

XII Colóquio Internacional de Gestão Universitária



Modelagem de um Instrumento de Medida de Avaliação do ENADE fundamentado na Teoria de Resposta ao Item (TRI): desenho para o MEES

O objetivo geral deste artigo é apresentar uma modelagem de um instrumento de medida de avaliação para o ENADE fundamentada na TRI e, especificamente, propõe-se um desenho para o Mapa Estratégico da Educação Superior (MEES). O MEES se constitui em uma representação visual de um sistema de gestão integrado para a operacionalização do Plano de Desenvolvimento Institucional de Instituições de Educação Superior (IES). A base conceitual que fundamenta este estudo é a Teoria de Resposta ao Item (TRI) e o delineamento metodológico do estudo, na sua concepção filosófica se insere na abordagem quântica, que integra, de forma, dual e complementar, a mensuração e a observação dos fenômenos. Almeja-se que o estudo de caso (próxima etapa desse trabalho) possa contribuir para verificar a viabilidade do uso da TRI como instrumento de medida de avaliação do ENADE. Consequentemente, o desenho para MEES, o qual se constitui na representação visual de um sistema de gestão integrado para a administração da educação superior, deverá corroborar para a operacionalização de algumas das principais diretrizes do SINAES, dentre as quais destacam-se: respeito à identidade e à diversidade institucionais em um sistema diversificado; globalidade; legitimidade; e continuidade.



1. Introdução

O Exame Nacional de Desempenho de Estudantes (ENADE) é um subsistema de avaliação do Sistema Nacional de Avaliação da Educação Superior (SINAES), que tem o propósito de aferir o desempenho dos estudantes dos cursos de graduação em relação aos conteúdos programáticos previstos nas diretrizes curriculares do respectivo curso, suas habilidades e competências.

O ENADE foi instituído como um componente curricular obrigatório dos cursos de graduação, onde participam da avaliação os estudantes em final de primeiro ano (ingressantes) e de último ano (concluintes) das áreas e cursos a serem avaliados. Para este ano de 2012 os ingressantes estão dispensados da realização da prova, conforme estabelece a respectiva Portaria Normativa, com obrigatoriedade apenas de concluintes, incluindo, ainda, estudantes do penúltimo semestre.

Com o advento do Conceito Preliminar de Curso - CPC, instituído pela Portaria Nº 4/2008, tornou-se possível a substituição das tradicionais avaliações *in loco* para renovação de reconhecimento dos cursos de graduação por um cálculo aonde os resultados do ENADE são preponderantes. Com base nos resultados dos CPCs de todos os cursos de uma instituição passou a ser calculado o Índice Geral de Cursos (IGC) de uma instituição. Trata-se de um indicador geral de qualidade da instituição. Em função destes novos conceitos e critérios, independentemente dos questionamentos levantados (SCHWARTZMAN, 2008), o ENADE passou a ter muito maior repercussão.

Tradicionalmente, o instrumento de medida de avaliação das provas utilizados se fundamenta na Teoria Clássica dos Testes (TCT). A Teoria de Resposta ao Item (TRI) é uma ferramenta estatística que surgiu para suprir as necessidades decorrentes das limitações da TCT. Uma das principais vantagens da TRI sobre a TCT é que ela permite a comparação entre populações, desde que submetidas a provas que tenham alguns itens comuns. Também proporciona a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes. Uma das principais características da TRI é que ela tem como elementos centrais os itens, e não a prova como um todo. Este é o diferencial da TRI em relação a TCT, a qual tem aumentado a sua aplicabilidade nas avaliações educacionais e de proficiências.

O objetivo geral deste artigo é apresentar uma modelagem de um instrumento de medida de avaliação para o ENADE fundamentada na TRI e, especificamente, propõe-se um desenho para o Mapa Estratégico da Educação Superior (MEES). O MEES se constitui em uma representação visual de um sistema de gestão integrado para a operacionalização do Plano de Desenvolvimento Institucional de Instituições de Educação Superior (IES). A seção um apresenta a TRI; A segunda seção descreve sobre o ENADE; a terceira seção a modelagem proposta; e as considerações finais uma apreciação crítica sobre o estudo.

2. A Teoria da Resposta ao Item

As avaliações educacionais tradicionalmente têm sido analisadas segundo os critérios da Teoria Clássica dos Testes (TCT). O surgimento da Teoria da Resposta ao Item (TRI) possibilitou uma nova forma de elaborar avaliações educacionais. O uso da TRI apresenta potencialidades efetivas que tem incentivado o incremento de seu uso como instrumento de mensuração, principalmente em sistemas de avaliação educacional.

De acordo com Andrade, Tavares e Valle (2000), TRI é uma metodologia que sugere formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta



a um item e seus traços latentes. Sob este prisma a avaliação é realizada a partir de observações de variáveis secundárias relacionadas à determinada característica do indivíduo que não pode ser observada diretamente. Essa relação é possível por meio da utilização de modelos matemáticos e métodos complexos de estimação. Esta seção expõe um breve histórico e estado da arte sobre a TRI, o modelo matemático - Modelo Logístico Unidimensional de Três Parâmetros (MLU3), e um paralelo da TRI com a TCT, explicitando as vantagens e desvantagens no uso das referidas teorias.

2.1 Histórico e Estado da Arte

A Teoria da Resposta ao Item começou a ser estudada na década de 50, por meio de alguns modelos matemáticos mais simples aplicados em variáveis dicotômicas. Os primeiros modelos desenvolvidos, propostos por Lord (1952), foram o modelo unidimensional de dois parâmetros e o modelo unidimensional de três parâmetros, baseados na distribuição normal acumulada (ogiva normal). Birnbaum (1968) aprimorou os modelos de Lord, substituindo, em ambos os modelos, a função ogiva normal pela função logística, que é mais conveniente matematicamente. Paralelamente, porém independentemente do trabalho de Lord, Rasch (1960) propôs o modelo unidimensional de 1 parâmetro, expresso também como modelo de ogiva normal e, posteriormente aprimorado por Wright (1968) substituindo, também, a função ogiva normal pela função logística. Outros modelos importantes para variáveis com mais de duas categorias de resposta foram propostos por Samegima (1969), Bock (1972), Andrich (1978), Masters (1982) e Muraki (1992). Bock & Zimowski (1997) introduziram os modelos logísticos de 1, 2 e 3 parâmetros para duas ou mais populações de respondentes, o que trouxe novas possibilidades para as comparações de rendimentos de duas ou mais populações submetidas a diferentes testes com itens comuns.

A TRI começou a ser utilizada no Brasil em 1996, primeiramente na pesquisa AVEJU, da Secretaria de Estado da Educação de São Paulo, e posteriormente no Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (SARESP) e no Sistema de Avaliação da Educação Básica (SAEB) do INEP/MEC (ANDRADE; TAVARES; VALLE, 2000).

A TRI passou a ser utilizada na avaliação das provas do Exame Nacional do Ensino Médio (ENEM), sob a responsabilidade do Instituto Nacional de Estudos e Pesquisas Educacionais do Ministério da Educação (INEP/MEC). Algumas das provas do ENEM já tinham sido analisadas previamente em estudos que utilizaram a TRI.

Francisco (2005), realizou um estudo de caso por meio da aplicação da TRI, com a finalidade de verificar o desempenho dos alunos formandos do curso de Matemática da Universidade Estadual do Centro-Oeste (UNICENTRO), em Guarapuava-PR, no período de 2000 até 2003, no Exame Nacional de Cursos (ENC). Foi estudado a estimação dos parâmetros dos modelos logísticos unidimensionais de 1, 2 e 3 parâmetros da TRI. Os resultados mostraram que, para a população em estudo (alunos formandos do curso de Matemática da UNICENTRO), em todos os anos analisados, a grande maioria dos itens era difícil. No entanto, observaram que a quantidade de itens difíceis diminuiu ao longo do período analisado. Uma limitação desse estudo foi a amostra utilizada para a calibração dos itens. Foram analisadas quatro provas (uma para cada ano), com número de itens igual a 25, 40, 30 e 40, e com número de respondentes igual a 46, 59, 41 e 41, respectivamente. Essa quantidade pequena de respondente pode afetar seriamente a estimativas dos parâmetros dos itens, tornando os resultados duvidosos. Outra restrição desse estudo é que os parâmetros



estimados não podem ser utilizados para avaliar a proficiência dos alunos de outra IES, que poderiam apresentar DIF (*Differential item functioning*), já que a população não é a mesma.

Oliveira (2006) utilizou a TRI para proceder a uma análise das propriedades psicométricas da prova do ENADE de 2004, aplicada aos alunos dos cursos de medicina. A amostra utilizada foi de 8.124 estudantes e as análises foram realizadas apenas com as questões objetivas tanto do componente de Formação Geral (FG) quanto do componente de Formação Específica (FE), totalizando 28 itens. Os resultados, baseados no coeficiente de fidedignidade que utiliza a fórmula KR- α de Kuder e Richardson (1967) mostraram que os dados se ajustam com maior confiabilidade ao Modelo Logístico Unidimensional de Um parâmetro (Modelo de Rasch) do que nos Modelos Logísticos Unidimensionais de 2 e 3 Parâmetros.

Nogueira (2008) aplicou a TRI para avaliar as questões da prova de formação geral (FG) do ENADE, em especial aquelas que envolvem conceitos estatísticos, visando estimar a proficiência dos estudantes nos conteúdos avaliados e o ajuste dos itens ao modelo de Rasch. A amostra utilizada foi de 403.512 estudantes de vários cursos que fizeram a prova do ENADE nos anos de 2004 e de 2005. Os resultados mostraram que as questões objetivas da prova de 2004 apresentam parâmetros de dificuldade mais altos em relação à prova de 2005, exigindo maior habilidade do estudante para que haja a probabilidade de acerto. Por outro lado, as questões discursivas de 2004 e 2005 apresentam parâmetros de dificuldade mais baixos, embora não sejam equiparáveis. O Modelo da TRI utilizado nesse estudo é mais restrito, pois considera que todos os itens possuem o mesmo nível de discriminação e apenas a dificuldade dos itens é avaliada, não levando em conta a possibilidade do acerto casual.

Primi *et al* (2009), apresentam outro enfoque da aplicação da TRI no ENADE para a determinação de pontos de corte, formando grupos de competências requeridas para a resolução de itens. A amostra utilizada foi de 26.613 estudantes que realizaram a prova do ENADE de Psicologia em 2006 e foram analisadas somente as questões referentes ao conteúdo específico. Os resultados mostraram que, de maneira geral, os estudantes concluintes concentram-se na competência mínima, enquanto que os ingressantes distribuem-se, em sua maioria, em fraca e competência mínima.

Primi *et al* (2010), expõem um estudo utilizando DIF (*Differential item functioning*) no ENADE. O DIF pode ser definido com sendo a observação, em pessoas com a mesma habilidade, de uma chance diferenciada de acerto de um item. Os parâmetros estimados pela TRI devem ser invariantes em relação à habilidade da amostra, mas quando isso não ocorre, tem-se essa situação anômala chamada de DIF.

O uso da TRI no ENEM proporcionou o seu destaque em âmbito nacional. No contexto internacional, a TRI vem sendo empregada amplamente por vários países: Estados Unidos, França, China, Holanda, Coreia do Sul e principalmente nos países participantes do Programa Internacional de Avaliação de Estudantes (PISA). O PISA utiliza o modelo de Rasch da Teoria de Resposta ao Item (TRI) e coloca os resultados em uma mesma escala de proficiências para cada área, ao longo dos anos (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2003, 2011; KLEIN, 2011).

Outro exemplo de avaliação utilizando a TRI é o exame de proficiência em língua inglesa (TOEFL). Este exame surgiu em 1964 e é largamente utilizado em todo o mundo. Desde o início de sua origem, este exame já avaliou mais de 25 milhões de alunos e tem sido administrado por mais de 4.500 centros em 165 países (KARINO, ANDRADE, 2010).



As avaliações educacionais tradicionalmente têm sido analisadas segundo os critérios da TCT. A TRI surge como uma metodologia alternativa utilizada em diversas áreas do conhecimento, e em larga escala em avaliações educacionais.

2.2 Modelo Matemático

Como foi mencionado na seção anterior, existem diversos modelos matemáticos que são utilizados nas aplicações da TRI. Propõe-se neste estudo a utilização do Modelo Logístico Unidimensional de Três Parâmetros (MLU3), porque é o mais indicado e aplicado nas avaliações educacionais de proficiência. O MLU3 é adequado para o ajuste de itens politômicos (itens com duas ou mais categorias) com uma única opção de resposta correta, o que permite que o item seja dicotomizado em duas categorias: certa e errada. Além disso, esse modelo permite modelar a probabilidade do acerto casual, ou seja, a probabilidade de um aluno com baixa proficiência acertar um determinado item. O MLU3 é dado por:

$$P(U_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \text{ para } i = 1, 2, \dots, I, \text{ e } j = 1, 2, \dots, n \quad (1),$$

onde:

U_{ij} é uma variável dicotômica (assume o valor 1 quando o indivíduo j responde corretamente o item i , ou assume o valor 0, caso contrário);

θ_j é o valor do traço latente (parâmetro da habilidade) do indivíduo j ;

$P(U_{ij} = 1 | \theta_j)$, também chamada de Função de Resposta do Item (FRI), é a probabilidade do indivíduo j responder corretamente o item i , dado que ele tem habilidade θ_j , ou seja, é a proporção de respostas corretas do item i dos indivíduos da população com habilidade θ_j ;

a_i é o parâmetro de discriminação (ou de inclinação) do item i ;

b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da habilidade;

c_i é o parâmetro de acerto casual, que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i ;

D é um fator de escala constante, igual a 1 se os parâmetros dos itens são estimados na métrica da Logística, ou igual a 1,7, se os parâmetros dos itens são estimados na métrica da ogiva Normal, que é a distribuição Normal acumulada, por aproximação (nesse estudo os parâmetros serão analisados pela métrica da Logística, considerando, portanto, $D = 1$);

e é a conhecida constante matemática igual a 2,718281...;

I é o número total de itens; e

n é a quantidade total de indivíduos na amostra.

O MLU3 é um modelo acumulativo, ou seja, a medida que o valor do traço latente aumenta, a probabilidade do indivíduo acertar o item também aumenta e vice-versa. A 0 apresenta um exemplo de uma Curva Característica do Item (CCI) de um MLU3 e a sua relação existente com os parâmetros dos itens a_i (inclinação da curva), b_i (posição do item na escala) e c_i (probabilidade de acerto casual de indivíduos com baixa habilidade). A CCI é o gráfico da função do modelo matemático, onde o eixo Y é a probabilidade de resposta correta de um indivíduo segundo o valor da sua habilidade (eixo X).

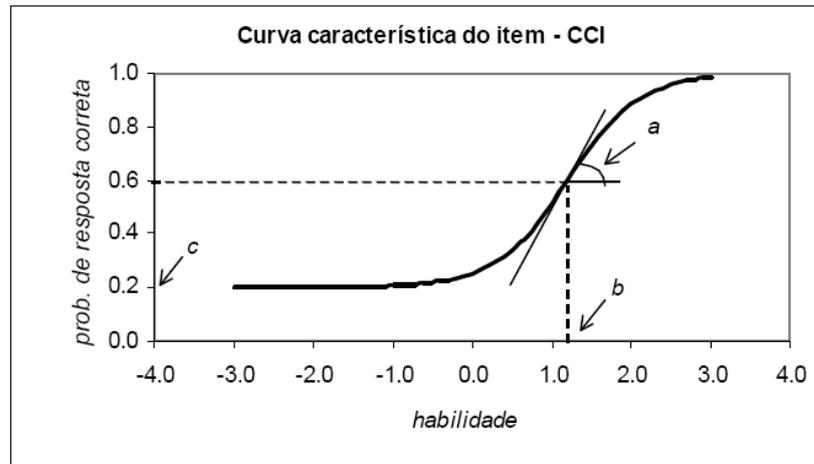


Figura 1: Relação entre os parâmetros dos itens e a CCI

Fonte: ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. (2000).

O traço latente (habilidade ou proficiência) do indivíduo (θ_j) é medido em uma escala arbitrária que varia teoricamente entre $-\infty$ e $+\infty$. Porém, o importante nessa escala não é a sua magnitude, mas as relações de ordem existentes (ANDRADE; TAVARES; VALLE, 2000). O traço latente, no modelo acumulativo, é especificado como um tipo de característica que apresenta uma probabilidade maior para indivíduos com θ_j maior, e uma probabilidade menor para indivíduos com θ_j menor. Ou seja, quanto maior for θ_j , maior será a probabilidade do indivíduo j acertar o item.

Segundo Baker (2001), o traço latente θ_j do indivíduo j é invariante em relação aos itens utilizados para estimá-la, desde que os itens sejam adequados, isto é, estejam calibrados (ou seja, possuam uma boa estimativa dos parâmetros), em uma métrica comum e medindo o mesmo traço latente (unidimensionalidade). Isso justifica o fato do resultado do θ_j ser o mesmo, independente dos itens que formam o questionário, o que não ocorre na TCT. Portanto, não importa se o teste é composto por itens difíceis ou fáceis, a estimativa da habilidade é a mesma. Isso é condizente com a realidade, já que a habilidade de um indivíduo, num determinado tempo t , é a mesma independente do grau de dificuldade do teste. Essa é a chamada propriedade de invariância do parâmetro de habilidade da TRI.

O parâmetro a_i mede a discriminação do item. Valores baixos de a_i indicam que o item tem pouco poder de discriminação, ou seja, a probabilidade de um indivíduo responder corretamente o item ou concordar com ele é aproximadamente a mesma para indivíduos com baixa ou alta proficiência. Por outro lado, valores altos de a_i indicam que o item tem grande poder de discriminação, dividindo os indivíduos praticamente em dois grupos: os que possuem habilidades abaixo do valor de b_i (com baixa probabilidade de acertar o item), e os que possuem habilidades acima do valor de b_i (com alta probabilidade de acertar o item). Não existe um valor exato de a_i para decidir se um item discrimina bem ou não. Em geral, na métrica logística, um item com a_i maior que 0,7 pode ser considerado aceitável, mas um valor maior ou igual a 1,0 indica que o item discrimina bem. Valores extremamente altos de a_i também não são adequados, pois provavelmente dividiria os indivíduos em dois grupos



distintos (os que têm θ_j maior que b_i e os que têm θ_j menor que b_i), mas não faria distinção entre os indivíduos dentro dos grupos.

O parâmetro mais importante do MLU3 é o b_i , parâmetro de dificuldade ou proficiência do item, que é medido na mesma unidade da escala da habilidade do indivíduo (θ_j). Ele representa o grau de dificuldade do item, ou seja, quanto maior seu valor, mais difícil o item é (somente indivíduos com habilidade alta terão uma boa probabilidade de acertá-lo), e vice-versa. Esse valor de b_i é que vai definir a posição do item na escala, por isso ele também é chamado de parâmetro de localização. Teoricamente, b_i pode assumir qualquer valor entre $-\infty$ e $+\infty$, entretanto, para valores muito altos ou baixos, o item pode não ser adequado, sendo usual os valores entre -3 e 3, na escala (0, 1), isto é, com média igual a zero e desvio padrão igual a um.

O parâmetro c_i é a probabilidade de um indivíduo com baixa proficiência ou com pouco (ou nenhum) conhecimento, em relação ao assunto que está sendo avaliado, responder corretamente ao item i . O parâmetro c_i é considerado quando existe a possibilidade de acerto casual, que é o caso do MLU3, e o seu valor depende da quantidade de alternativas que o item apresenta.

Todo item fornece uma informação à avaliação na TRI, através da Função de Informação do Item (FII), que permite analisar a quantidade de informação que um item fornece para a medida do traço latente analisado e reflete a qualidade do item. Maiores detalhes podem ser encontrados em Andrade, Tavares e Valle (2000).

2.3 Um paralelo entre a TCT E A TRI

O Quadro 1 apresenta um estudo comparativo entre a TRI e a TCT, destacando as vantagens no uso da TRI.

TRI	TCT
Fornecer informações mais precisas sobre o desempenho dos respondentes já que o traço latente do indivíduo não depende da dificuldade das questões do teste.	O escore do indivíduo depende essencialmente dos itens que compõem o teste.
Analisa com mais eficácia as questões utilizadas (itens), que devem ter índices aceitáveis em relação aos parâmetros (dificuldade, discriminação, etc).	Calcula os índices (dificuldade, discriminação, etc) utilizando contagem de escores e percentuais.
Permite obter índices de precisão do item (função de informação do item - FII) e do teste (função de informação do teste - FIT) mais sofisticados	Os índices de precisão são menos sofisticados (Alfa de Crombach e Erro Padrão da Medida)
Possui a possibilidade de considerar o acerto casual.	Não possui a possibilidade de considerar o acerto casual.
Oferece alternativas adequadas no caso da existência de itens que não se aplicam ou que não foram respondidos pelos indivíduos	Não oferece alternativas adequadas no caso da existência de itens que não se aplicam ou que não foram respondidos pelos indivíduos
Os indivíduos que acertam a mesma quantidade de itens possuem o mesmo escore, se e somente se acertaram os mesmos itens na mesma prova. Se acertarem itens diferentes, a nota será diferente de acordo com os itens que foram acertados.	Os indivíduos que acertam a mesma quantidade de itens possuem o mesmo escore, tendo acertado itens diferentes (fáceis ou difíceis) na mesma prova.
Permite a comparação através do escore entre os indivíduos que responderam questionários com itens diferentes para medir o mesmo traço latente.	Não permite a comparação através do escore entre os indivíduos que responderam questionários com itens diferentes para



	medir o mesmo traço latente.
Consegue determinar qual seria a resposta de um indivíduo a um determinado item que ele não respondeu em um questionário.	Não consegue determinar qual seria a resposta de um indivíduo a um determinado item que ele não respondeu em um questionário.
Os índices de dificuldade e de discriminação de um item dependem dos seus parâmetros devidamente calibrados, e não de quem ou quantos irão responder.	Os índices de dificuldade e de discriminação, por serem uma proporção de acertos, dependem da quantidade e da qualidade dos respondentes, ou seja, depende da amostra.
Não necessita de amostra representativa para estimar os seus índices.	Necessita de amostra representativa para estimar os seus índices.
Cada respondente tem seu próprio erro padrão de medida (EPM)	O erro padrão de medida (EPM) é o mesmo para todos os escores dos respondentes.
Os testes curtos podem ser mais confiáveis que os testes longos.	Os testes mais longos são mais confiáveis do que os curtos
Formatos mesclados de itens (por exemplo, dicotômicos e politômicos nominais e graduais) na TRI podem resultar em excelentes escores de testes.	Formatos mesclados de itens (por exemplo, dicotômicos e politômicos nominais e graduais) na TCT conduzem a impacto desequilibrado nos escores total do teste.
A confiabilidade é examinada em nível do item, portanto não há uma única estimativa de confiabilidade, mas pode ser estimado um índice composto da confiabilidade	Analisa a confiabilidade em função de uma medida como um todo, geralmente por meio do Alfa de Cronbach

Quadro 1- Estudo Comparativo entre a TRI e a TCT - Vantagens

Síntese elaborada com base nos estudos de ANDRADE; TAVARES; VALLE (2000); EMBRETSON; REISE (2000), HAMBLETON; SWAMINATHAN; ROGERS (1991); OLEA *et al.* (1996); SÃO PAULO *et al.* (2007); VENDRAMINI; SILVA; CANEL (2004).

O Quadro 2 apresenta um estudo comparativo entre a TRI e a TCT, destacando as desvantagens no uso da TRI.

TRI	TCT
O usuário pode encontrar dificuldades encontradas com a aplicabilidade da sua própria natureza, tanto do ponto de vista teórico, devido a problemas de difícil solução no campo da estimação, como do ponto de vista computacional.	É mais simples e mais fácil de usar, não depende de métodos sofisticados de estimação
Em alguns casos, a utilização da TRI pode não contribuir muito.	Em alguns casos, as análises da TCT do item poderão fornecer toda a informação que é necessária.
As análises da TRI exigem um <i>software</i> especial e em geral diversos programas diferentes são necessários para executar os testes das suposições e estimação dos parâmetros, bem como os testes para ajuste do modelo.	As análises da TCT não exigem um <i>software</i> especial
Em quase todos os casos existe a necessidade de uma amostra bastante grande, dependendo da quantidade de parâmetros que serão estimados e do padrão das respostas.	É possível obter resultados com amostras pequenas.

Quadro 2- Estudo Comparativo entre a TRI e a TCT - Desvantagens

Síntese elaborada com base nos estudos de ANDRADE; TAVARES; VALLE (2000); EMBRETSON; REISE (2000), HAMBLETON; SWAMINATHAN; ROGERS (1991); OLEA *et al.* (1996); SÃO PAULO *et al.* (2007); VENDRAMINI; SILVA; CANEL (2004).

Os estudos têm mostrado que, na maioria das aplicações, a TRI tem obtido bons resultados. Particularmente, na avaliação educacional, geralmente realizada em larga escala,



os resultados tem sido sempre satisfatórios. Os ganhos na utilização da TRI incluem: comparar alunos que responderam provas com questões diferentes, estimar a resposta de um aluno a um item que ele não respondeu ou que não tenha sido aplicado a ele, e identificar o nível em que o aluno se encontra na escala de proficiência. A TRI também permite a comparação do desempenho dos alunos entre as escolas, tipo de rede (municipal, estadual, federal e particular, privada). Devido às propriedades da TRI, foi possível a criação do Índice de Desenvolvimento da Educação Básica (IDEB) no ENEM, um indicador para medir a qualidade de cada escola, de cada rede de ensino. O indicador é calculado com base no desempenho do estudante em avaliações do INEP e em taxas de aprovação. Assim, para que IDEB de uma escola ou rede cresça é preciso que o aluno aprenda, não repita o ano e frequente a sala de aula.

Recentemente, houve um caso no ENEM onde alguns alunos fizeram provas que, devido a dificuldades na logística, houve vazamentos nas respostas. Com a utilização da TRI, apenas esses alunos se submeteram a realização de outra prova novamente, sem prejuízo de nota para aqueles que não apresentaram problemas na sua prova. Se fosse a abordagem clássica, todos os alunos teriam que fazer novamente a prova, pois as questões seriam diferentes, o que prejudicaria a nota dos alunos que não refizessem a prova.

3. O Exame Nacional de Desempenho de Estudantes (ENADE)

O Ministério da Educação (MEC) instituiu o Sistema Nacional de Avaliação da Educação Superior (SINAES), por meio da lei nº 10.861, de 14 de Abril de 2004, com o propósito precípua de promover a melhoria contínua da qualidade da educação superior (BRASIL, 2004).

O SINAES se constitui dos seguintes subsistemas de avaliação: avaliação institucional, a qual integra a auto-avaliação e a avaliação externa; a avaliação de cursos de graduação e a avaliação do desempenho de estudantes (SINAES, 2009, p.153).

A avaliação do desempenho dos estudantes é realizada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), sob a orientação da Comissão Nacional de Avaliação da Educação Superior (CONAES), mediante a aplicação do ENADE.

O ENADE foi instituído legalmente em 2004 pelo SINAES, conforme estabelece o Art. 5 da Lei 10.861/2004, de onde aponta-se alguns parágrafos que evidenciam a sua concepção:

Art. 5º A avaliação do desempenho dos estudantes dos cursos de graduação será realizada mediante aplicação do Exame Nacional de Desempenho dos Estudantes - ENADE.

§ 1º O ENADE aferirá o desempenho dos estudantes em relação aos conteúdos programáticos previstos nas diretrizes curriculares do respectivo curso de graduação, suas habilidades para ajustamento às exigências decorrentes da evolução do conhecimento e suas competências para compreender temas exteriores ao âmbito específico de sua profissão, ligados à realidade brasileira e mundial e a outras áreas do conhecimento.

§ 2º O ENADE será aplicado periodicamente, admitida a utilização de procedimentos amostrais, aos alunos de todos os cursos de graduação, ao final do primeiro e do último ano de curso.

§ 3º A periodicidade máxima de aplicação do ENADE aos estudantes de cada curso de graduação será trienal.

§ 4º A aplicação do ENADE será acompanhada de instrumento destinado a levantar o perfil dos estudantes, relevante para a compreensão de seus resultados.

§ 5º O ENADE é componente curricular obrigatório dos cursos de graduação, sendo inscrita no histórico escolar do estudante somente a sua situação regular com relação a essa obrigação, atestada pela sua efetiva participação ou, quando for o caso, dispensa oficial pelo Ministério da Educação, na forma estabelecida em regulamento.



§ 8º A avaliação do desempenho dos alunos de cada curso no ENADE será expressa por meio de conceitos, ordenados em uma escala com 5 (cinco) níveis, tomando por base padrões mínimos estabelecidos por especialistas das diferentes áreas do conhecimento.

§ 10. Aos estudantes de melhor desempenho no ENADE o Ministério da Educação concederá estímulo, na forma de bolsa de estudos, ou auxílio específico, ou ainda alguma outra forma de distinção com objetivo similar, destinado a favorecer a excelência e a continuidade dos estudos, em nível de graduação ou de pós-graduação, conforme estabelecido em regulamento (MINISTÉRIO DA EDUCAÇÃO, 2004)

Posteriormente, a PORTARIA NORMATIVA Nº 40, de 12 de dezembro de 2007, veio instituir o *e-MEC*, sistema eletrônico de fluxo de trabalho e gerenciamento de informações relativas aos processos de regulação, avaliação e supervisão da educação superior no sistema federal de educação, e o Cadastro *e-MEC* de Instituições e Cursos Superiores, bem como consolidar disposições sobre indicadores de qualidade, banco de avaliadores (Basis) e o Exame Nacional de Desempenho de Estudantes (ENADE), dentre outras disposições.

A cada ano o Ministério da Educação, por Portaria Normativa, defini os cursos que participarão da prova e estabelece as regulamentações complementares relacionadas as provas daquele ano.

Conforme já previsto pela Portaria Normativa Nº 40, de 12/12/2007, no § 3 do seu Art.33-F, os ingressantes poderiam ser dispensados da prova.

Os alunos ingressantes que tiverem realizado o ENEM, aplicado com metodologia que permita comparação de resultados entre edições do exame, poderão ser dispensados de realizar a prova geral do ENADE, mediante apresentação do resultado válido. (MINISTÉRIO DA EDUCAÇÃO, 2007).

Pela Portaria Normativa Nº 6, de 14 de março de 2012, os ingressantes ficam dispensados da prova neste ano de 2012 (MINISTÉRIO DA EDUCAÇÃO, 2012). A expectativa é de que, com a universalização do ENEM, o Índice de Diferença de Desempenho – IDD seja calculado com base no perfil do ingressante em função do desempenho no Ensino Médio e não na mesma prova aplicada a ingressantes e concluintes de um mesmo curso na Educação Superior.

Os dados sobre o desempenho dos estudantes da IES no ENADE, o IDD, e do Questionário Socioeconômico dos estudantes, coletados na aplicação do ENADE serão examinados pelas Comissões Externas de Avaliação das Instituições e pelas Comissões Externas de Avaliação de Cursos (SINAES, 2004). Estes dados vão, preponderantemente, compor, com a inclusão de dados do corpo docente, o Conceito Preliminar de Curso, o CPC, que poderá substituir as avaliações *in loco* para a renovação do reconhecimento dos cursos de graduação do país. Assim, com a Portaria Nº 4, de 05 de agosto de 2008, o desempenho dos estudantes nas provas do ENADE, bem como as suas avaliações sobre a IES, expressas no Questionário Sócio-econômico, entregues na ocasião da prova, passaram a informar os dados com base nos quais os cursos poderão ter o seu reconhecimento renovado.

Com a Portaria Normativa Nº 12, do Ministério da Educação, de 05 de julho, também de 2008, o CPC de todos os cursos de graduação de uma instituição, acrescidos dos dados de seus cursos de pós-graduação *stricto sensu*, passaram a compor o Índice Geral de Cursos que é uma espécie de “nota” sobre a qualidade de uma IES. Tal índice permite comparar as instituições entre si e favorece o ranqueamento das mesmas.

Não tem sido poucos os questionamentos levantados no meio acadêmico sobre a validade técnica e política destas medidas e critérios avaliativos. Dentre eles destaca-se a do Prof. Simon Schwartzman (2008) que critica aos critérios adotados para o cálculo do CPC, como uma decisão discricionária, feita sem uma justificativa explícita.



Tais medidas passaram a ser criticadas também como um desvirtuamento das políticas de avaliação da Educação Superior Brasileira estabelecidas pelo SINAES.

Cabe ainda destacar que a importância atribuída pelo SINAES ao processo de autoavaliação realizado pelas IES é absolutamente relegada no momento da construção do IGC, não obstante também integre a avaliação institucional. Assim, o desempenho de uma instituição de ensino resume-se ao resultado obtido no ENADE, somado a opinião do aluno a respeito de alguns insumos, uma vez que os dados do Censo representam menos de 15% do CPC, e os demais processos de avaliação não são considerados em sua composição. Ratifica-se o exposto, na medida em que a composição do CPC e do IGC, derivado deste, não contempla as 10 dimensões do SINAES, tais como Responsabilidade Social, Comunicação, Gestão da IES e Sustentabilidade Financeira, considerando, apenas, a avaliação que o estudante faz da instituição. (POLIDORI, Marlis Morosini et al, 2011. P.270)

Os questionamentos que se destacam, como os acima apresentados, referem-se principalmente ao uso dos resultados do ENADE e a validade destes usos. Não são indagações sobre questões técnicas ou sobre os critérios estatísticos adotados. Uma vez que o uso dos resultados do ENADE na composição do CPC e do IGC estão sendo polemizados se torna relevante a revisão da validade técnica e dos critérios estatísticos adotados de forma a repensar em soluções alternativas para o aperfeiçoamento do SINAES. É o que pretende este estudo com a construção da modelagem de um Instrumento de Medida de Avaliação do ENADE.

4. A Administração da Educação Superior no Brasil

A administração da educação superior constitui-se de cinco elementos que devem atuar de maneira dialógica e indissociada: o ensino, a pesquisa, a extensão, a gestão e a cultura e arte. O tipo de IES, universidade, faculdade ou centro universitário irá determinar em quais desses elementos deverão priorizar as suas atividades. De qualquer forma, entende-se que todos os elementos integram três dimensões: pedagógica, avaliativa, estratégica e informacional. Considerando os elementos supracitados que integram a administração da educação superior, a repercussão da hegemonia da abordagem clássica na gestão das suas atividades fins e meios, faz-se necessário o estudo de novas concepções em paradigmas alternativos.

O SINAES tem o propósito de realizar uma avaliação diagnóstica, formativa e regulatória das IES. O Plano de Desenvolvimento Institucional (PDI) deve considerar a missão, os objetivos e as metas da instituição, bem como as propostas de desenvolvimento das suas atividades, definindo claramente os procedimentos relativos à qualificação do corpo docente, inclusive quanto a eventuais substituições, assim como o regime de trabalho, o plano de carreira, a titulação, a experiência profissional no magistério superior e a experiência profissional não acadêmica, levando em conta as condições de formação em pós-graduação de docentes na região, o projeto pedagógico dos cursos e as outras atribuições acadêmicas dos docentes (SINAES, 2009).

Corrêa (2007) e Corrêa *et al.* (2008) relatam que as suas experiências como avaliadores de cursos de graduação, demonstram que muitos dos instrumentos do SINAES estão cumprindo uma função meramente formal. Não estão sendo construídos coletivamente com a participação efetiva da comunidade acadêmica, conforme preconizam as diretrizes do SINAES. Na fase exploratória deste estudo constatou-se por meio da realização de um *benchmarking* dos Projetos Pedagógicos Institucionais (PPIs) de algumas IES, que uma parcela significativa foi construída desarticuladamente dos demais instrumentos do SINAES. Verificou-se também que um problema comum no processo de elaboração do PDI e do PPI e nos instrumentos de avaliação, é que os mesmos não são acompanhados de um plano de



implementação operacional.

Constata-se a necessidade de uma otimização dos instrumentos de planejamento e avaliação do MEC, com o propósito de: evitar duplicidade de informações e permitir a convergência de esforços e a sua implementação de maneira dinâmica, sistêmica e integrada. O sistema de gestão integrado apresentado na seção seguinte se constitui em uma proposta metodológica para a operacionalização do PDI.

5. Concepção do Mapa Estratégico da Educação Superior (MEES) fundamentado em um Sistema de Gestão Integrado: uma Proposta Metodológica para a Operacionalização do PDI

O Mapa Estratégico da Educação Superior (MEES) é a representação visual de um sistema de gestão integrado para as IES, o qual se constitui em uma proposta metodológica para operacionalização do PDI fundamentado na abordagem quântica (cf. Figura 2). Sob esta ótica o PDI é concebido como um sistema dual e complementar de medição de desempenho e observação dos padrões arquétipos comportamentais.



Figura 2 - Mapa Estratégico da Educação Superior Fundamentado em um Sistema de Gestão Integrado: Uma proposta metodológica para a operacionalização do PDI fundamentado na Abordagem Quântica.

Fonte: Corrêa et al (2007, 2008, 2011)

Incorpora aos instrumentos de planejamento e avaliação do Sistema Nacional de Avaliação da Educação Superior, uma concepção alternativa à educação superior. O PDI é construído coletivamente a partir das bases da instituição, por meio dos seguintes instrumentos: Plano de Desenvolvimento da Unidade (PDU) e seus subsistemas: Projeto Pedagógico da Unidade (PPU), Avaliação da Unidade (AU) e Plano Operacional da Unidade (POI).

A abordagem quântica à educação superior fundamenta-se em quatro princípios:



sistêmico, orgânico, holográfico e ecológico. O princípio sistêmico concebe o sistema de gestão como interativo, o qual visa integrar as políticas, diretrizes e estratégias institucionais com as setoriais. O princípio holográfico visualiza a instituição como um holograma, os Planos de Desenvolvimento das Unidades (PDUs) estão contidos no PDI e vice-versa, simultaneamente. O princípio ecológico extrapola as questões ambientais. Os propósitos e valores do sistema universitário definem a ecologia do ambiente. O princípio orgânico elege como requisitos de um sistema sob a ótica quântica a flexibilidade e a adaptabilidade das estratégias da educação superior às constantes mutações do ambiente. O mapa estratégico da educação superior visa apresentar a sociedade os objetivos, indicadores, metas e iniciativas estratégicas com transparência e continuidade.

6. Modelagem de um Instrumento de Medida de Avaliação do ENADE fundamentado na Teoria de Resposta ao Item (TRI): Desenho para o MEES

A Figura 3 expõe o desenho para o MEES de um instrumento de medida de avaliação do ENADE.

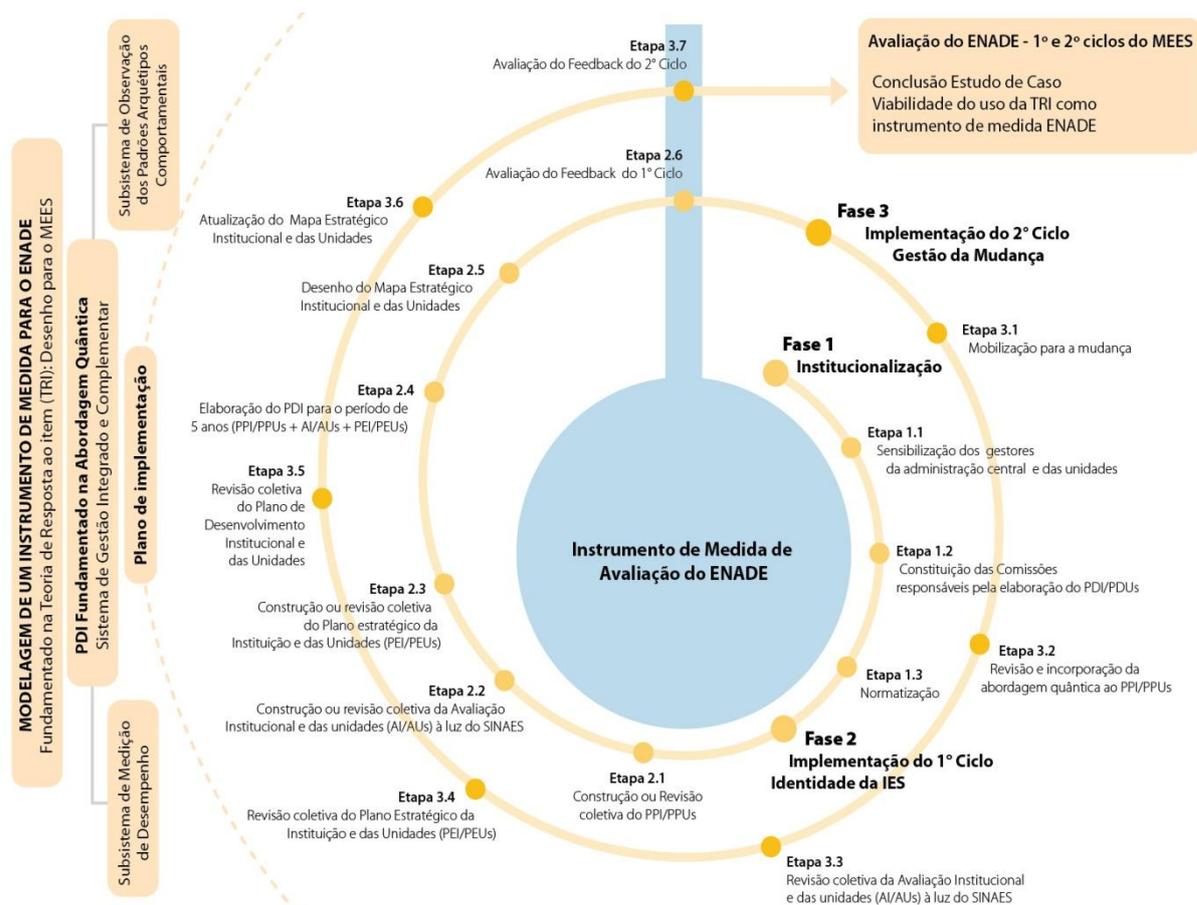


Figura 3 - Modelagem de um Instrumento de Medida de Avaliação do ENADE fundamentado na Teoria de Resposta ao Item (TRI): desenho para o MEES

O Plano de implementação do MEES compõe-se basicamente de três fases(cf. Figura 3). A fase 1 - a Institucionalização consiste na formalização do comprometimento da administração com a implementação do MEES; fase 2 – a implementação do primeiro ciclo



tem o propósito de reconhecer a identidade institucional, por meio do comportamento organizacional; fase 3 – a implementação do terceiro ciclo tem a finalidade de implementar a gestão da mudança, por meio da revisão de todas etapas do MEES. Visa potencializar os padrões comportamentais pró-ativos e minimizar os reativos.

A Figura 4 apresenta a incorporação do instrumento de medida do ENADE na avaliação do MEES. O desenho consiste basicamente na contribuição do instrumento de medida de avaliação para o ENADE nas etapas 2.2 e 3.3, dos primeiro e segundo ciclos do MEES. As referidas etapas consistem na construção e revisão coletiva da Avaliação Institucional alinhada à Avaliação das Unidades à luz do SINAES. Neste contexto, a avaliação do desempenho de estudantes será feita de forma integrada e interativa com os demais subsistemas de avaliação do SINAES: auto-avaliação, avaliação do ensino de graduação e pós-graduação da instituição e da(s) unidade(s).

O delineamento metodológico da modelagem ao ser classificado nas distintas concepções filosóficas da ciência se enquadra na abordagem quântica.

Corrêa, Cunha e Sutilli (2003), inserem a abordagem quântica como um paradigma alternativo da ciência, na taxionomia das suas concepções filosóficas, a qual agrega o ponderável, previsível, linear e passível de medição e determinação ao imponderável, imprevisível e não linear, passível de observação. Advogam que esta concepção científica possui características comuns à abordagem sistêmica, tais como o enfoque sistêmico e a relação de causalidade probabilística entre as variáveis. Apresenta como diferenciais a observação enquanto ato de percepção, na relação observador e objeto observado, a dualidade e a complementaridade dos fenômenos.



Figura 4 - Incorporação do instrumento de medida do ENADE fundamentado na Teoria de Resposta ao



Item(TRI) na avaliação do MEES

O método de estudo a ser aplicado no desenvolvimento experimental da modelagem será o estudo de caso. Segundo Ludke e André (1986), o estudo de caso é um estudo de um caso, seja ele simples e específico ou complexo e abstrato. É sempre bem delimitado, devendo ter seus contornos claramente definidos no desenrolar do estudo. O caso pode ser similar a outros, mas é ao mesmo tempo distinto, pois tem um interesse próprio, singular. “[...]Por sua flexibilidade, é recomendável nas fases iniciais de uma investigação sobre temas complexos, para a construção de hipóteses ou reformulação do problema”(GIL, 1989, p.79).

O estudo de caso da aplicação experimental do MEES se constitui das etapas descritas no centro da figura 4. O tamanho da amostra para ter eficácia na utilização da TRI deverá ser grande. Por isso optou-se por realizar a escolha de uma amostra representativa de um curso por região do país. Os dados serão coletados do banco de itens, de testes do ENADE já realizados. Será solicitada autorização do Instituto Nacional de Estudos e Pesquisas Anísio Teixeira do Ministério da Educação (MEC/INEP). O curso de administração é o curso priorizado para ser objeto do estudo de caso, já que um dos autores desta pesquisa já participa por dois anos consecutivos como membro das comissões assessoras de área deste curso no ENADE. Cabe destacar que a Portaria Normativa Nº 40, de 12/12/2007, no § 2 do seu Art.33-D, regulamenta que o banco de itens é elaborado por um corpo de especialistas, conforme orientação das Comissões Assessoras de área, para composição das provas do ENADE (MINISTÉRIO DA EDUCAÇÃO, 2007).

Após a definição da amostra será feita a estimativa dos parâmetros dos itens e a definição da escala de medida por meio da utilização do Modelo Logístico Unidimensional de Três Parâmetros (MLU3), já descrito na subseção 2.2 deste estudo. A análise dos resultados nos permitirão verificar a viabilidade da aplicação da TRI como instrumento de medida de avaliação do ENADE. A interação da TRI com o MEES acontece de maneira natural porque ambas tem com base o enfoque da abordagem sistêmica, ou seja, visualizar a parte no todo e o todo na parte.

7. Considerações Finais

Constata-se que uma polêmica entre os pares na comunidade científica sobre o cálculo do conceito do ENADE, do Conceito Preliminar de Cursos (CPC) e do Índice Geral de Cursos (IGC) e do cálculo da diferença entre o índice de desempenho esperado e o atingido no exame nacional de ensino médio (ENEM).

Por outro lado a comunidade científica questiona também as distorções e o mal uso pela mídia destes índices, bem como a supervalorização e a desarticulação de outros subsistemas do SINAES. Pontualmente destacam a pouca atenção a auto-avaliação institucional e a desarticulação desta a avaliação de cursos e ao ENADE.

Almeja-se que o estudo de caso (próxima etapa desse trabalho) possa contribuir para verificar a viabilidade do uso da TRI como instrumento de medida de avaliação do ENADE. Conseqüentemente, o desenho para MEES, o qual se constitui na representação visual de um sistema de gestão integrado para a administração da educação superior, deverá corroborar para a operacionalização de algumas das principais diretrizes do SINAES, dentre as quais destacam-se: respeito à identidade e à diversidade institucionais em um sistema diversificado; globalidade; legitimidade; e continuidade. Pretende-se incentivar a reflexão sobre a importância de todos os subsistemas de avaliação do SINAES serem operacionalizados de



maneira integrada. Destaca-se também neste contexto a relevância da avaliação construída coletivamente com a participação efetiva da comunidade acadêmica e sociedade conforme preconizam as diretrizes do SINAES.

Referências Bibliográficas

- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da resposta ao item: conceitos e aplicações**. São Paulo: ABE - Associação Brasileira de Estatística, 2000.
- ANDRICH, D. A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573, 1978.
- BAKER, F. B. **The Basics of Item Response Theory**. 2 ed. USA: ERIC Clearinghouse on Assessment and Evaluation, 2001.
- BRASIL. Lei nº 10.861, de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior - SINAES - e dá outras Providências. **Diário Oficial da União**, Brasília, DF, DOU nº72, de 15/4/2004, seção 1, p. 3-4.
- BIRNBAUM, A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F.M. Lord & M.R. Novick. **Statistical Theories of Mental Test Scores**. Reading, MA: Addison-Wesley, 1968.
- BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51, 1972.
- BOCK, R. D.; ZIMOWSKI, M. F. Multiple Group IRT. In **Handbook of Modern Item Response Theory**. W.J. van der Linder e R.K. Hambleton Eds. New York: Springer-Verlag, 1997.
- CORRÊA, A. C.; CUNHA, C.; SUTILLI, V. *The new paradigms of the science and their implications in the evolution and revolution of the knowledge*. In: 12TH INTERNATIONAL CONGRESS OF LOGIC, METHODOLOGY AND PHILOSOPHY OF SCIENCE, 2003. Oviedo. Anais... Oviedo: DLMPS - IUHPS, 2003.
- CORRÊA, A. C. *Sistema de Gestão para as IES com base no Balanced Scorecard como um Sistema Complexo Adaptativo: Uma proposta metodológica para a operacionalização do PDI*. In: VII Colóquio de Gestão Universitaria de América Del Sur, 2007, Mar Del Plata. Argentina. **Anais...** Mar Del Plata, 2007, 1a.
- CORREA, A. C., ERDMANN, R. H., MELO, P. A. de., RISSI, M., TECCHIO, E. L. *Mapa Estratégico para as IES fundamentado em um Sistema de Gestão Integrado: uma proposta metodológica para a implementação do PDI*. In: VIII Colóquio Internacional sobre Gestão Universitaria en América del Sur, 2008, Assunción. Paraguay. **Anais...** Universidade Tecnológica Intercontinental, 2008. p. 1-16.
- CRISÓSTOMO, V. L.; FREIRE, F. S.; BARBOSA, G. C. Análise dos indicadores de gestão das IFES e o desempenho discente no ENADE. **Avaliação**. Campinas; Sorocaba, SP, v. 16, n. 2, p. 317-344, jul. 2011.
- EMBRETSON, S. E.; REISE, S. P. **Item Response Theory for Psychologists**. New Jersey, USA: Lawrence Erlbaum Associates, 2000.
- FRANCISCO, R. **Aplicação da Teoria da Resposta ao Item (TRI) no Exame Nacional de Cursos (ENC) da Unicentro**. 2005. 144 f. Dissertação (Mestrado em Ciências) Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2005.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 1989.



- HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. **Fundamentals of item response theory**. Newbury Park, CA: Sage, 1991.
- KARINO, C. A., ANDRADE, D. F. Entenda a Teoria de Respostas ao Item (TRI), utilizada no Enem. **Nota Técnica**, 2010. Disponível em:
<http://www.senado.gov.br/comissoes/CE/AP/AP20101116_NotaTecnica_INEPMEC_TeoriaRespostasAoItem.pdf>. Acesso em fev. 2012.
- KLEIN, R. Uma re-análise dos resultados do PISA: problemas de comparabilidade. **Aval. Pol. Públ. Educ.**, Rio de Janeiro, v. 19, n. 73, p. 717-742, out./dez. 2011.
- KUDER, F. G.; RICHARDSON, M. W. The theory of test reliability. IN: Mehrens. William A. e Ebel, Robert L. **Principles of educational and psychological measurement**. Chicago: Rand McNally, 1967
- LORD, F. M. A theory of test scores (No. 7). **Psychometric Monograph**, 1952.
- LUDKE, M., ANDRÉ, M. E. D. **Pesquisa em Educação: abordagens qualitativas**. São Paulo: EPU, 1986.
- MASTERS, G. N. A Rasch model for partial credit scoring. **Psychometrika**, 47, 149-174, 1982.
- MINISTÉRIO DA EDUCAÇÃO, Diário Oficial da União. **Artigo 5 e parágrafo V do artigo 6 da lei 10.861, de 14 de Abril de 2004**. Seção 1, n.72; Brasília, p.3-4, 2004.
- _____, Diário Oficial da União. **Portaria Normativa Nº 40, DE 12 DE DEZEMBRO DE 2007**. Seção 1, Nº 249: Brasília, p 23, 2007.
- _____, Diário oficial da União. **Portaria Normativa Nº 6, de 14 de março de 2012**. Seção 1. Nº 52: Brasília, p 1-5, 2012.
- MURAKI, E. A generalized partial credit model: Application of an EM algorithm. **Applied Psychological Measurement**, 16, 159-176, 1992.
- NOGUEIRA, S. O. **ENADE: Análise de Itens de Formação Geral e de Estatística pela TRI**. 2008. 166 f. Dissertação (Mestrado em Psicologia) - Programa de Pós-Graduação Stricto Sensu em Psicologia, Universidade São Francisco, Itatiba, 2008.
- OLEA, J.; PONSODA, V.; REVUELTA, J.; BELCHI, J. Propiedades psicométricas de un test adaptativo informatizado de vocabulario inglés. **Estudios de Psicología**, n. 55, p. 61-73, 1996.
- OLIVEIRA, K. S. **Avaliação do exame nacional de desempenho do estudante pela teoria de resposta ao item**. 2006. 96 f. Dissertação (Mestrado em Psicologia) - Programa de Pós-Graduação Stricto Sensu em Psicologia, Universidade São Francisco, Itatiba, 2006.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Pisa 2003 Data Analysis Manual**. Paris, 2005. Disponível em:
<<http://www.oecd.org/dataoecd/53/22/35014883.pdf>>. Acesso em: ago. 2011.
- _____. **Pisa 2009 assessment framework: key competencies in reading, mathematics and science**. Paris, 2009. Disponível em:
<http://www.oecd.org/document/44/0,3746,en_2649_35845621_44455276_1_1_1_1,00.html#TOC>. Acesso em: ago. 2011.
- PRIMI, R., NUNES, C. H. S. S., SILVA, M. C. R., CARVALHO, L. F., MIGUEL, F. K. & VENDRAMINI, C. M. M. Aplicação da Teoria de Resposta ao Item na Interpretação das Notas do ENADE de Psicologia. **Revista de Educação AEC**, 38, 115-124, 2009.
- PRIMI, R.; CARVALHO, L. F.; MIGUEL, F. K.; SILVA, M. C. R. Análise do funcionamento diferencial dos itens do Exame Nacional do Estudante (ENADE) de psicologia de 2006. **Psico-USF**, v. 15, n. 3, p. 379-393, set./dez. 2010.



- POLIDORI, M. M., *et al.* Políticas de Avaliação da Educação Superior. **Educação e Realidade**. Porto Alegre, v.36,n.1,p.253-278, jan/abr, 2011.
- RASCH, G. **Probabilistic Models for Some Intelligence and Attainment Tests**. Copenhagen : Danish Institute for Educational Research, 1960.
- SAMEJIMA, F. A. Estimation of latent ability using a response pattern of graded scores. **Psychometric Monograph**, 17. 1969.
- SÃO PAULO, E.; MIRANDA, B. S.; MOREIRA NETO, J. G.; PAIXÃO, L. A. R. Aplicação do Modelo de Crédito Parcial Generalizado na Avaliação do Projeto Sesi – Por um Brasil Alfabetizado. **Revista Eletrônica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 5, n. 2e, p. 24-38, 2007.
- SCAGLIONE, V. L. T.; COSTA, M. N. Avaliação da educação superior e a gestão universitária: padrões de qualidade definidos pelas instituições de ensino superior, pelo MEC e pela sociedade, incluindo ENADE, IDD, CPG e IGC. **XI Congresso de Gestão Universitária “Gestão universitária, Cooperação Internacional e Compromisso Social”**. Florianópolis, 2011, 1CD-ROM.
- SCHWARTZMAN, S. **O “conceito preliminar” e as boas práticas de avaliação do ensino superior**. Brasília: ABMES, 2008.
- SISTEMA NACIONAL DE AVALIAÇÃO DA EDUCAÇÃO SUPERIOR (SINAES): da concepção à regulamentação / [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira]. 5. ed., revisada e ampliada. Brasília, 2009.
- VENDRAMINI, C. M. M.; SILVA, M. C.; CANEL, M. Análise de Itens de uma Prova de Raciocínio Estatístico. **Psicologia em Estudo**, Maringá, v. 9, n. 3, p. 487-498, set./dez 2004.
- WRIGHT, B. D. **Sample-free test calibration and person measurement**. Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N. J. Educational Testing Service, 1968.