC

R-Square:	
	Estimate
CxB_1	0.695
CxB_2	0.599
CxB_3	0.614
Infra_1	0.610
Infra_2	0.510
Infra_3	0.561
Local_1	0.657
Local_2	0.391
Q_Ens_1	0.568
Q_Ens_2	0.838
Q_Ens_3	0.431
Q_Ens_4	0.547

Figura 25 - R² das variáveis indicadoras do terceiro modelo de AFC

```
> fitMeasures(cfa_GSP3, c("gfi", "agfi"))
gfi agfi
0.932 0.889
```

Figura 26 - Medidas GFI e AGFI do terceiro modelo de AFC

Em comum com o segundo modelo, ainda rejeitamos a hipótese nula para o teste qui-quadrado e de comparação com o modelo nulo. Os índices CFI e NNFI aumentaram de 0,775 e 0,704 para 0,802 e 0,727. As medidas GFI e AGFI sofreram um acréscimo mudando seu valor de 0,915 e 0,873 para 0,932 e 0,889, ou seja, no

limite do valor ideal para essas medidas. O RMSEA aumentou de 0,057 para 0,065 e o RMSR diminuiu de 0,084 para 0,80.

Concluimos que as medidas de validação do modelo não convergem para uma decisão unânime. Porém, as reduções do modelo fatorial obtiveram sucesso em obter um modelo melhor ajustado. Apesar de nem todas as medidas atingirem os valores indicados pela literatura, aconteceram melhoras substanciais nos seus valores. Pensando assim, consideraremos o instrumento de medida com esses quatro fatores como “bom o suficiente para a construção do modelo de equações estruturais”. Além disso, avaliando a relação entre os quatro fatores e a variável resposta no ambiente hipotético, existem relações, pelo menos teóricas, entre cada uma das variáveis latentes (Custo-benefício, Localização, Infraestrutura e Qualidade de Ensino) com a satisfação do aluno.

O aluno da IES parece ser propenso a se sentir infeliz com a instituição que estuda se ele sentir que o preço que está pagando não vale a experiência de aprendizado que ele está recebendo. De forma similar, o aluno ficará frustrado se sentir que o percurso até o local de estudo é demorado ou de difícil acesso. É plausível, também, pensar que a infraestrutura é um determinante da satisfação do aluno já que se acredita que o espaço físico de estudo em bom estado contribui com o bem-estar do aluno e com a qualidade de ensino. Similarmente, o aluno pode considerar que um local com boa infraestrutura vale o preço que ele paga. Isso daria sentido aos altos valores de correlação encontrados entre os construtos de Infraestrutura, Custo-Benefício e Qualidade de Ensino. Por fim, e talvez mais importante, o aluno não continuará matriculado se sentir que a qualidade do ensino não é boa o suficiente.

Para finalizar essa parte, a tabela 2 apresenta, em resumo, a evolução das medidas de validação ao excluir cada um dos construtos. A tabela 3 mostra a comparação dos modelos obtidos com os outros dois bancos de dados (Brasília e São Paulo), indicando que, ao se retirar as mesmas variáveis latentes retiradas no modelo da IES Metrop. SP, obtém-se modelos melhores nas três pesquisas.

Tabela 2 – Evolução das medidas de validação da AFC ao retirar construtos*

Medida de Validação	AFC Completa	AFC sem <i>Tradição</i>	AFC sem <i>Tradição e Empregabilidade</i>
CFI	0,716	0,775	0,802
NNFI	0,645	0,704	0,727
GFI	0,883	0,915	0,932
AGFI	0,838	0,873	0,889
RMSEA	0,055	0,057	0,065
RMSR	0,079	0,084	0,080

*Hipóteses nulas do teste qui-quadrado e do teste sobre o modelo nulo sempre são rejeitadas com $p < 0,001$

Tabela 3 – Comportamento do modelo AFC antes e depois da retirada dos fatores para as bases de dados de São Paulo e Brasília*

Medida de Validação	SP AFC Completa	SP AFC sem <i>Tradição e Empregabilidade</i>	BSB AFC Completa	BSB AFC sem <i>Tradição e Empregabilidade</i>
CFI	0,774	0,858	0,796	0,851
NNFI	0,718	0,804	0,745	0,795
GFI	0,916	0,951	0,914	0,942
AGFI	0,884	0,921	0,880	0,906
RMSEA	0,048	0,059	0,042	0,056
RMSR	0,077	0,071	0,088	0,073

* Hipóteses nulas do teste qui-quadrado e do teste sobre o modelo nulo sempre são rejeitadas com $p < 0,001$

3.4.2 Modelo de equações estruturais

Levando em consideração os resultados da AFC, será agora construído o modelo de equações estruturais. O novo diagrama de caminhos considerando os resultados da seção 4.3.1 é mostrado na figura 27. Nessa parte do trabalho, similar a seção anterior, será mostrado o código do R somente para o primeiro modelo e utilizaremos novamente como exemplo o banco de dados da IES localizada na Metrop. SP, porém, dessa vez, serão mostrados os resultados para todos os três bancos de dados.

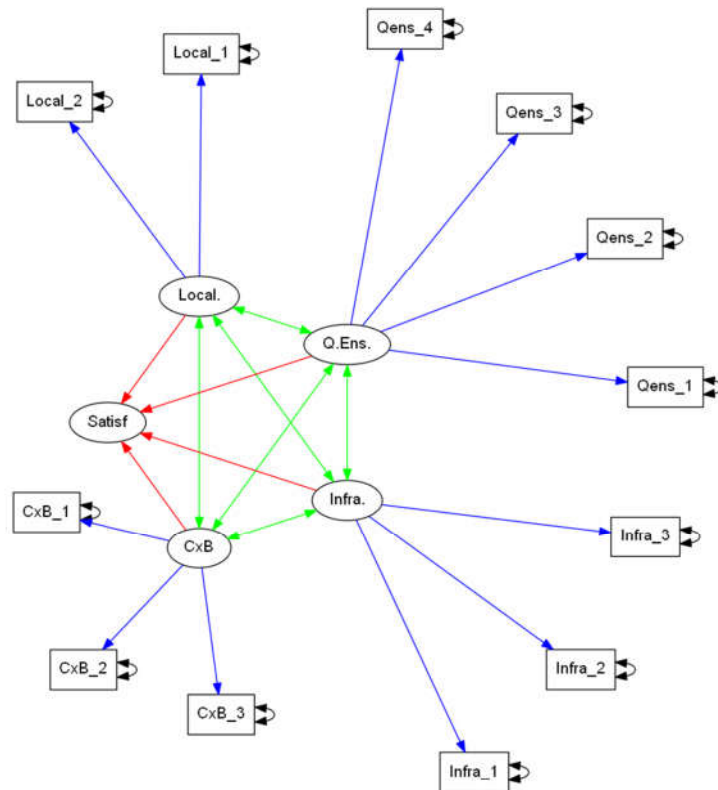


Figura 27 - Novo modelo de equações estruturais para explicar a satisfação geral.

O código para especificar e estimar um modelo de equações estruturais no *lavaan*, apresentado na figura 28, não difere muito do código para realizar a AFC. Inclui-se somente uma linha no modelo especificado indicando quais variáveis compõem a satisfação e altera-se a função de *cfa* para *sem*. Os resultados para o primeiro modelo estão nas figuras 29 a 35.

```

modelo_sem='CxB =~ CxB_1 + CxB_2 + CxB_3
Infra =~ Infra_1 + Infra_2 + Infra_3
Local =~ Local_1 + Local_2
Q_Ens =~ Q_Ens_1 + Q_Ens_2 + Q_Ens_3 + Q_Ens_4
SATIS_GERAL ~ CxB + Infra + Local + Q_Ens'

sem_GSP=sem(modelo_sem, estimator="WLS", data=GSP)
summary(sem_GSP, fit.measures=TRUE, standardized=TRUE, rsquare=TRUE)

```

Figura 28 - Código do R e funções do pacote *lavaan* para realizar a MEE

lavaan (0.5-20) converged normally after 56 iterations	
Number of observations	1141
Estimator	WLS
Minimum Function Test Statistic	312.291
Degrees of freedom	56
P-value (Chi-square)	0.000
Model test baseline model:	A
Minimum Function Test Statistic	1298.045
Degrees of freedom	78
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.790
Tucker-Lewis Index (TLI)	0.707
	B
Root Mean Square Error of Approximation:	
RMSEA	0.063
90 Percent Confidence Interval	0.057 0.070
P-value RMSEA <= 0.05	0.001
	C
Standardized Root Mean Square Residual:	
SRMR	0.078
	D
Parameter Estimates:	
Information	Expected
Standard Errors	Standard

Figura 29 - Medidas de validação do modelo da Metrop. SP

Latent variables:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
CxB =~						
CxB_1	1.000				1.414	0.832
CxB_2	0.966	0.038	25.740	0.000	1.366	0.763
CxB_3	0.984	0.031	31.669	0.000	1.392	0.784
Infra =~						
Infra_1	1.000				1.329	0.780
Infra_2	0.951	0.038	25.124	0.000	1.263	0.721
Infra_3	0.993	0.036	27.817	0.000	1.320	0.757
Local =~						
Local_1	1.000				1.020	0.803
Local_2	0.669	0.079	8.428	0.000	0.683	0.624
Q_Ens =~						
Q_Ens_1	1.000				1.141	0.761
Q_Ens_2	1.320	0.043	30.498	0.000	1.507	0.916
Q_Ens_3	0.878	0.041	21.530	0.000	1.002	0.661
Q_Ens_4	1.009	0.042	23.914	0.000	1.151	0.730

Figura 30 - Escores fatoriais para o modelo da Metrop. SP

Regressions:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
SATIS_GERAL ~						
CxB	-0.049	0.037	-1.329	0.184	-0.070	-0.077
Infra	0.243	0.068	3.576	0.000	0.323	0.355
Local	-0.046	0.035	-1.328	0.184	-0.047	-0.052
Q_Ens	0.353	0.067	5.235	0.000	0.403	0.443

Figura 31 – Coeficientes estimados do impacto de cada variável latente na satisfação geral do aluno para o modelo da Metrop. SP

Covariances:						
	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB ~						
Infra	1.439	0.081	17.673	0.000	0.766	0.766
Local	0.575	0.066	8.758	0.000	0.398	0.398
Q_Ens	1.140	0.074	15.347	0.000	0.706	0.706
Infra ~						
Local	0.425	0.062	6.807	0.000	0.314	0.314
Q_Ens	1.321	0.076	17.356	0.000	0.872	0.872
Local ~						
Q_Ens	0.411	0.064	6.477	0.000	0.353	0.353

Figura 32 - Medidas de covariância/correlação para o modelo da Metrop. SP

Variances:						
	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB_1	0.892	0.077	11.557	0.000	0.892	0.309
CxB_2	1.339	0.094	14.288	0.000	1.339	0.418
CxB_3	1.216	0.091	13.396	0.000	1.216	0.386
Infra_1	1.137	0.081	14.051	0.000	1.137	0.392
Infra_2	1.477	0.096	15.415	0.000	1.477	0.481
Infra_3	1.298	0.076	17.162	0.000	1.298	0.427
Local_1	0.573	0.129	4.443	0.000	0.573	0.355
Local_2	0.733	0.066	11.182	0.000	0.733	0.611
Q_Ens_1	0.944	0.055	17.250	0.000	0.944	0.420
Q_Ens_2	0.434	0.045	9.718	0.000	0.434	0.160
Q_Ens_3	1.296	0.071	18.180	0.000	1.296	0.563
Q_Ens_4	1.164	0.071	16.285	0.000	1.164	0.468
SATIS_GERAL	0.423	0.022	18.849	0.000	0.423	0.510
CxB	2.000	0.114	17.540	0.000	1.000	1.000
Infra	1.765	0.108	16.357	0.000	1.000	1.000
Local	1.041	0.159	6.544	0.000	1.000	1.000
Q_Ens	1.302	0.099	13.168	0.000	1.000	1.000

Figura 33 - Medidas de variância para o modelo da Metrop. SP

R-Square:	
	Estimate
CxB_1	0.691
CxB_2	0.582
CxB_3	0.614
Infra_1	0.608
Infra_2	0.519
Infra_3	0.573
Local_1	0.645
Local_2	0.389
Q_Ens_1	0.580
Q_Ens_2	0.840
Q_Ens_3	0.437
Q_Ens_4	0.532
SATIS_GERAL	0.490

Figura 34 - R² das variáveis indicadoras para o modelo da Metrop. SP

```
> fitMeasures(sem_GSP3, c("gfi", "agfi"))
gfi agfi
0.927 0.881
```

Figura 35 - Medidas GFI e AGFI para o modelo da Metrop. SP

Considerando os resultados da figura 29, os testes qui-quadrado e de comparação com o modelo nulo rejeitam a hipótese nula. As medidas CFI e NNFI possuem valores iguais a 0,790 e 0,707. Os índices RMSEA e RMSR são iguais a 0,063 e 0,078. Finalmente, os valores das medidas GFI e AGFI, presentes na figura

35, estão iguais a 0,927 e 0,881 respectivamente. Esses resultados são esperados dados os resultados do modelo de AFC para a mesma base de dados.

Na figura 30, a relação de cada variável indicadora com sua variável latente é significativa com $p < 0,001$. Esse resultado é importante e está de acordo com o encontrando na AFC.

Cabe ressaltar que a figura 31 apresenta os resultados de um modelo que tem uma variável dependente. Os resultados aqui indicam que não existe um efeito significativo das variáveis latentes *Custo-Benefício* e *Localização* na satisfação do aluno (variável dependente) devido à falta de evidência de que seus coeficientes difiram de zero ($p > 0,05$), o que leva a não rejeição da hipótese nula.

As medidas de covariância e correlação estimadas na figura 32 são similares as encontradas na AFC e possuem alta significância estatística ($p < 0,001$). Importante aqui é o alto nível de associação entre as variáveis *Custo-Benefício* e *Infraestrutura*, *Custo-Benefício* e *Qualidade de Ensino*, e *Qualidade de Ensino* e *Infraestrutura*. Isso não significa que uma influência a outra diretamente, porém, para alunos dessa instituição essas variáveis “movem-se” de forma conjunta. Por exemplo, uma percepção de melhora na qualidade de ensino resulta em um sentimento de melhora no custo-benefício de estudar na IES.

Os resultados das figuras 33 e 34 são mistos. O ideal é que na coluna *Std.lv* e *Std.all* da figura 33, os valores sejam todos menores que 0,5 (Kline, 2011) ou maiores que 0,5 na figura 34. Isso não é visualizado para as variáveis observadas “*Fica próxima de estações de transporte público*” (*Local_2*) e “*Exige esforço do aluno para se formar*” (*Q_Ens_3*). A implicância disso é que existe evidência de que essas variáveis não medem o que nós presumimos que elas medem, ou seja, a variável *Local_2* possivelmente não está relacionada com o construto *Localização*. Esse problema é denominado **validação convergente** (Kline, 2011). Outro problema é referente a variável dependente *Satis_Geral*. O valor de R^2 baixo (0,490) pode ser causado pelo fato de as variáveis consideradas não serem o suficiente para explicar toda a variabilidade na satisfação dos alunos. Em um próximo modelo considerar variáveis adicionais poderá ajudar a elevar esse coeficiente.

Por fim, nas figuras 37 e 38 é mostrado o diagrama de caminhos estimado para as soluções padronizadas e não padronizadas utilizando a função *semPath* do pacote *semPlot* do R. Esse pacote é muito útil pois ele lê com facilidade os objetos criados pelas funções do *lavaan* e possui uma representação gráfica

agradável. Entretanto, com muitos fatores e/ou variáveis indicadoras, os gráficos podem ficar muito agrupados dificultando a leitura. O código está apresentado na figura 36.

```
require(semPlot)
# Padronizados
semPaths(sem_GSP, what = "par", fade = F, layout = "tree2")
# Não padronizados
semPaths(sem_GSP, what = "std", fade = F, layout = "tree2")
```

Figura 36 - Código do R e funções do *semPlot* para gerar os diagramas de caminhos

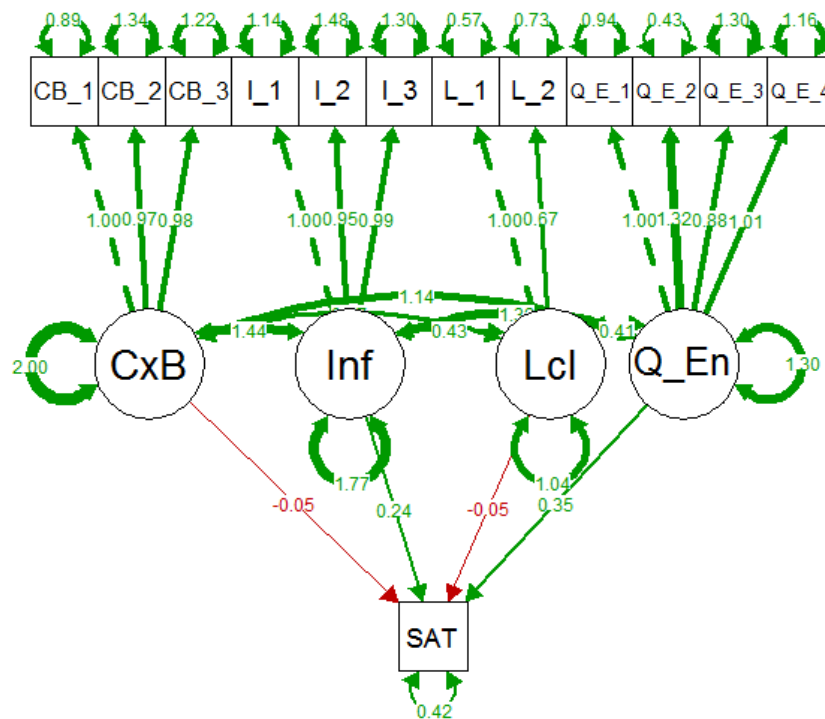


Figura 37 - Gráfico gerado pelo pacote *semPlot* para a solução não padronizada

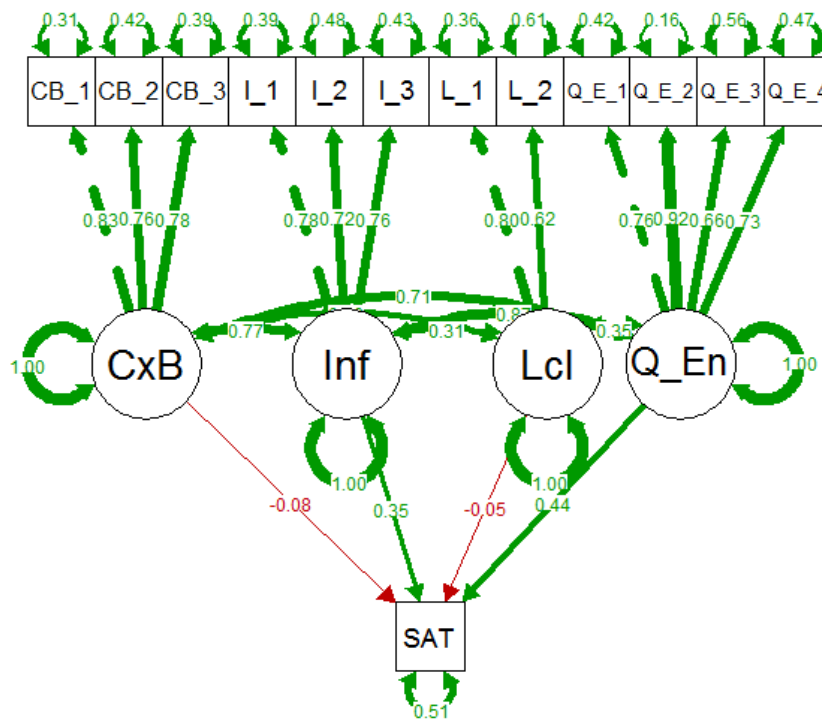


Figura 38 - Gráfico gerado pelo pacote *semPlot* para a solução padronizada

Agora serão considerados os resultados para a IES localizada em Brasília. Os resultados estão nas figuras 39 a 45.

Para esse o modelo com os dados da IES localizada em BSB, os testes continuam tendo suas hipóteses nulas rejeitas e os valores das estatísticas de validação também foram reduzidos levemente. O CFI e o NNFI agora são iguais a 0,843 e 0,782. As medidas RMSEA e RMSR são iguais a 0,054 e 0,072 respectivamente. A medida de ajuste geral GFI é igual a 0,942, enquanto a medida AGFI têm seu valor igual a 0,905.

lavaan (0.5-20) converged normally after 60 iterations	
Number of observations	1785
Estimator	WLS
Minimum Function Test Statistic	345.651
Degrees of freedom	56
P-value (Chi-square)	0.000
Model test baseline model:	A
Minimum Function Test Statistic	1928.557
Degrees of freedom	78
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.843
Tucker-Lewis Index (TLI)	0.782
Root Mean Square Error of Approximation:	
RMSEA	0.054
90 Percent Confidence Interval	0.048 0.059
P-value RMSEA <= 0.05	0.118
Standardized Root Mean Square Residual:	
SRMR	0.072
Parameter Estimates:	
Information	Expected
Standard Errors	Standard

Figura 39 - Medidas de validação do modelo de BSB

Latent variables:	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB =~						
CxB_1	1.000				1.560	0.864
CxB_2	0.776	0.027	29.191	0.000	1.211	0.692
CxB_3	0.980	0.026	37.054	0.000	1.529	0.832
Infra =~						
Infra_1	1.000				1.207	0.767
Infra_2	0.979	0.034	28.746	0.000	1.182	0.680
Infra_3	0.949	0.030	31.205	0.000	1.146	0.725
Local =~						
Local_1	1.000				1.202	0.698
Local_2	1.099	0.066	16.715	0.000	1.321	0.780
Q_Ens =~						
Q_Ens_1	1.000				0.948	0.725
Q_Ens_2	1.237	0.042	29.406	0.000	1.172	0.884
Q_Ens_3	0.896	0.042	21.325	0.000	0.850	0.606
Q_Ens_4	1.027	0.045	23.006	0.000	0.973	0.734

Figura 40 – Escores fatorais do modelo de BSB

Regressions:	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
SATIS_GERAL ~						
CxB	-0.041	0.018	-2.271	0.023	-0.064	-0.074
Infra	0.339	0.058	5.844	0.000	0.410	0.472
Local	-0.021	0.020	-1.009	0.313	-0.025	-0.029
Q_Ens	0.290	0.066	4.418	0.000	0.275	0.317

Figura 41 - Coeficientes estimados do impacto de cada variável latente na satisfação geral do aluno do modelo de BSB

Covariances:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
CxB ~						
Infra	1.229	0.065	18.790	0.000	0.652	0.652
Local	0.850	0.070	12.084	0.000	0.453	0.453
Q_Ens	0.791	0.055	14.374	0.000	0.535	0.535
Infra ~						
Local	0.714	0.061	11.716	0.000	0.492	0.492
Q_Ens	0.985	0.059	16.793	0.000	0.861	0.861
Local ~						
Q_Ens	0.534	0.051	10.402	0.000	0.469	0.469

Figura 42 - Medidas de covariância/correlação do modelo de BSB

Variances:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
CxB_1	0.830	0.073	11.318	0.000	0.830	0.254
CxB_2	1.594	0.080	19.939	0.000	1.594	0.521
CxB_3	1.040	0.076	13.700	0.000	1.040	0.308
Infra_1	1.023	0.053	19.183	0.000	1.023	0.412
Infra_2	1.624	0.079	20.569	0.000	1.624	0.538
Infra_3	1.183	0.057	20.923	0.000	1.183	0.474
Local_1	1.522	0.112	13.599	0.000	1.522	0.513
Local_2	1.122	0.106	10.558	0.000	1.122	0.391
Q_Ens_1	0.810	0.044	18.537	0.000	0.810	0.474
Q_Ens_2	0.382	0.031	12.491	0.000	0.382	0.218
Q_Ens_3	1.241	0.068	18.378	0.000	1.241	0.632
Q_Ens_4	0.812	0.047	17.422	0.000	0.812	0.462
SATIS_GERAL	0.380	0.019	19.720	0.000	0.380	0.504
CxB	2.435	0.101	24.032	0.000	1.000	1.000
Infra	1.457	0.080	18.208	0.000	1.000	1.000
Local	1.445	0.126	11.502	0.000	1.000	1.000
Q_Ens	0.899	0.074	12.208	0.000	1.000	1.000

Figura 43 - Medidas de variância do modelo de BSB

R-Square:	
	Estimate
CxB_1	0.746
CxB_2	0.479
CxB_3	0.692
Infra_1	0.588
Infra_2	0.462
Infra_3	0.526
Local_1	0.487
Local_2	0.609
Q_Ens_1	0.526
Q_Ens_2	0.782
Q_Ens_3	0.368
Q_Ens_4	0.538
SATIS_GERAL	0.496

Figura 44 - R² das variáveis indicadoras do modelo de BSB

```
> fitMeasures(sem_BSB3, c("gfi", "agfi"))
gfi agfi
0.942 0.905
```

Figura 45 - Medidas GFI e AGFI para o modelo de BSB

Igual ao modelo para os dados da IES da região metropolitana de São Paulo, os escores fatoriais de cada variável indicadora na variável latente são altamente significativos ($p < 0,001$) conforme observarmos na figura 40.

Considerando o efeito das variáveis latentes na satisfação geral, vistos na figura 41, somente a variável *Localização* não foi significativa ($p=0,313$), já os demais construtos foram todos significativos ($p<0,05$).

Na figura 42, constatamos que para o modelo da IES de Brasília, somente existe uma correlação considerável entre as variáveis *Infraestrutura* e *Qualidade de Ensino* ($Std.all = 0,861$).

Avaliando as variâncias não explicadas do modelo nas figuras 43 e 44, pode-se observar que a variável Q_Ens_3 sofre de validação convergente, assim como as variáveis “É fácil de chegar” (Local_1), “Oferece as melhores bolsas de estudo” (CxB_2) e “Dispõe de ótimos serviços de atendimento aos alunos” (Infra_2). O mesmo problema de um R^2 baixo para a satisfação geral também acontece nesse modelo.

Os diagramas de caminhos gerados pelo *semPlot* são apresentados nas figuras 46 e 47.

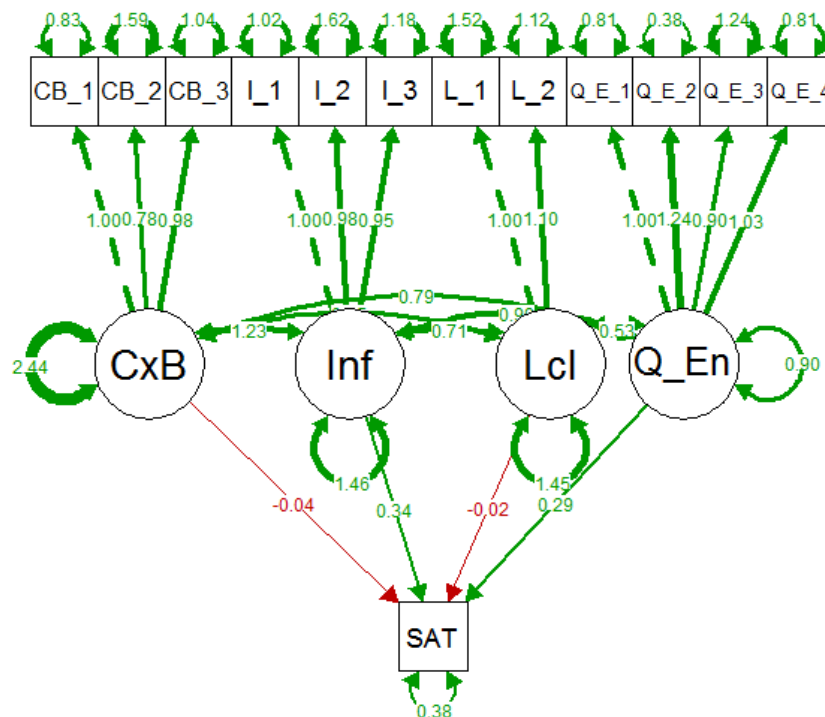


Figura 46 - Diagrama de caminhos para a solução não padronizada utilizando dados da IES de BSB

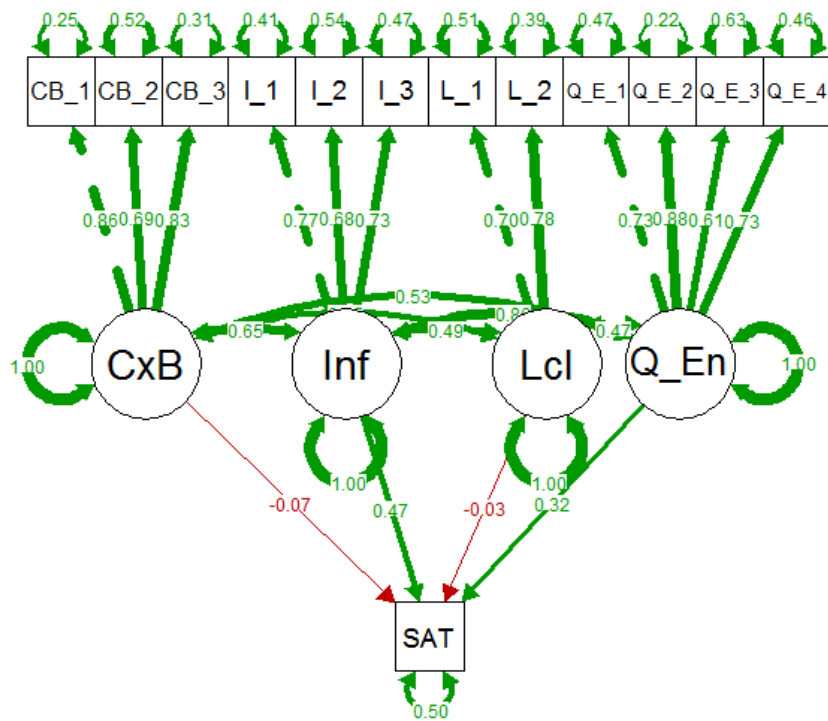


Figura 47 - Diagrama de caminhos para a solução padronizada utilizando dados da IES de BSB

Por fim, será construído o modelo da IES de SP. Os resultados estão nas figuras 48 a 54.

Novamente, como evidenciado pela figura 48, os testes qui-quadrado e de comparação com o modelo nulo são rejeitados e as estatísticas de validação CFI e NNFI são iguais a 0,844 e 0,782 respectivamente. Os valores de RMSEA e RMSR possuem valores iguais a 0,058 e 0,069. Finalmente, as medidas GFI e AGFI são equivalentes a 0,946 e 0,913 respectivamente.

lavaan (0.5-20) converged normally after 54 iterations	
Number of observations	3792
Estimator	WLS
Minimum Function Test Statistic	779.310
Degrees of freedom	56
P-value (Chi-square)	0.000
Model test baseline model:	A
Minimum Function Test Statistic	4705.109
Degrees of freedom	78
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.844
Tucker-Lewis Index (TLI)	0.782
	B
Root Mean Square Error of Approximation:	
RMSEA	0.058
90 Percent Confidence Interval	0.055 0.062
P-value RMSEA <= 0.05	0.000
	C
Standardized Root Mean Square Residual:	
SRMR	0.069
	D
Parameter Estimates:	
Information	Expected
Standard Errors	Standard

Figura 48 - Medidas de validação do modelo de SP

Latent variables:	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB =~						
CxB_1	1.000				1.676	0.884
CxB_2	0.805	0.016	51.010	0.000	1.350	0.760
CxB_3	0.991	0.016	62.248	0.000	1.661	0.857
Infra =~						
Infra_1	1.000				1.301	0.802
Infra_2	0.896	0.021	42.528	0.000	1.165	0.647
Infra_3	0.871	0.020	42.919	0.000	1.133	0.679
Local =~						
Local_1	1.000				0.977	0.626
Local_2	1.432	0.070	20.536	0.000	1.399	0.811
Q_Ens =~						
Q_Ens_1	1.000				1.074	0.741
Q_Ens_2	1.229	0.026	47.253	0.000	1.320	0.873
Q_Ens_3	0.919	0.027	34.052	0.000	0.987	0.644
Q_Ens_4	0.978	0.023	42.278	0.000	1.051	0.737

Figura 49 – Escores fatorais do modelo de SP

Regressions:	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
SATIS_GERAL ~						
CxB	-0.002	0.013	-0.160	0.873	-0.003	-0.004
Infra	0.283	0.030	9.296	0.000	0.369	0.398
Local	-0.068	0.021	-3.296	0.001	-0.067	-0.072
Q_Ens	0.309	0.030	10.175	0.000	0.331	0.358

Figura 50 - Coeficientes estimados do impacto de cada variável latente na satisfação geral do aluno do modelo de SP

Covariances:						
	Estimate	Std.Err	Z-value	P(> z)	std.lv	std.all
CxB ~						
Infra	1.441	0.047	30.464	0.000	0.661	0.661
Local	0.708	0.045	15.852	0.000	0.432	0.432
Q_Ens	1.027	0.042	24.355	0.000	0.571	0.571
Infra ~						
Local	0.672	0.040	16.756	0.000	0.528	0.528
Q_Ens	1.125	0.041	27.506	0.000	0.805	0.805
Local ~						
Q_Ens	0.528	0.036	14.802	0.000	0.503	0.503

Figura 51 - Medidas de covariância/correlação do modelo de SP

Variances:						
	Estimate	Std.Err	Z-value	P(> z)	Std.lv	Std.all
CxB_1	0.789	0.050	15.738	0.000	0.789	0.219
CxB_2	1.334	0.051	26.384	0.000	1.334	0.423
CxB_3	1.001	0.053	18.722	0.000	1.001	0.266
Infra_1	0.937	0.038	24.432	0.000	0.937	0.356
Infra_2	1.883	0.050	37.757	0.000	1.883	0.581
Infra_3	1.499	0.044	34.046	0.000	1.499	0.539
Local_1	1.484	0.063	23.426	0.000	1.484	0.608
Local_2	1.019	0.088	11.578	0.000	1.019	0.342
Q_Ens_1	0.945	0.032	29.350	0.000	0.945	0.450
Q_Ens_2	0.544	0.027	20.001	0.000	0.544	0.238
Q_Ens_3	1.375	0.044	31.023	0.000	1.375	0.586
Q_Ens_4	0.927	0.032	29.095	0.000	0.927	0.456
SATIS_GERAL	0.461	0.015	29.975	0.000	0.461	0.538
CxB	2.808	0.071	39.462	0.000	1.000	1.000
Infra	1.692	0.057	29.653	0.000	1.000	1.000
Local	0.955	0.070	13.564	0.000	1.000	1.000
Q_Ens	1.154	0.054	21.296	0.000	1.000	1.000

Figura 52 - Medidas de variância do modelo de SP

R-Square:	
	Estimate
CxB_1	0.781
CxB_2	0.577
CxB_3	0.734
Infra_1	0.644
Infra_2	0.419
Infra_3	0.461
Local_1	0.392
Local_2	0.658
Q_Ens_1	0.550
Q_Ens_2	0.762
Q_Ens_3	0.414
Q_Ens_4	0.544
SATIS_GERAL	0.462

Figura 53- R² das variáveis indicadoras do modelo de SP

```
> fitMeasures(sem_SP3, c("gfi", "agfi"))
gfi agfi
0.946 0.913
```

Figura 54 - Medidas GFI e AGFI para o modelo de SP

Similar aos outros modelos, os escores das variáveis observadas nas variáveis latentes são altamente significativos ($p < 0,001$), como apresenta a figura 49.

Na figura 50, observamos que para o modelo da IES de SP, somente a variável *Custo-Benefício* não possui relação estatisticamente significativa com a satisfação geral ($p=0,873$).

Considerando as covariâncias e correlações na figura 51, novamente somente a correlação entre os construtos *Infraestrutura* e *Qualidade de Ensino* é considerada muito alta.

Enfim, avaliando a proporção de variância não explicada das variáveis indicadoras nas figuras 52 e 53, esse modelo possui problemas com algumas variáveis em comum com o modelo de BSB, são as variáveis indicadoras *Infra_2*, *Local_1* e *Q_Ens_3*. Além disso, encontrou R^2 menor que 0,5 na variável observada “*Tem ótimas instalações físicas para estudo (salas de aula, bibliotecas, laboratórios, etc)* ” (*Infra_3*). Similarmente aos modelos anteriores, a variável *Satis_Geral* também possui R^2 baixo ($R^2=0,462$).

Similar aos outros modelos, as figuras 55 e 56 apresentam os gráficos gerados pelo R com auxílio do *semPlot*.

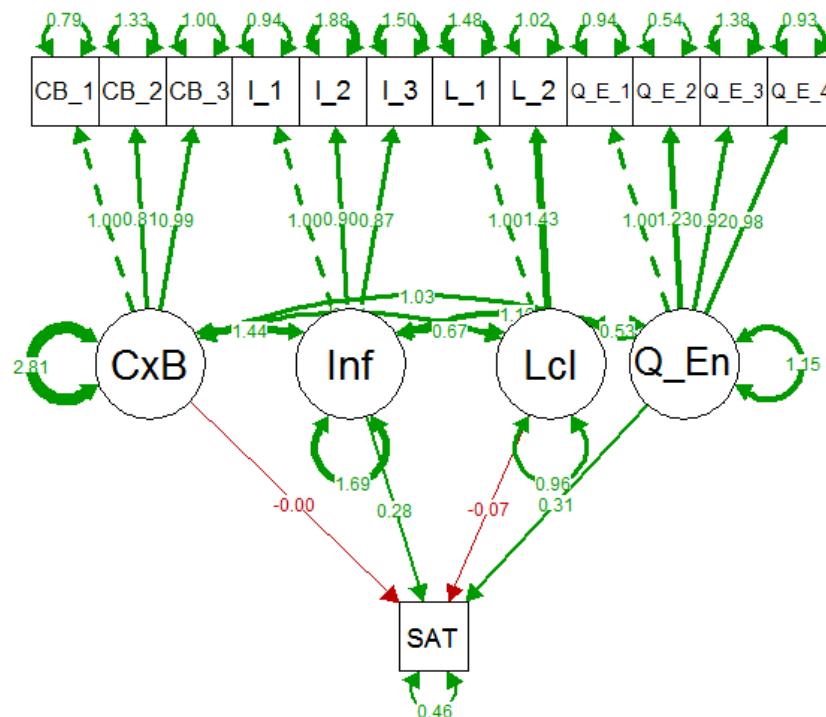


Figura 55 – Diagrama de caminhos gerados pelo pacote *semPlot* para o modelo da IES de SP e com valores não padronizados

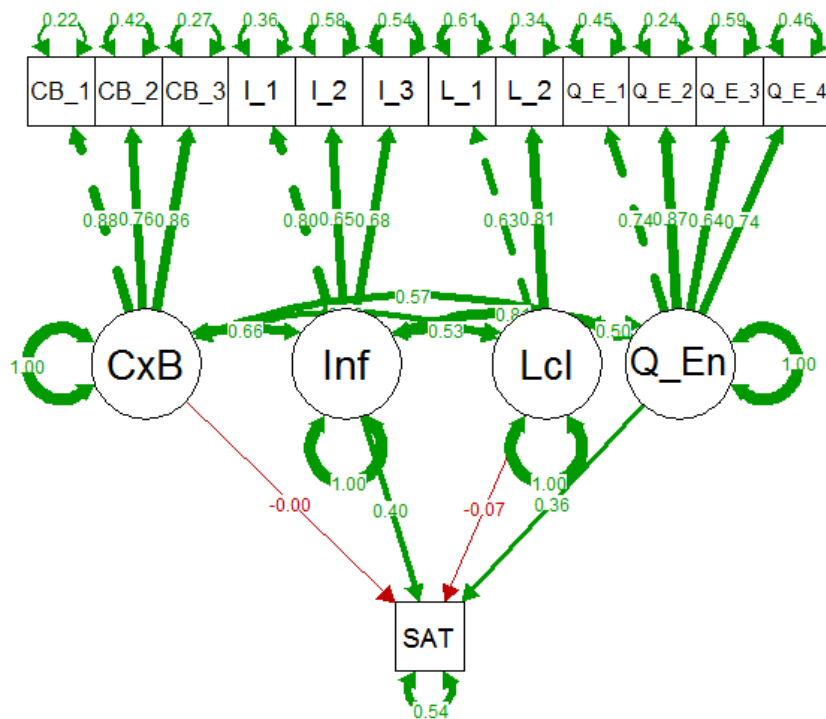


Figura 56 - Diagrama de caminhos gerados pelo pacote *semPlot* para o modelo da IES de SP e com valores padronizados

3.5 Interpretação das estimativas

A interpretação das estimativas de um modelo de equações estruturais depende se os valores estão padronizados ou não (Kline, 2011).

No caso de valores padronizados, interpretamos de forma similar a uma regressão linear. Por exemplo, considerando o modelo construído para a IES de SP, temos que a variável “*Oferece a melhor qualidade de ensino*” (Q_Ens_2) possui um coeficiente estimado de 1,229 com a variável latente *Qualidade de Ensino*. Logo, um aumento de 1 ponto no escore da variável latente causa um acréscimo de 1,229 pontos na variável indicadora. Consideramos o aumento na variável latente pois o modelo construído considera que a variável latente se manifesta pelas variáveis indicadoras consideradas. (Figura 49)

Em outro exemplo, consideraremos o efeito da variável latente *Qualidade de Ensino* na satisfação do aluno no mesmo modelo. Nesse caso, um aumento de 1 unidade no escore da *Qualidade de Ensino*, resulta em um aumento de 0,309 pontos na satisfação geral do aluno.

Considerando variáveis não-padronizadas, os coeficientes são interpretados como correlações. Utilizando os mesmos exemplos, a variável observada Q_Ens_2 possui uma correlação igual a 0,873 (coluna “Std.all” da figura 49) com o construto *Qualidade de Ensino*. Similarmente, o construto *Qualidade de Ensino* possui uma correlação de 0,358 com a satisfação geral do aluno. (Figura 50)

No entanto devemos ter cuidado com essas interpretações já que as variáveis nesse trabalho estão em escala ordinal e essas interpretações consideram que as variáveis estão em escalas contínuas. Na literatura referenciada, é mais comum interpretar somente o sinal e a força dos coeficientes.

3.6 Relato dos resultados

De acordo com o roteiro proposto por Kline (figura 3), nesta seção serão relatados os principais resultados obtidos com a MEE. As tabelas 4 e 5 apresentam resumos dos modelos ajustados para os três bancos de dados.

Tabela 4 - Resumo das medidas de validação dos três modelos de equações estruturais ajustados*

Medida de Validação	Metrop. SP	BSB	SP
CFI	0,79	0,843	0,844
NNFI	0,707	0,782	0,782
GFI	0,827	0,942	0,946
AGFI	0,881	0,905	0,913
RMSEA	0,063	0,054	0,058
RMSR	0,078	0,072	0,069
R ² Satisfação Geral	0,490	0,496	0,462

*Hipóteses nulas do teste qui-quadrado e do teste sobre o modelo nulo sempre são rejeitadas com $p < 0,001$

Tabela 5 - Coeficientes das variáveis latentes regredidas na satisfação geral do aluno

Construto / IES	Metrop. SP	BSB	SP
CxB	-0,049*	-0,041	-0,002*
Infra.	0,243	0,339	0,283
Local.	-0,046*	-0,021*	-0,068
Q.Ens.	0,353	0,290	0,309

*Teste de significância do coeficiente não rejeitado com $p > 0,05$.

Reitera-se pelos valores da Tabela 4 que esse modelo é aceitável. As medidas de validação do modelo estão em valores intermediários, chegando somente próximo aos valores considerados ideais (0,95 para CFI, NNFI, GFI e AGFI, e menor que 0,05 para RMSEA e RMSR). Também, o modelo falha em explicar boas proporções da satisfação do aluno. Isso é justificado pelo baixo R^2 para a variável dependente nos três modelos. Apesar disso, os resultados são coerentes e nenhum valor absurdo foi encontrado.

Vemos pelo tamanho dos coeficientes na tabela 5 que os fatores mais importantes na satisfação do aluno, independente da região e IES que estuda, são a *Infraestrutura* e a *Qualidade de Ensino*. Ademais, contrariando a fase de especificação do modelo, os coeficientes de *Localização* e *Custo-Benefício* são negativos nos três modelos. Contudo, quando os valores são significativos, eles são também muito próximos a zero, indicando que essas variáveis têm pouquíssimo impacto na satisfação do aluno.

Outro resultado interessante é de que a variável *Qualidade de Ensino* não possui o maior coeficiente para os alunos da IES de Brasília onde a percepção de uma melhor *Infraestrutura* é mais importante. Importante lembrar, porém, que existe um alto grau de associação entre essas duas variáveis medida pela correlação, então, podemos interpretar que melhoras em uma variável gera uma percepção de melhoria na outra.

5. Considerações Finais

É importante que fique claro que o presente texto tem o propósito de ser um referencial rápido para que pessoas com certa familiaridade em análise de dados possam consultar e tenham acesso a uma leitura menos detalhista e mais prática e direta sobre a MEE.

Quanto à utilização dos recursos computacionais, uso do R e do *lavaan*, conforme o próprio autor do pacote, o *lavaan* foi criado com o propósito de ser uma ferramenta de fácil utilização para pesquisadores não familiarizados com o programa a partir de comandos intuitivos e simples (Rosseel, 2012). Para as análises realizadas nesse trabalho, a parte mais difícil foi importar para o R as planilhas com os dados. Desconsiderando isso, a ajuda encontrada nas referências de Pereira (2013), Rosseel (2012) e na documentação do próprio R foram suficientes para conseguir os resultados com êxito. Pode-se observar que as formatações dos resultados provenientes do *software* são esteticamente agradáveis, e as informações suficientes, apesar da necessidade de precisar executar um comando a mais para obter medidas de validação auxiliares.

Contrariando a literatura sobre MEE, que indica que as variáveis usadas na modelagem devem ser criadas com o objetivo de uso dessa técnica, as perguntas do questionário usado nesse trabalho não foram construídas com o propósito de construir um modelo de equações estruturais. Consequentemente os dados não são totalmente adequados para a utilização da técnica. Entretanto, por existir a disponibilidade dos dados e adequação “parcial” das perguntas, decidiu-se testar a técnica.

Mas, acredita-se que isso não invalida os resultados encontrados, eles são até motivadores e criam a necessidade de modificar alguns pontos para futuros estudos, entre eles:

- **Modificar a variável dependente do modelo:** Foi possível perceber que as variáveis latentes aqui consideradas têm um impacto na satisfação do aluno, porém elas foram primeiramente concebidas com o propósito de avaliar a imagem da IES. Entretanto, devido à falta de uma pergunta que medisse isso, se decidiu estudar o impacto das cinco variáveis latentes na satisfação do aluno. Uma pergunta no questionário que solicitasse ao respondente a atribuição de uma nota de avaliação geral da IES seria uma melhor variável dependente.

- **Considerar novos atributos:** Independente de qual é a variável dependente do modelo, acredita-se que mais perguntas para medir cada variável latente ajude a aumentar a proporção de variância explicada, tanto da variável de interesse como de cada variável indicadora, pois esse foi um dos problemas encontrados durante a validação do modelo.
- **Considerar relações mais avançadas:** Durante esse trabalho não foi considerado modelos não recursivos, variáveis mediadoras e nem atributos com mais de um fator como possíveis soluções de alguns dos problemas encontrados (mais especificamente da multicolineariedade). Um exemplo de uma relação que poderia ser interessante é testar se o construto de *Tradição* pode ser mediador de alguns outros construtos, como a *Qualidade de Ensino*.

Trazendo para o contexto de pesquisa de mercado, uma das intenções desse trabalho foi começar a entender, a partir de dados reais, uma variável importante no mercado de ensino superior privado do Brasil: a satisfação do aluno. A empresa que cedeu os dados é atuante no ramo de pesquisa de mercado com foco no mercado educacional, então, criar novos métodos, seja para apresentar para os clientes ou para a construção de inteligência interna, ajuda a entender outras relações encontradas nas pesquisas.

Referências

EOM, Sean B.; WEN, H. Joseph ; Ashil, Nicholas. *The Determinants of Students' Perceived Learning Outcomes and Satisfaction in University Online Education: An Empirical Investigation*. **Decision Sciences Journal of Innovative Education**, v. 4, p. 215-235, 2006

EPSKAMP. Sacha., *semPlot: Path diagrams and visual analysis of various SEM packages' output*. URL <http://CRAN.R-project.org/package=semPlot>.

EVERITT, Brian., HOTHORN, Torsten. 2011. *An Introduction to Applied Multivariate Analysis with R*. Springer .New York, 2011

FARIAS, Salomão Alencar de; SANTOS, Rubens da Costa. Modelagem de equações estruturais e satisfação do consumidor: uma investigação teórica e prática. **Rev. adm. contemp.**, Curitiba , v. 4, n. 3, p. 107-132, Dec. 2000 .

GANSNER, Emden R., NORTH, Stephen C. *An open graph visualization system and its applications to software engineering*. **Software – Practice and Experience**, v 30, p. 1203-1233 ,2000.

HAIR, Joseph. F.; ANDERSON, Rolph. E; TARHAM, Ronald. L; BLACK, William. C; BABIN, Barry. J. **Análise Multivariada de dados** 6 Ed Editora Bookman. Porto Alegre, 2009.

KLINE, Rex B. **Principles and Practice of Structural Equation Modeling**. 3 Ed. The Guilford Press. New York, 2011.

LEMKE, C., **Modelos de Equações Estruturais com Ênfase em Análise Fatorial Confirmatória no Software AMOS**. Universidade Federal do Rio Grande do Sul, Instituto de Matemática, Departamento de Estatística. Monografia. Porto Alegre, 2005

MALHOTRA, Naresh K. Pesquisa de marketing: Uma orientação aplicada 6 Ed. Editora Bookman. Porto Alegre, 2012

MARTÍNEZ-CARO, Eva. *Factors affecting effectiveness in e-learning: An analysis in production management courses*. **Computer Applications in Engineering Education**, v 19, p. 572-581, 2011

PEREIRA, Suziane dos Santos, **Modelagem de Equações Estruturais no Software R**, Universidade Federal do Rio Grande do Sul, Instituto de Matemática, Departamento de Estatística. Monografia. Porto Alegre, 2013

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

ROSSEEL, Yves., *Lavaan: An R Package for Structural Equation Modeling*.
Journal of Statistical Software, 2012.

SILVA, Juliane. F, **Modelagem de equações estruturais: apresentação de uma metodologia**, Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Programa de Pós-Graduação em Engenharia de Produção. Tese. Porto Alegre, 2006

URDAN, André Torres; RODRIGUES, Arnaldo Rocha. O modelo do índice de satisfação do cliente Norte-Americano: um exame inicial no Brasil com equações estruturais. **Rev. adm. contemp.**, Curitiba , v. 3, n. 3, p. 109-130, Dec. 1999

Anexos

Formulas dos índices de ajuste do modelo

Goodness-of-fit Index (GFI)

$$GFI = 1 - \frac{F_M}{F_0}$$

Onde:

F_0 : Função de discrepância do modelo nulo.

F_m : Função de discrepância do modelo ajustado.

Adjusted Goodness-of-fit Index (AGFI)

$$AGFI = 1 - (1 - GFI) \frac{d_b}{gl}$$

Onde:

$$d_b = \sum_{(g=1)}^G p^{*(g)}$$

$p^{*(g)}$: Número de momentos amostrais no grupo g.

gl = Número de graus de liberdade do modelo.

Normal Fitted Index (NFI)

$$NFI = 1 - \frac{\chi_M^2}{\chi_N^2}$$

χ_M^2 : Valor da estatística qui-quadrado do modelo ajustado.

χ_N^2 : Valor da estatística qui-quadrado do modelo nulo.

Comparative Fit Index (CFI)

$$CFI = 1 - \frac{(\chi_M^2 - Gl_M)}{(\chi_N^2 - Gl_N)}$$

Onde:

Gl_M : Graus de liberdade do modelo ajustado.

Gl_N : Graus de liberdade do modelo nulo.

Non-Normal Fitted Index (NNFI)

$$NNFI = \frac{\left(\frac{\chi_N^2}{Gl_N}\right) - \left(\frac{\chi_M^2}{Gl_M}\right)}{\left(\frac{\chi_N^2}{Gl_N}\right) - 1}$$

Root Mean Square Error of Approximation (RMSEA)

$$RMSEA = \sqrt{\frac{\chi_M^2 - Gl_M}{(N - 1)}}$$

Onde:

χ_M^2 : Valor da estatística qui-quadrado do modelo ajustado.

χ_N^2 : Valor da estatística qui-quadrado do modelo nulo.

N = Tamanho da amostra