

---

# Avaliação Informatizada Adaptativa do ENADE Pelo MOODLE: Evidências de Validade

## Adaptive Computerization Assessment of ENADE by MOODLE: Evidence of Validity

---

Lucio Ferreira Santana

Centro Universitário FIEO

Daniel Bartholomeu

Centro Universitário Salesiano – UNISAL – Americana.

José Maria Montiel

Centro Universitário FIEO – UNIFIEO/SP.

Gleiber Couto

Universidade Federal de Goiás no Campus de Catalão - UFG/CAC.

Arthur Almeida Berberian

Centro Universitário FIEO

Fernando Pessoto

Centro Universitário Salesiano – UNISAL – Americana.

**Resumo:** Testes informatizados adaptativos adquirem relevância quanto às possibilidades de estudos. Este estudo teve como objetivo relacionar dados de uma avaliação do ENADE do Curso do Direito. Participaram 92 pessoas. Foi utilizado dois instrumentos de medida de múltipla escolha, um em papel e lápis e outro adaptado ao modelo informatizado adaptativo em ambiente *MOODLE*. Os resultados demonstraram que os itens estiveram dentro dos parâmetros de ajuste ao modelo investigado. A correlação entre os níveis de dificuldade dos itens nas duas formas, revelaram que os parâmetros dos itens se mantiveram os mesmos entre uma e outra forma de aplicação. Considera-se que estudos que se propõem a analisar os itens, recorrendo a TRI como meio de análise do modelo de Rasch, permitam aferir evidência de validade de construto. Sugere-se que novos estudos a respeito do ENADE, em modelo informatizado adaptativo, pois o ambiente *MOODLE* demonstrou ser um meio útil e eficaz de avaliação.

**Palavras-chave:** Avaliação Educacional. TRI. Modelo de Rasch.

**Abstract:** Adaptive computer tests acquire relevance as to the possibilities of studies. This study aimed to relate data from an ENADE evaluation of the Law Course. 92 people participated. Two multi-choice measurement instruments were used, one in paper and pencil and another adapted to the adaptive computer model in a *MOODLE* environment. The results showed that the items were within the fit parameters to the investigated model. The correlation between the difficulty levels of the items in the two forms, revealed that the parameters of the items remained the same between one and another form of application. It is considered that studies that propose to analyze the items, using TRI as a means of analysis of the Rasch model, allow to gauge evidence of construct validity. It is suggested that new studies regarding ENADE, in an adaptive computer model, since the *MOODLE* environment has proved to be a useful and effective means of assessment.

**Keywords:** Educational Assessment. TRI. Rasch Model.

SANTANA, Lucio Ferreira; BARTHOLOMEU, Daniel; MONTIEL, José Maria; COUTO, Gleiber; BERBERIAN, Arthur Almeida; POSSOTO, Fernando. Avaliação informatizada adaptativa do ENADE pelo MOODLE: evidências de validade, São Paulo, v. 20, n. 2, p. 222-238, mai./ago. 2017.

## 1 Introdução

No que se refere especificamente ao sistema educacional brasileiro, o Ministério da Educação (MEC) tem dedicado na última década um cuidado especial ao processo de avaliação deste contexto, por exemplo as inúmeras provas desenvolvidas para se aferir a qualidade do ensino em diversos níveis, desde o ensino fundamental (Sistema de Avaliação do Ensino Básico – SAEB), médio (Exame Nacional do Ensino Médio – ENEM) e superior (Exame Nacional de Cursos – ENC). Para esse último, mais recentemente foi idealizado o Sistema Nacional de Avaliação do Ensino Superior (SINAES) (informações disponíveis em [www.inep.gov.br](http://www.inep.gov.br)). O objetivo fundamental do SINAES é a avaliação das Instituições de Ensino Superior (IES), integrando diferentes dimensões cujos elementos centrais são, a auto-avaliação orientada, a avaliação da instituição, avaliação dos cursos de graduação e o ENADE. Assim, é possível analisar de um ponto de vista mais extenso, a instituição como um todo, de uma expectativa mais reduzida, restringindo-se ao desempenho do estudante por meio do ENADE.

Uma qualidade peculiar desses sistemas de avaliação em larga escala para instituições educacionais incide na medida de qualidade das instituições, definida por indicadores do quanto cada instituição colaborou para o desenvolvimento de seus alunos em termos de habilidades acadêmicas, competências profissionais e conhecimento. A questão principal que radica dessa discussão é a validade das medidas utilizadas na aferição dessa qualidade ou, em outros termos, da eficácia escolar. Tal medida tem sido nomeada, na literatura científica, como medida de “valor agregado”, acentuado como “uma medida do avanço médio do aluno, durante o período de tempo em que ele é exposto a determinado ambiente educativo, comparativamente com o que obteria em outras escolas da amostra” (Ferrão, 2003, p.16; Jesus & Laros, 2004; Raudenbush, 2004b; Rubin, Stuart & Zanutto, 2004; Soares, Ribeiro & Castro, 2001).

A despeito de apresentar uma definição operacional, a validade de tais avaliações deve ser estabelecida para cada sistema de avaliação designado aos diferentes níveis de educação. Nesse contexto, estudos devem ser delineados no sentido de se testar as interpretações feitas com base nos instrumentos de medida, averiguando se estão funcionando adequadamente, conforme se foi planejado. Assim, não só o instrumento, mas os delineamentos de coleta de informação são testados. Na psicometria, tais estudos são abordados dentro do julgamento de validade. Além da preparação adequada de uma prova, para que essa apresente êxito na avaliação (cumpra seus objetivos), é fundamental atestar-se a qualidade científica das próprias (American Educational Research Association AERA, American Psychological Association – APA, National Council on Measurement in Education – NCME, 1999; Urbina, 2004). As informações providas por esses estudos agregam dados respeitáveis no aprimoramento dos sistemas de avaliação (nesse caso em larga escala), já que aprofundam o debate da legalidade da medida utilizada e recomendam algumas limitações das inferências feitas com base nos resultados obtidos por tal sistema, recomendando novas formas de se avaliar com vistas a aumentar sua eficácia. Apesar disso, evidenciam-se poucos trabalhos dessa natureza feitos com as provas do

---

MEC, fato que se torna inquietante ao se considerar o impacto social de tais sistemas bem como de seus resultados.

Dentre as possibilidades de estudos de validade para o ENADE, a investigação de alternativas viáveis de administração e redução da prova são particularmente interessantes, pois tendem a reduzir seu custo e tornar mais fiáveis as avaliações. Neste contexto, os Testes informatizados adaptativos adquirem relevância, já que possibilitam uma avaliação reduzida e com estimativas boas das habilidades das pessoas. Os TAIs são administrados via computador. Testes adaptativos (TA) são aqueles cuja aplicação dos itens se adapta ao sujeito que está respondendo ao teste, que busca encontrar um teste ótimo para cada examinando, ou seja, um teste aplicado com o mínimo de perguntas possível e que aproximar-se para o mesmo resultado de um teste comum. Para isso, é necessário estimar a proficiência do indivíduo de forma iterativamente durante a administração do teste e, assim, só precisam ser selecionados os itens que mensurem eficientemente a proficiência do examinado. A noção básica de um TA é imitar automaticamente o que um sábio examinador faria, ou seja, um TAI tem por intenção administrar itens, de um banco de itens previamente calibrados, que satisfaçam ao nível de capacidade do examinando. Geralmente, esses itens são escolhidos de acordo com o modelo da Teoria de Resposta ao Item (TRI), que é adotado para descrever o comportamento da resposta do indivíduo. Ao contrário dos testes papel-e-lapis, diferentes examinados podem receber diferentes testes de tamanhos variados. (COSTA, 2009 apud WAINER et al., 1990, p. 3).

Outra vantagem de um TAI é a flexibilidade para aplicação de baterias de testes. Ao oposto do exame tradicional, um TAI não determina que todos os examinados façam a prova ao mesmo tempo. Em uma bateria de testes, por exemplo, o examinando que finalizar a prova pode passar diretamente para a prova seguinte sem precisar esperar os outros. Além disso, o aplicador do teste pode fornecer as instruções do teste de forma virtual. Para Moreira (2011, p. 98), "a maior vantagem de um CAT em relação a um teste tradicional consiste na maior eficiência na estimação do nível de habilidade do respondente com um menor número de itens do que os testes tradicionais". Para atingir esse objetivo, dois aspectos básicos são necessários: o método de estimação das habilidades dos respondentes e o critério de seleção dos itens. Ainda Olea et al. (2004), descreve que a maioria dos CATs utiliza estratégias como, um critério de partida, para determinar o primeiro item a ser apresentado; um método estatístico (geralmente Bayesiano ou Máxima Verossimilhança) para estimar a proficiência do indivíduo e a precisão associada; um procedimento para selecionar o próximo item; um critério para finalizar o teste.

Seguindo tais apontamentos quando se opta pela utilização do TAI, deve-se primeiramente se preocupar com qual será o tipo de ambiente ou sistema gerenciador de conteúdo que devemos utilizar para nos auxiliar. Ainda é possível optar pelo Ambiente Virtual de Aprendizagem (AVA) ao qual possui diversos recursos e ferramenta, facilitadoras para a gerência centralizada de novas implementações de conteúdos, usuários, banco de questões e os já existentes de maneira automatizada; publicação de conteúdo rapidamente; permitir reuso ou duplicação desses conteúdos de maneira que não necessite a criação desse novamente a fim de favorecer a utilização mais rápida por parte dos alunos (Ellis, 2009). Ainda o mesmo autor

---

descreve uma metodologia, chamada Metodologia de Avaliação de Ambientes Virtuais de Aprendizagem (MAAVA), destinada à avaliação de um ambiente configurado para um curso, fundamentado na importância de atrelar os objetivos educativos das ferramentas do AVA aos objetivos pedagógicos, dos quais se pretende que os alunos alcancem ao final de curso e/ou disciplina e; na importância de ter um AVA com boa usabilidade, comunicabilidade e sociabilidade, para não afetar o atendimento aos objetivos pedagógicos.

Em continuidade, a literatura tem apontado que para essa transição de paradigma, o surgimento de novos sistemas de gestão de conteúdos tem sido muito recorrente e pertinente, por exemplo, (CMS – *content management system*), em código de fonte aberta (*open source*), como é o caso do MOODLE (<http://www.MOODLE.org>), lançado em 1999, utilizado também para ambientes educativos e colaborativos *online*, com intuito de fomentar um espaço de colaboração, onde seus usuários poderiam intercambiar saberes, experimentando, criando novas interfaces para o ambiente em grande comunidade aberta. Quando utilizados em educação, tendem a ser designados como sistemas (ou plataformas) em linha de gestão da aprendizagem e de trabalho colaborativo (<http://MOODLE.fct.unl.pt/>), possui um sistema de julgamento que auxilia o profissional a configurar e a aplicar a avaliação. As questões são guardadas por categorias em uma base de dados e podem ser reutilizadas em outros questionários e avaliações e até mesmo em outros cursos ou disciplinas como no caso da educação.

PULINO (2005) menciona como vantagem, em relação a outros sistemas gerenciadores de cursos, que o MOODLE é um software livre que apresenta todas as funcionalidades e os objetivos educacionais requerido em um LMS. Com o advento da informática e a popularização dos computadores, versões informatizadas de vários testes do tipo “papel e lápis” passaram a ser desenvolvidas. Contudo, na grande maioria dos casos, o que foi feito, foi apenas uma mudança nos processos de aplicação do teste, que passou a ser disponibilizado na tela do computador, e de resolução do teste, que passou a ser efetivado por intermédio do computador. De tal modo, um teste informatizado é o mesmo teste tradicional do tipo “papel e lápis” efetivado por meio do computador. Essa foi considerada a primeira geração de testes informatizados, segundo Olea e Hontangas (1999). Uma alternativa para as avaliações do tipo “papel e lápis” tradicional ou informatizada é a emprego dos chamados TAI's.

Esta alternativa é particularmente viável em investigações de competências e de larga escala, considerando a maior rapidez e eficiência desta modalidade de avaliação, bem como praticidade na forma de coletar os dados. No caso do Brasil, provas governamentais como o ENADE tendem a ter alto impacto na definição de políticas públicas educacionais no ensino superior e, apesar disso, muito ainda resta a ser feito com estas provas, sobretudo no que se refere à novos estudos de validade e de viabilidade metodológica (Bartholomeu, 2011). Na tentativa de buscas por trabalhos envolvendo metodologias de avaliação de AVAs, é notória a carência de pesquisas que venham a sugerir um formato de avaliação.

Como mencionado anteriormente bem como em decorrência da carência de estudos com esta temática, existem vantagens de se ter um método adaptativo de avaliação do ENADE como sua praticidade, já que pode ser respondida por qualquer computador ou aparelho de

celular, bem como que o modelo adaptativo otimiza o uso dos itens, tendo maior precisão na estimativa das habilidades reais dos estudantes e favorecendo melhores avaliações da qualidade dos cursos, melhorando estas avaliações aplicadas no contexto escolar, otimizando o tempo dos estudantes ao responder o questionário, tornando ele menos cansativo e mais preciso. Ao mesmo tempo, o uso do *MOODLE* para este fim pode facilitar para as IESs que desejem empregar um modelo gratuito para se obter as estimativas de habilidades dos seus estudantes nesta prova, usando de recursos gratuitos para analisar e contemplar aspectos públicos.

Seguindo tais postulados, esse trabalho se coloca no sentido de se propor uma metodologia para se estudar evidências de validade de exames de larga escala produzidos no INEP, individualmente, o ENADE. Tal exame é mesclado por questões de formação geral e específica e tem o objetivo de aferir as habilidades acadêmicas e competências profissionais que estudantes ingressantes e concluintes de Instituições de Ensino Superior apresentam. Assim, dadas as vantagens que as TAIs apresentam e a necessidade de novas investigações de validade do ENADE este trabalho se insere procurando evidências de validade para o ENADE “Direito” a partir de sua estrutura interna e fazendo a adaptação informatizada e adaptativa deste instrumento a partir do software livre *MOODLE* como uma alternativa barata para esta finalidade. Neste sentido, será que o sistema adaptativo do *MOODLE* é melhor do que a aplicação de papel e lápis do ENADE Direito? Esta é a pergunta de pesquisa que este trabalho se atém. Com isso, o objetivo deste estudo é relacionar dados de uma avaliação do ENADE do Curso do Direito, informatizada adaptativa e de papel e lápis visando evidências de validade para este teste.

## **2 Método**

### **2.1 Participantes**

Foram participantes 92 adultos com idade variando de 18 a 70 anos do ensino superior do 1º e 6º semestre do curso de Direito de um Centro Universitário da Grande São Paulo, com 64,13 % dos participantes sendo do sexo feminino.

### **2.2 Instrumentos**

#### **2.2.1 ENADE Direito – INEP**

O ENADE apresenta 40 questões que avaliam conhecimentos, competências profissionais e habilidades acadêmicas de estudantes do Ensino Superior. Desse teste, 10 questões são relativas ao componente de formação geral, e as 30 restantes são itens específicos da área do curso em avaliação. Cada questão respondida corretamente receberá um ponto. O estudo utilizou 26 questões de Competência Específico/Múltipla Escolha, extraídas do questionário do ENADE. As demais questões, que são dissertativas e conhecimento geral, não foram utilizadas, por não contemplarem o objetivo do presente estudo. Sua aplicação foi

realizada via “papel e lápis” definição atribuída por sistema de avaliação tradicional neste tipo de ambiente.

### **2.2.2 Avaliação do ENADE Direito informatizado adaptativa**

Cabe ressaltar que este instrumento é feito em papel e lápis e suas questões foram inseridas no MOODLE como questionário adaptativo, instalado no módulo de atividade. Foi utilizado nesse estudo o ambiente MOODLE versão 2.7+ (Build:20140522), com plugin Adaptive Quiz versão:2015060900, ao qual foram adequados os questionários e disponibilizados aos respondentes no ambiente MOODLE.

O Modulo de atividade Adaptive Quiz é um novo módulo do MOODLE, que permite criar, modificar e administrar questionário no estilo adaptativo. Um algoritmo é utilizado para estimar a capacidade do participante nas escolhas das perguntas dentro de certo número de erros, do nível mais fácil até o mais difícil. Os parâmetros para utilização do algoritmo são configuráveis, dentro outros parâmetros são necessários a utilização de categorias do banco de questões e tags (marcadores), a fim de determinar as questões que serão exibidas durante a tentativa do participante individualmente. Entretanto as categorias de questões podem atribuir o conjunto de perguntas do *Adaptive Quiz* e as *tags* são utilizadas para atribuir os níveis de dificuldades as perguntas. Foi verificar a implementação deste algoritmo, nos arquivos de *plugin* do *MOODLE ADAPTIVE QUIZ*, e pode ser encontrado no arquivo salvo no diretório “..\MOODLE\mod\adaptivequiz\catalogo.class.php”.

O conjunto de parâmetros permite que o aplicador possa determinar os níveis de dificuldade das perguntas. Inicialmente o nível para primeira pergunta é selecionado nos parâmetros e, em seguida, o algoritmo seleciona o próximo nível de dificuldade ou retrocede o nível dependendo da resposta certa ou errada e sim por diante até o final do questionário. No início de uma tentativa o conjunto de perguntas é indeterminado. A quantidade de perguntas que serão apresentadas está entre o número mínimo e máximo de perguntas determinadas nos parâmetros. Quando o número mínimo de perguntas respondidas for atingido e para que aja a sua parada por completo, o participante do teste tem que atingir o nível de capacidade de precisão requerida ou atingir o número máximo de perguntas respondidas. Assim, atendendo a estratégia sugerida por Olea et al. (2004).

### **2.3 Procedimentos**

Todos os instrumentos de avaliação anteriormente descritos foram aplicados em sessões previamente agendadas junto aos participantes e à Instituição, após consentimento dos responsáveis pela Instituição e pelos alunos, e assentimento desses últimos. Foram utilizados os laboratórios de informática da própria Instituição com agendamento prévio para a aplicação do teste informatizado e teste no papel, sendo que o pesquisador levou lápis, caneta, papéis e folhas de teste, bem como termos de consentimento. As avaliações respeitaram as atividades cotidianas, de modo que não comprometeu as atividades estabelecidas pela Instituição e foram realizadas durante o período de aulas (noturno), em horário cedido pelo

professor, usou em média noventa minutos para aplicação. Os dois instrumentos foram aplicados um em seguida do outro, iniciando pelo teste em lápis e papel e o de computador em seguida, durando 45 minutos cada um. Cabe mencionar que o estudo foi aprovado pelo comitê de ética em pesquisa do Centro Universitário FIEO – UNIFIEO/SP.

Os procedimentos de coletas de dados ocorreram em 2 etapas; na primeira foi aplicado o instrumento de avaliação “ENADE”, em formato de papel e lápis em grupo; e num segundo momento, a avaliação informatizada em computadores nos laboratórios cedidos pela Instituição. Nesse caso, ocorreu em 40 máquinas ao mesmo tempo (1 para cada avaliando). Assim, a aplicação ocorreu em grupos compostos de 40 (quarenta) pessoas cada. Ambas as etapas tiveram duração de 45 minutos cada.

## **2.4 Resultados E Discussão**

Para a análise dos dados, inicialmente, os dados obtidos serão tabulados em planilha. Em seguida serão feitas as análises descritivas das variáveis obtidas no ENADE informatizado e em papel e lápis. Esses instrumentos serão submetidos a procedimentos de análise de sua estrutura interna, visando a validá-la por duas análises estatísticas a saber: análise fatorial confirmatória, e TRI pelo modelo de Rasch. Em seguida, serão submetidos a análise de correlação de Pearson, visando a identificar se as duas formas do teste (informatizada e papel e lápis) são equivalentes (paralelas). Cabe mencionar que os resultados serão apresentados em 2 (duas) etapas conforme a seguir.

### **2.4.1 Etapa 1. Estimação dos parâmetros dos itens do ENADE Direito 2006**

Para estimação dos parâmetros, foram usados dados de um estudo prévio do ENADE da prova de Direito de 2006, disponibilizados no site do INEP. Esse curso contou com 4491 sujeitos com idades de 21 a 63 anos com a maioria deles (74%) com até 23 anos. Quanto ao sexo, 57,5% eram mulheres. Ainda, 58,7% estudavam no turno noturno. O nível de escolaridade dos pais em sua maioria (cerca de 50%) foi nível superior. A maior parte das pessoas eram provenientes de famílias que apresentavam até 9 salários. Cerca de 53% estudaram em instituições privadas de ensino e 75% do total de instituições estudadas apresentavam incentivo para pesquisa. Em relação ao conceito das instituições que os sujeitos eram provenientes, 43% apresentou conceito 3.

Nesse estudo foi estabelecido os parâmetros ao modelo informatizado adaptativo dos itens dos componentes específicos da prova de ENADE direito de 2006, esses parâmetros podem ser observados na Tabela 1. Percebe-se o nível de dificuldade dos itens da prova, bem como índices de ajuste de infit e outfit que demonstraram bons ajustes não excedendo parâmetros estabelecidos por Linacre (2002) de 1,5. Assim, a prova demonstrou bom ajuste ao modelo de Rasch. Os níveis de dificuldade foram classificados e inseridos no MOODLE a partir da Tabela 2 seguindo a especificação deste software que não aceita valores negativos. Ainda cabe mencionar que as cores apresentadas tanto na tabela 1, como na tabela 2, tem como propósito a compreensão dos níveis de dificuldades ao se inserir cada questão, sendo apenas

ilustrativa. Após a estimação dos parâmetros, estes foram incluídos na base do *MOODLE* pelo método adaptativo e foram rodados nessa avaliação informatizada e também foram feito escalonamentos dos parâmetros, por limitação do aplicativo, por não receber valores de estimação de itens negativos. Vale ressaltar que a prova foi aplicada no formato de papel e lápis e informatizado adaptativo em estudantes deste curso para serem correlacionadas as notas no Estudo 2 e estimadas a quantidade de informação fornecida por um e outro formato visando estabelecer pontos fortes e fracos de cada modalidade de aplicação. Para maiores detalhes sobre essas estimativas veja os anexos.

Tabela 1. ENADE 2006 - DIREITO.

TABLE 14.1 ENADE 2006 - DIREITO												
ZOU088WS.TXT Sep 3 18:21 2008												
INPUT: 92413 PS 68 IS MEASURED: 76557 PS 26 IS 2 CATS WINSTEPS 3.61.1												
P: REALSEP.: 1.31 REL.: .63 ... I: REALSEP.: 89.87 REL.: 1.00												
I STATISTICS: ENTRY ORDER												
ENTRY	RAW		MODEL		INFIT		OUTFIT		PTME	EXACT		
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	A	OBS%	EXP%	I
CORR.												
17	27855	76209	.09	.01	1.06	9.9	1.07	9.9	.26	64.9	68.0	item_CE_11
18	36248	76186	-.43	.01	1.02	9.9	1.03	8.0	.31	62.6	64.2	item_CE_12
19	25725	76166	.23	.01	1.06	9.9	1.08	9.9	.25	67.4	69.6	item_CE_13
20	18955	76008	.70	.01	1.05	9.9	1.09	9.9	.24	75.2	76.2	item_CE_14
21	51887	76158	-1.39	.01	.98	-6.7	.99	-1.8	.34	71.5	70.4	item_CE_15
22	37385	76102	-.50	.01	.95	-9.9	.93	-9.9	.41	67.5	64.0	item_CE_16
23	13214	75573	1.20	.01	1.10	9.9	1.27	9.9	.13	82.3	82.8	item_CE_17
24	37192	75920	-.49	.01	.91	-9.9	.90	-9.9	.44	69.2	64.1	item_CE_18
25	40052	76129	-.65	.01	.99	-6.1	.99	-4.6	.36	65.3	64.1	item_CE_19
26	37629	75675	-.52	.01	.94	-9.9	.93	-9.9	.41	67.5	64.0	item_CE_20
27	27988	75315	.06	.01	1.03	9.9	1.05	9.9	.29	66.6	67.7	item_CE_21
28	12609	75100	1.26	.01	1.14	9.9	1.40	9.9	.07	82.8	83.4	item_CE_22
29	19118	75851	.69	.01	1.00	-1.1	1.01	2.3	.31	76.5	75.9	item_CE_23
30	20270	75553	.59	.01	1.08	9.9	1.16	9.9	.20	73.9	74.6	item_CE_24
31	19192	75868	.68	.01	1.03	6.0	1.07	9.9	.27	75.8	75.9	item_CE_25
32	12737	75908	1.25	.01	1.00	.6	1.09	9.6	.25	83.8	83.5	item_CE_26
33	32099	75886	-.18	.01	.93	-9.9	.91	-9.9	.43	70.1	65.4	item_CE_27
34	37886	75718	-.54	.01	.94	-9.9	.93	-9.9	.41	66.9	64.0	item_CE_28
35	32355	75825	-.20	.01	.97	-9.9	.97	-8.4	.37	67.1	65.3	item_CE_29
36	24481	75299	.29	.01	1.02	6.4	1.04	9.3	.29	70.3	70.5	item_CE_30
37	44870	75721	-.96	.01	.92	-9.9	.89	-9.9	.44	69.4	65.7	item_CE_31
38	17231	75762	.84	.01	1.03	5.6	1.08	9.9	.26	78.0	78.0	item_CE_32
39	18547	75622	.73	.01	.98	-4.8	.99	-1.5	.33	77.2	76.5	item_CE_33
40	40662	75752	-.70	.01	.98	-9.9	.97	-8.4	.37	65.7	64.2	item_CE_34
41	39064	75700	-.61	.01	.94	-9.9	.93	-9.9	.41	67.5	64.1	item_CE_35
42	52277	75863	-1.43	.01	.95	-9.9	.94	-9.9	.37	72.5	71.0	item_CE_36
MEAN	29904.9	75802.6	.00	.01	1.00	-.8	1.03	.6		71.5	70.5	
S.D.	11526.6	279.4	.77	.00	.06	8.7	.11	8.9		5.8	6.5	
TABLE 14.3 ENADE 2006 - DIREITO												
ZOU088WS.TXT Sep 3 18:21 2008												
INPUT: 92413 PS 68 IS MEASURED: 76557 PS 26 IS 2 CATS WINSTEPS 3.61.1												

Tabela 2 Agrupamento por itens de dificuldade

Level 1	Dificuldata -2	adpq_1	Qtde Itens 6
level 2	Dificuldata -1	adpq_2	Qtde Itens 6
Level 3	Neutro	adpq_3	Qtde Itens 2
Level 4	Dificuldade 1	adpq_4	Qtde itens 6
Level 5	Dificuldade 2	adpq_5	Qtde Itens 6

### 2.4.2 Etapa 2. Resultados da aplicação do ENADE – Direito em papel e lápis e informatizado adaptativo

Os indicadores do ENADE no modelo informatizado adaptativo e papel e lápis para os Alunos do Ensino Superior foram investigados quanto à sua adequação ao modelo de Rasch. Tal modelo considera somente o parâmetro de dificuldade dos itens para se estimar a adequação deles a uma escala, partindo do pressuposto da TRI. Foram demonstrados abaixo. O ajuste do modelo leva em consideração a estrutura geral. Se você remover alguns itens "desajustados" e reprocessar a análise TRI a distribuição vai ser mais normal, mas ainda haverá itens com altos resíduos. Devido a isso, a abordagem de "modo de ajuste" não é uma boa maneira de examinar item ajustados. Em caso afirmativo, qual é a ferramenta adequada para verificar o ajuste do item? É o "IN.MSQ" e "OUT.MSQ" que significam "*InFit*" e "*outfit*." Foi utilizado nessa abordagem o "*outfit*".

### 2.4.3 ENADE papel e lápis

A precisão fornecida por esse modelo para os itens resultou em um coeficiente de 0,26 para os itens e 0,48 para as pessoas, o que favorece a interpretação de que as pessoas, ao responderem os itens, tendem a informar sobre os conhecimentos com maior precisão. Todavia, os itens não foram capazes de avaliar os conhecimentos dos alunos com a mesma magnitude. O erro médio de medida para itens e pessoas, respectivamente foi de 0,04 (DP= 0,01) e 0,08 (DP = 0,01). Essas e outras informações constam da Tabela 3.

Tabela 3 - Parâmetros de Ajustes de Itens e das Pessoas (N = 184) papel e lápis

Parâmetros	Itens				Pessoas			
	Dificuldade	<i>Infit</i>	<i>Outfit</i>	Erro	Habilidade	<i>Infit</i>	<i>Outfit</i>	Erro
<b>Média</b>	0,00	1,00	1,00	0,04	-1,07	1,00	1,00	0,08
<b>DP</b>	0,74	0,06	0,14	0,01	0,67	0,14	0,24	0,01
<b>Máximo</b>	1,31	1,09	1,44		0,36	1,31	1,68	
<b>Mínimo</b>	-1,65	0,87	0,79		-2,72	0,74	0,32	
<b>Precisão</b>	0,26				0,48			

Os itens forneceram média de *outfit* de 1,00 (DP= 0,14), indicando um bom ajuste dos itens nesse aspecto, de forma geral. Analisando mais detidamente esses dados, observa-se que a variação de *outfit* esteve entre os valores 1,44-0,79, sugerindo que todos os itens estiveram dentro dos parâmetros de ajuste ao modelo. Nenhum item apresentou índices acima de 1,50 (Bond & Fox, 2001). O procedimento de análise empregado para se interpretar as pontuações gerais brutas produzidas com o ENADE (provenientes da soma dos escores nos fatores) foi o uso do mapa do construto ou mapa de respostas esperadas, baseado na TRI, mais especificamente, no modelo de Rasch.

### 2.4.4 ENADE Informatizado e Adaptativo

A precisão fornecida por esse modelo para os itens resultou em um coeficiente de 0,40 para os itens e 0,81 para as pessoas, o que favorece a interpretação de que as pessoas, ao responderem os itens, tendem a informar sobre os conhecimentos com maior precisão. Todavia, os itens não foram capazes de avaliar os conhecimentos dos alunos com a mesma magnitude. O erro médio de medida para itens e pessoas, respectivamente foi de 0,70 (DP= 0,01) e 0,40 (DP = 0,01). Essas e outras informações constam da Tabela 4.

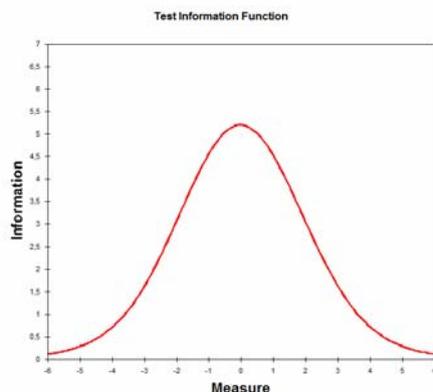
Tabela 4 Parâmetros de Ajustes de Itens e das Pessoas (N = 184) adaptativo

Parâmetros	Itens				Pessoas			
	Dificuldade	Infit	Outfit	Erro	Habilidade	Infit	Outfit	Erro
<b>Média</b>	0,00	0,99	1,03	0,70	-1,05	0,99	0,95	0,04
<b>DP</b>	1,06	0,16	0,36	0,01	0,90	0,21	0,33	0,01
<b>Máximo</b>	1,96	1,36	2,43		1,38	1,50	2,40	
<b>Mínimo</b>	-1,97	0,25	0,54		-3,22	0,58	0,49	
<b>Precisão</b>	0,40				0,81			

Os itens forneceram média de *outfit* de 1,03 (DP= 0,36), indicando um bom ajuste dos itens nesse aspecto, de forma geral. Analisando mais detidamente esses dados, observa-se que a variação de *outfit* esteve entre os valores 2,43-0,54, sugerindo que nem todos os itens estiveram dentro dos parâmetros de ajuste ao modelo. Alguns itens apresentaram índices acima de 1,50 (Bond & Fox, 2001). Isso revela que para a análise de casos extremos há mais discordância do modelo para o dado empírico, em outras palavras, pessoas que deveriam acertar itens mais fáceis ou mais difíceis não acertam, não é tão preciso para pessoas com resultados extremos. Provavelmente isso decorre pelo fato de que nem todos os sujeitos responderam ao inventário por completo. O fato é que ele é mais preciso para avaliar pessoas com habilidades medianas. Ao mesmo tempo não houveram incrementos de precisão significativos entre as duas formas.

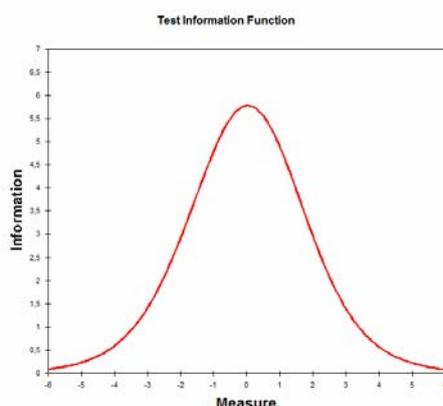
Em uma situação dicotômica, a variância é  $p(1-p)$  enquanto  $p$  = valor do parâmetro. Com base nos valores dos parâmetros do item, pode-se calcular e traçar a função das informações de teste (TIF) para os itens como mostrado na Figura 1 e 2.

Figura 1. Test Information Function Teste Adaptativo.



Fonte: SANTANA, 2016

Figura 2. Test Information Function Teste papel e lápis.



Fonte: SANTANA, 2016

Conforme apresentado, é possível observar as duas Figuras, que estas TIFs são diferentes entre si. No teste adaptativo, a maioria das informações é dispersa em relação ao  $\theta$  zero, isso nos indica que os itens são menos precisos, enquanto a quantidade de informação nas extremidades inferiores são as mesmas, podemos notar que cai lentamente. No teste papel e lápis, a maioria das informações é centralizada em relação ao  $\theta$  zero, isso nos indica que os itens são mais precisos, enquanto a quantidade de informação nas extremidades inferiores também são as mesmas, podemos notar que cai rapidamente. A Função de Informações do Item (IIF) é um pico em algum momento que mostra a precisão desigual ao longo da escala, tal teste seria melhor para estimar a capacidade de armazenamento dos investigados cuja habilidade é cair perto do pico do teste de informação. Em alguns testes, a TIF são bastante plana na região central da escala. Tais testes estimam alguma gama de capacidade de armazenamento com quase igual precisão e fora desta faixa com menos precisão.

A TIF é simplesmente a soma de todos os IIFs no teste. Enquanto IIF pode nos expor a informação e a precisão de um determinado parâmetro de item, o TIF pode nos expor a mesma coisa no nível de exame. Quando há mais de um formulário alternativo para o mesmo exame, o

TIF pode ser usado para equilibrar formulários alternativos. O objetivo é fazer com que todas as formas alternativas tenham os mesmos valores de TIF em todos os níveis de  $\theta$ . A correlação entre os níveis de dificuldade dos itens nas duas formas foi 0,82, revelando que os parâmetros dos itens se mantiveram os mesmos entre uma e outra forma de aplicação. Já a habilidade das pessoas teve correlação de 0,40, sugerindo que os dois formatos tenderam a avaliar as pessoas de maneira diferente. Dentre os fatores que podem explicar este fato, há que se considerar que os algoritmos de parada e seleção dos itens no *MOODLE* são feitos desconsiderando valores negativos (mas somente categorias dos níveis de dificuldade dos itens). Além disso, a redução do erro padrão em casos extremos tardou a acontecer, elevando os valores de *outfit* de itens e principalmente de pessoas no modelo informatizado, sendo esta uma limitação deste modelo no *MOODLE* que poderia ser reavaliada, sobretudo com um banco de itens pequeno como este ( $n=26$ ). Apesar disso, a curva de informação do modelo informatizado apresentou maior amplitude, possibilitando uma melhor análise e interpretação das habilidades das pessoas em um intervalo maior. O aumento no número de itens favoreceria uma estimativa destas habilidades com maior precisão, melhorando níveis de precisão e problemas de *outfit* identificados.

Há que se ressaltar também que somente o item 13 apresentou funcionamento diferencial significativo entre formas do teste, revelando que na forma de papel e lápis ele foi mais difícil. Halkits (1993) apresenta um CAT para medir a competência dos estudantes de enfermagem em três áreas: cálculo, princípio de administração de drogas e efeitos dos medicamentos. Este modelo foi apresentado como uma possibilidade explicativa para os dados ora obtidos no *MOODLE*. Construiu um banco de itens para cada área de conhecimento, calibrando suas dificuldades conforme o de papel-e-lápis anterior mente administrado a 4496 examinados. Como administrar um teste-tomador, uma estimativa de capacidade inicial (pseudo-Bayesiano) é fornecido pela atribuição a cada participante em dois itens fictícios com a dificuldade média,  $D^0$ , do banco de subitens. Assim estimativa de capacidade inicial de cada aluno é a dificuldade do item média.

O primeiro item para resposta é selecionado aleatoriamente perto 0,5 logits menos do que a capacidade de estimativa inicial. Isso produz uma chance de 62% de sucesso de acerto no item, ao qual o participante pode não estar familiarizado com o CAT, produzindo assim à oportunidade de exatidão de sucesso em todo o CAT. A seleção dos próximos itens por ser aleatório melhora na segurança do teste, impedindo que os participantes tenham testes idênticos. Após a resposta do primeiro item, é feito o cálculo da média de competência e erro padrão são estimados. Mais uma vez, um item é selecionado entre os que estão perto de 0,5 logits mais fácil do que a competência estimada. De, pois que este item é respondido, a medida de competência é novamente revista e mais um item é selecionado e administrado, e o processo tem seu prosseguimento.

A medida de erro padrão adotada também não foi restritiva o suficiente para alguns sujeitos, sendo que estes responderam a uma quantidade maior de itens. Talvez a quantidade de itens no banco tenha sido pequena demais sendo necessária uma versão com mais itens para novas estimativas do erro padrão que sejam mais adequadas para uma parte dos

participantes. Isso pode estar associado também à maior quantidade de desajustes das pessoas na versão informatizada em comparação à papel e lápis. Apesar disso, a versão informatizada via *MOODLE* foi capaz de fornecer uma quantidade maior de informação com uma quantidade menor de itens de modo geral, fato este que, se associado com a baixa correlação dos  $\theta$  estimados via informatizada e papel e lápis indicaria a viabilidade do modelo informatizado adaptativo.

Visualizando a função de erro padrão, note que depende apenas da diferença entre o número de respostas certas e erradas e o número total de respostas, não em outras características tais como que a igualdade entre as resposta certas e erradas, para um número fixo de perguntas, o erro-padrão será menor. Conforme discutido em Wright (1988), a fórmula para calcular a medida de capacidade é dada por: Medida de capacidade =  $H/L + \ln(R/W)$ , onde  $H$  é a soma de tudo questionar dificuldades respondidas,  $L$  é o número de perguntas respondidas,  $R$  é o número de respostas certas, e  $W$  é o número de respostas erradas.

Segundo as descrições anteriormente apresentadas é possível entender que esta medida não é afetada pela ordem de resposta, e sim pela dificuldade total e o numero de respostas certas e erradas. Esta medida é controlada pelo algoritmo de teste apresentando alternância entre as perguntas respondidas pelo usuário certo/errado e pode não ser aplicado a outros algoritmos. Na prática isso significa que a capacidade de medir não deve excepcionalmente ser afetada por um conjunto pequeno de respostas certas ou erradas. Conforme discutido por Linacre (2000), a medida da capacidade do teste-tomador pode convergir com a dificuldade da pergunta no qual o teste-tomador tem uma probabilidade de 50% de responder uma pergunta corretamente. Por exemplo, dado um teste com 10 perguntas e o teste-tomador que respondeu todas as perguntas 5 corretamente e a pergunta 6 de forma errada, medida de capacidade do teste-tomador cairia perto de 5.5.

No caso específico do *MOODLE* há que se considerar que as dificuldades no critério de parada identificadas estão associadas a cinco critérios, quais seja o banco de itens já atingiu sua capacidade total. (ocorre quando há pequena quantidade de itens associadas ao teste); O tamanho máximo é atingido. (ocorre quando atinge o numero Maximo de itens cadastrados nos parâmetros); A medida de capacidade é estimada com precisão suficiente. (ocorre quando cada resposta fornece mais informações estatísticas sobre a medida de capacidade diminuindo o desvio padrão de medida e aumentando sua precisão.); A medida da capacidade é suficientemente longe. (ocorre quando não há mais itens à esquerda); O teste-tomador esta mostrando comportamento fora-teste. (ocorre quando o algoritmo do CAT detecta respostas irrelevantes da mesma opção ou resposta padrão de opções, responder rapidamente e também muito lentamente). Assim, a quantidade de desajustes de pessoas ao modelo, bem como a quantidade de itens parecem ter sido os principais problemas identificados no modelo atual.

Por fim, de um ponto de vista psicométrico, este estudo refere-se a validade, que na opinião de Anastasi e Urbina (2000), referem-se primeiramente ao que o teste mede. Entretanto, essa definição é mais compreensiva, dizendo respeito, essencialmente ao uso específico para o qual a prova foi criada, sendo determinados seus métodos de pesquisa em razão de tais objetivos de testagem. Que vem ao encontro do que diz (Stevens, 1946) que

medir é inserir número para apresentar a variação de uma variável qualquer, a partir de algumas regras bem definidas, seja qual for, e este trabalho teve como objetivo evidenciar validade entre os testes, para os parâmetros de ajustes dos itens papel e lápis indicam que eles foram respondidos conforme as expectativas do modelo Rasch (Rasch, 1960), pois os valores de *infit* e *outfit* estão dentro do intervalo considerado por Linacre (2002) como de ajuste adequado, também para os parâmetros de ajustes dos itens do modelo adaptativo indicam os mesmos e a precisão fornecida por esses modelos de itens favorece a interpretação de que as pessoas, ao responderem os itens, tendem a informar sobre os conhecimentos com maior precisão.

A precisão e eficiência de um teste são preocupações primordiais e, muitas vezes, concorrentes na avaliação aplicada, havendo uma tendência para usar testes (questionários) curtos como alternativa para redução do tempo do teste e dos custos segundo (Makransky e Glas, 2013), com tudo pode ser observado pelos achados que teste informatizado adaptativos compre estes princípios, pois não é necessário que todo o questionário seja aplicado ao respondente para que obtenha uma avaliação desejada. Outro ponto importante também observado por Olea et al. (2004), a respeito de um Teste Adaptativo Computadorizado que o *plugin adaptive quiz* do *MOODLE* nos trás características imprescindíveis a randomização das questões do banco de questões, a aleatoriedade entre os itens das questões, proporcionando ao respondente exclusividade em seu questionário.

## 2.5 Considerações Finais

A presente pesquisa, tendo por base as descrições observadas na literatura no que tange a hipótese que poderia haver correlação entre os testes, sendo que alguma diferença poderia ser mais comum em um ou em outro teste, foi observado que a correlação entre os itens de dificuldades dos testes foi significativa e as habilidades das pessoas teve uma correlação baixa, levantou-se o funcionamento diferencial dos itens (DIF) que compunham o teste Informatizado Adaptativo, utilizando o modelo de Rasch (1960) que permite investigar a probabilidade de acerto aos itens, por pessoas com determinada habilidade, tendo em vista suas dificuldades, observa-se que neste teste em comparação com o papel e lápis nem todos os itens são passados para os respondentes podendo haver alguma discrepância entre os resultados. Pode-se considerar que estudos que se propõem a analisar os itens, recorrendo a TRI como meio de análise, permitem aferir evidência de validade de construto ao instrumento analisado, pois evidenciam o quanto cada item estaria contribuindo para mensurar o construto.

Outro aspecto a ser mencionado é a análise do DIF, que possibilitou concluir que a medida de compreensão investigada na presente pesquisa não apresentou vieses de gênero, não havendo necessidade de modificá-la. Esses dados reforçam a confiabilidade do uso do teste Informatizado Adaptativo acerca de sua eficácia no caso do ENADE Direito, mostrando-se uma alternativa viável e de baixo custo para instituições e governo visando melhorar os sistemas de avaliação do ENADE. Todavia, não se descarta a necessidade de investigações futuras que visem aferir novas evidências de validade para o teste Informatizado Adaptativo no *MOODLE*.

### 3 Referências

AAKER, ET AL **Marketing Research** (7th Ed.), New York: John Wiley & Sons, Inc, 2001. Disponível em: <[http://www.inf.ufsc.br/~verav/Ensino\\_2013\\_2/O\\_uso\\_de\\_questionarios\\_em\\_trabalhos\\_cient%EDficos.pdf](http://www.inf.ufsc.br/~verav/Ensino_2013_2/O_uso_de_questionarios_em_trabalhos_cient%EDficos.pdf)>, Acesso em: 11 dez. 2015.

ADAPTIVE QUIZ ENGINE. Disponível em: <<http://rlcommunity.remote-learner.net/mod/book/view.php?id=70&chapterid=1030>>, Acesso em: 11 dez, 2015.

AERA; APA; NCME (American Educational Research Association; American Psychological Association; National Council on Measurement in Education). **Standards for Educational and Psychological Testing**. Washington, DC: AERA, 1999.

ANASTASI, A., & URBINA, S. **Testagem Psicológica**. Porto Alegre: Artes Médicas. Andrade, D. F., Tavares, H. R., & Valle, R. C. Teoria da Resposta ao Item: conceitos e aplicações. São Paulo: Associação Brasileira de Estatística, 2000.

BARTHOLOMEU, D., SILVA, M. C. R., PIRES, S. D., & PRIMI, R.. **Estabelecimento de pontos de corte pela TRI para uma prova equalizada com o ENADE**. Em F. C.-Capovilla. (Org.). Transtornos de aprendizagem: Progressos em avaliação e intervenção preventiva e remediativa (p. 229-238), v. 1, São Paulo: Memnon, 2011.

BOND, T. G., & FOX, C. M. **Applying the Rasch model: Fundamental measurement in the human sciences**. London: Lawrence Erlbaum Associates, 2001

BRASIL. **Sistema nacional de avaliação do ensino superior - sinaes**. 2015. Disponível em: <[www.inep.gov.br](http://www.inep.gov.br)>, Acesso em: 21 abr. 2015

CHAGAS, D. A.; LISBOA, R. P.; FURTADO, E. S. **Framework MAAVA Metodologia de Avaliação de Ambientes Virtuais de Aprendizagem**. Anais do Simpósio Brasileiro de Informática na Educação, p. 856–859, 2011.

COMPUTER-ADAPTIVE TESTING: **A Methodology Whose Time Has Come**. Disponível em : <<http://www.rasch.org/memo69.pdf>>, Acesso em: 11 dez. 2015.

COSTA, D. R. **Métodos estatísticos em testes adaptativos informatizados**. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009. Disponível em: <<http://www.pg.im.ufrj.br/teses/Estatistica/Mestrado/121.pdf>>. Acesso em: 03 Jul, 2016.

DOUGIAMAS, M. **Reading and Writing for Internet Teaching**, 1999. Disponível em: <<http://dougiamas.com/writing/readwrite.html>>, Acesso em: 11 dez. 2015.

ELLIS, R. **Field Guide to Learning Management Systems, ASTD Learning Circuits**, 2009. Disponível em: <[http://www.astd.org/NR/rdonlyres/12ECDB99-3B91-403E-9B15-7E597444645D/23395/LMS\\_fieldguide\\_20091.pdf](http://www.astd.org/NR/rdonlyres/12ECDB99-3B91-403E-9B15-7E597444645D/23395/LMS_fieldguide_20091.pdf)>. Acessado em: 15 abril de 2016

FERRÃO, M. E. **Introdução aos modelos de regressão multinível em educação**. Campinas: Komedi, 2003.

FRANCISCATO, F. T. et al. **Avaliação dos Ambientes Virtuais de Aprendizagem MOODLE, TeEduc e Tidia - Ae: um estudo comparativo**. Novas Tecnologias na Educação, v. 6, n. 2, p. 5–10, 2008.

HALKITIS, P. N. **CAT algorithm**. Rasch Measurement Transactions, v. 6:4, p.254-5, 1993.

JESUS, G. R. & LAROS, J. A.. **Eficácia escolar: regressão multinível com dados de avaliação em larga escala**. Avaliação Psicológica, 3(2), 93-106, 2004.

JESUS, F. DE; JUNIOR, M.; TEZZA, R. **Algoritmo de um teste adaptativo informatizado com base na teoria da resposta ao item para a estimação da usabilidade de sites de e-commerce**. Produção, v. 23, n. 2010, p. 525–536, 2013.

LANDEIRA-FERNANDEZ, J., & PRIMI, R. (2002). **Comparação do desempenho entre calouros e formandos no Provão de Psicologia** 2000. Psicologia: Reflexão e Crítica, 15(1), 219-234.

LEGOINHA, P., Brilha, J. & Neves, L. **Geologia e Internet em Portugal. Ciências da Terra (UNL)**, Lisboa, n° especial IV, pp. 9-15, 2000. Disponível em: <<http://www.geopor.pt/geotic/papers/legoinha.html>>, Acesso em: 11, dez, 2015

LEGOINHA, P. **O MOODLE e as comunidades virtuais de aprendizagem The MOODLE and the virtual learning communities.** *SciencesNew York*, v. 1, p. 1–4, 2006.

LEITE, E. A. M. et al. **Avaliação Assistida , Feedbacks e Questionários do MOODLE.** *Anais do XXII SBIE - XVII WIE*, n. Novembro 2011, p. 2303–2313, 2011.

LINACRE, J. M. **Computer-Adaptive Testing : A Methodology Whose Time Has Come. Development of Computerized Middle School Achievement Test**, n. 69, 2000.

LINACRE, J. M. **What do infit and outfit, mean-square and standardized mean?** *Rasch Measurement Transactions*, 16(2), 887, 2002.

MAKRANSKY, G. & GLAS, C.A.W. **The Applicability of Multidimensional Computerized Adaptive Testing for Cognitive Ability Measurement in Organizational Assessment.** *International Journal of Testing*, 13(2), 123-139, 2013.

MOODLE, 2015. Disponível em: <<http://www.MOODLE.org>>, Acesso em: 15 fev 2015.

MOREIRA, F. **Sistemática para a implantação de testes adaptativos informatizados baseados na teoria da resposta ao item.** Tese (Doutorado) — Universidade Federal de Santa Catarina, Florianópolis, 2011. Disponível em: <<http://repositorio.ufsc.br/handle/123456789/95506>>. Acesso em: 03 Jul, 2016.

OLEA, J. et al. **Un test adaptativo informatizado para evaluar el conocimiento de inglés escrito: diseño y comprobaciones psicométricas.** *Psicothema*, v. 16, n. 3, p. 519–525, 2004. Disponível em: <http://www.unioviado.es/reunido/index.php/PST/article/view/8230>. Acesso em: 03 jul, 2016.

PORTUGUAL, FCT (Faculdade de Ciências e Tecnologia), 2015. Disponível em: <<http://MOODLE.fct.unl.pt/>>. Acesso em: 15 Out, 2015

PULINO, A. R., **MOODLE, um sistema de gerenciamento de cursos.** Brasília/DF: Departamento de Engenharia Civil e Ambiental, Universidade de Brasília. Sob Licença da Creative Commons (2005).

RASCH, G. **Probabilistic models for some intelligence and attainment tests.** Copenhagen: Danish Institute for Educational Research, 1960

RAUDENBUSH, S. W. **Schooling, statistics, and poverty: can we measure school Improvement?** New Jersey: Educational Testing Service, 2004

RAUDENBUSH, S. W. **What are value-added models of estimating and what does this imply for statistical practice.** *Journal of Educational and Behavioral Statistics*, 29(1), 121-129, 2004b.

RUBIN, D. B. **Estimating causal effects of treatments in randomized and nonrandomized studies.** *Journal of Educational Psychology*, 66(5), 688-701, 1974

RUBIN, D. B., STUART, E. A. & ZANUTTO, E. L. **A potential outcomes view of value-added assessment in education.** *Journal of Educational and Behavioral Statistics*, 29, 103-116, 2004.

SANTANA, L. F. **Avaliação informatiza adaptativa do ENADE pelo MOODLE: evidências de validade.** Centro Universitário FIEO, 2016.

SALVADOR, JOSÉ; GONÇALVES, J. O **MOODLE Como Ferramenta De Apoio a Uma Disciplina Presencial De Ciências Exatas.** *Anais do XXXIV Congresso Brasileiro de Ensino de Engenharia*, v. 1, n. 1941, p. 7122 – 7131, 2006.

SOARES, J. F., RIBEIRO, L. & CASTRO, C. M. **Valor agregado de instituições de ensino superior em Minas Gerais para os cursos de Direito, Administração e Engenharia Civil.** *Dados*, 44(2) 363-396, 2001.

STEVENS, S.S. **On the Theory of Scales of Measurement.** *Science* 103 (2684): 677– 680, 1946. Disponível em: doi:10.1126/science.103.2684.677. PMID 17750512. Acessado em: 16 Nov, 2016

URBINA, S. **Essentials of Psychological Testing.** New Jersey: John Wiley & Sons, Inc., 2004.

WAINER, H. et al. **Computerized adaptive testing: A primer.** Lawrence Erlbaum Associates, Inc, 1990. Disponível em: <<http://www.springerlink.com/index/M58R14228306306V.pdf>>. Acesso em: 03 Jul, 2016.

Recebido em dezembro de 2016

Aprovado para publicação em agosto de 2017

**Lucio Ferreira Santana**

Programa de Pós-Graduação em Psicologia Educacional – Centro Universitário FIEO – UNIFIEO, Brasil, lucioifs@gmail.com

**Daniel Bartholomeu**

Programa de Pós-Graduação em Psicologia - Centro Universitário Salesiano – UNISAL – Americana, Brasil, d\_bartholomeu@yahoo.com.br

**José Maria Montiel**

Programa de Pós-Graduação em Psicologia Educacional – Centro Universitário FIEO – UNIFIEO, Brasil, montieljm@hotmail.com

**Gleiber Couto**

Programa de Pós-Graduação em Psicologia - Universidade Federal de Goiás no Campus de Catalão - UFG/CAC, Brasil, gleibercouto@gmail.com

**Arthur Almeida Berberian**

Programa de Pós-Graduação em Psicologia Educacional – Centro Universitário FIEO – UNIFIEO, Brasil, arthru.berberian@gmail.com

**Fernando Pessoto**

Programa de Pós-Graduação em Psicologia - Centro Universitário Salesiano – UNISAL – Americana, Brasil, fpessoto@gmail.com