

Validação de um modelo de avaliação no uso da modelagem computacional em atividades educacionais

Alexandre Neves Louzada^{1,3}, Marcos da Fonseca Elia^{1,2}, Fabio Ferrentini Sampaio^{1,2}, Andre Luiz Pestana Vidal³, Ricardo Rodrigues³

¹Programa de Pós-Graduação em Informática/UFRJ

²Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais –iNCE/UFRJ
Caixa Postal 2324 Rio de Janeiro – RJ – Brasil

³Fundação de Apoio à Escola Técnica do Estado do Rio de Janeiro – RJ - Brasil

professorlouzada@globo.com, {melia, ffs}@nce.ufrj.br,
avidal@firjan.org.br, cursoidepro@yahoo.com.br

Abstract. *The main aim of this paper is adapting and testing in a Brazilian public school the ACE model of evaluating student performance, proposed by Borkulo, to be used in a teaching-learning process involving computer modeling systems. This model deals with different types of reasoning in two dimensions. Besides to adapt the model and to introduce innovative methodological procedures and instruments for collecting and analyzing data, our main results showed that the ACE model is superior for discriminating the students on the top and bottom of the ability scale of scientific reasoning, when compared to written tests.*

Resumo. *Este trabalho tem por objetivo adaptar e testar em uma escola pública brasileira o modelo de avaliação “ACE”, desenvolvido por Borkulo especificamente para ser aplicado em um processo de ensino-aprendizagem envolvendo a modelagem computacional. Este modelo lida com diferentes tipos de raciocínio envolvidos em duas dimensões. Além de adaptar o modelo e de introduzir novos procedimentos metodológicos e instrumentos de observação e de análise, nossos resultados mostraram que o modelo ACE é mais efetivo na discriminação dos alunos que estão na parte superior e inferior da escala de habilidades de raciocínio científico, quando comparados aos testes escritos.*

1. Introdução

Diferentes pesquisadores no Brasil e no mundo têm ressaltado a importância do uso de programas de computador denominados *sistemas ou ambientes de modelagem computacional* na área educacional [Oliveira, 2010; Borkulo, 2009; De Jong & Van Joolingen, 2007]. Tais ambientes possibilitam a construção de um modelo e a observação de seu comportamento pela simulação de seu funcionamento.

Estas características presentes nos sistemas de modelagem computacional são úteis para auxiliar os alunos na obtenção de habilidades como apresentar hipóteses, aceitar ou refutar argumentos, compreender processos naturais, fazer avaliações qualitativas e quantitativas, etc. A importância destas habilidades pode ser exemplificada no ensino de ciências, onde os alunos precisam utilizá-las para entender fenômenos do mundo real. O Grupo de Informática Aplicada à Educação do NCE/UFRJ

(GINAPE)¹ tem colaborado em fomentar o uso da modelagem computacional através de diferentes pesquisas, utilizando os ambientes de modelagem JlinkIt² e WLinkIt³.

No entanto, para que o potencial de tais ambientes possa ser aproveitado é necessário levar em consideração fatores como a forma adequada de utilizá-los e a avaliação dos resultados, pois o uso de um ambiente de modelagem deve estar associado a estratégias de avaliação, que não só estimule como também avalie consistentemente a aprendizagem.

A partir das considerações feitas, este trabalho tem como objetivo, primeiramente, aplicar um modelo de avaliação desenvolvido Borkulo (2009) especificamente para o uso didático de modelagem dinâmica e, em seguida, procurar responder a seguinte questão: como validar um processo de avaliação voltado para um ambiente de aprendizagem institucionalizado baseado em modelagem computacional?

2. Descrição do Modelo ACE

Com o objetivo de verificar os resultados específicos de aprendizagem que podem ser obtidos por meio da modelagem computacional, Borkulo (2009) fez um levantamento bibliográfico sobre os processos de raciocínio envolvidos na modelagem computacional. A partir deste trabalho, propôs, desenvolveu e testou como parte de sua pesquisa de doutorado, o modelo denominado "ACE" (Figura 1).

Complexidade	Tipos de raciocínio		
	Aplicar	Criar	Avaliar
	Simulação mental	Construir/estender um modelo	Raciocínio científico
Simple Baixo número de variáveis e relações	Predizer ou explicar as conseqüências de uma relação direta	Criar a partir de um modelo	Testar uma relação direta
Complexo Alto número de variáveis e relações	Predizer ou explicar o comportamento de uma estrutura, relação indireta ou ciclo	Criar um modelo como uma solução global	Testar uma relação indireta ou modelo
		Domínio específico	Domínio geral

Figura 1. O modelo ACE, adaptado de Borkulo (2009)

O modelo "ACE" descreve os processos de raciocínio envolvidos na modelagem em três dimensões: tipo de raciocínio, complexidade e especificidade de domínio. A dimensão "tipos de raciocínio" inclui a aplicação (A - Apply), criação (C - Create), e avaliação (E - Evaluate) de um modelo de forma a modificá-lo e gerar novas simulações. A dimensão da complexidade faz uma distinção entre o raciocínio com modelos de menor ou maior grau de complexidade. A dimensão do domínio descreve em que medida o raciocínio é dependente do contexto (específico e geral).

¹ Endereço eletrônico: <http://www.nce.ufrj.br/GINAPE>

² Endereço eletrônico: <http://www.nce.ufrj.br/ginape/JlinkIt/index.htm>

³ <http://www.nce.ufrj.br/ginape/wlinkit/index.htm>

A questão de como medir de forma válida os resultados da aprendizagem obtidos a partir do uso didático de modelagem dinâmica, foi verificada pela autora através da aplicação de um teste sobre modelagem com estudantes de diferentes níveis de conhecimento: ensino médio, graduandos do primeiro ano de Psicologia e de Engenharia Física (sendo que estes últimos haviam completado um curso sobre modelagem); concomitantemente com a realização de uma atividade sobre um domínio específico (aquecimento global) utilizando a modelagem computacional com o apoio do ambiente Co-Lab⁴.

O teste verificou os diferentes tipos de raciocínio em situações simples e complexas dentro do domínio em questão, usando questões objetivas e discursivas, todas corrigidas segundo um critério de correção dicotômico (certo ou errado).

Uma análise das respostas produziu evidências de que as habilidades de raciocínio aplicação (A), criação (C) e avaliação (E) previstas pelo modelo são válidas. Produziu também evidências que sugerem a existência de uma quarta sub-dimensão: reprodução (R - Reproduce), que diz respeito à habilidade do aluno em transferir o aprendido para novos contextos. Outros resultados importantes que merecem destaque foram: alunos com experiência prévia em modelagem e também com conhecimento do domínio enfrentam menor dificuldade para trabalhar com modelos complexos. Há ainda menção a um estudo avaliativo posterior sobre o impacto do uso da modelagem dinâmica na aprendizagem tradicional em relação à “aprendizagem investigativa”

3. Proposta de Pesquisa

A presente pesquisa procura reproduzir o trabalho de Borkulo descrito na seção 2, em termos dos seus objetivos e do modelo utilizado (ACE). Entretanto, algumas adaptações foram necessárias para adequá-lo à realidade das escolas brasileiras:

1. O nosso desenho experimental envolve uma aplicação institucional para todos os 151 alunos cursando a matéria de Física Térmica da 2ª Série de uma escola pública de nível médio, o que nos permite fazer comparações entre o modelo de avaliação ACE e os instrumentos tradicionais de avaliação. Já a situação experimental proposta originalmente por Borkulo (2009) é bem diferente, pois utiliza como campo de pesquisa experimental amostras de alunos cursando diferentes níveis escolares, o que a obrigou utilizar a teoria de resposta ao item (TRI), dentre outros procedimentos, para poder estabelecer comparações.
2. Uma proposta com características institucionais não permite que sejam utilizadas amostras randômicas nem, tampouco, utilizar grupos de controle. Assim, o nosso estudo foi organizado na forma de um quase-experimento usando o modelo conhecido na literatura como de “um único caso ABAB” [Cohen et al, 2000] que consiste em, alternadamente, aplicar (A) e deixar de aplicar (B) uma nova situação experimental a um mesmo grupo de alunos, criando uma sequência experimental ABAB... No presente caso, a situação experimental A seria o uso durante aproximadamente 4 semanas da modelagem computacional como recurso didático e a situação B seria a utilização, com o mesmo período de 4 semanas, dos recursos que tradicionalmente vêm sendo utilizados pela escola em estudo. Além disso, começou-se o curso com a situação didática tradicional B.

⁴ Endereço eletrônico: <http://www.co-lab.nl>

3. O sistema de avaliação da escola em tela prevê seis avaliações sob a forma de provas (P_1 , P_2 , P_3 , P_4 , P_5 e P_6) que foram aplicadas sempre após a conclusão de um ciclo didático, formando portanto a sequência “ $BP_1AP_2BP_3AP_4BP_5AP_6$ ” quando se considera o ensino-aprendizado e a avaliação.

Estas provas foram preparadas em conjunto pelos pesquisadores externos e pelos professores-pesquisadores da escola responsáveis por ministrar o curso aos alunos, sendo que todas foram planejadas segundo uma mesma matriz de referência (Quadro 1) para permitir comparações.

Quadro 1 - Matriz de referência das provas

MATRIZ DE REFERÊNCIA	PROVA 1	PROVA 2	PROVA 3	PROVA 4	PROVA 5	PROVA 6
A – CATEGORIA						
MÚLTIPLA ESCOLHA	7	1,2,4,5,6,8	1,2	4,5	4,8,9,10	4,5
SEMIABERTA	1,2,3,4,5,6,8,9,10	3,7,9,10	3,4,5,6,7,8,9,10	1,2,3,6,7,8,9,10	1,2,3,5,6,7	1,2,3,6,7,8,9,10
B – FOCO						
MODELAGEM	Não aplicada	2,4	Não aplicada	1,2	Não aplicada	1,2
FÓRMULA ANALÍTICA	5	5	5,6,7	3	2,6,7	6,7,8,9,10
FÓRMULA COM GRÁFICO	7,8,9,10	6	8,9,10	8,9,10	3,9	5
OUTRAS	1,2,3,4,6	1,3,7,8,9,10	1,2,3,4	4,5,6,7	1,4,5,8,10	3,4
C – CONTEXTO						
ESCOLARIZADO	1,2,3,5,6,7,8,9,10	1,2,3,4,5,6,8	2,3,5,6,7,8,9,10	3,4,5,6,7,8,9,10	2,3,4,6,7,8,9,10	2,4,5,6,7,8,9,10
COTIDIANO	4	7,9,10	1,4	1,2	1,5	1,3
D – AUTORIA						
PROFESSOR	1,2,3,4,5	1,2,3,7,9	1,2,3,4	1,2,3,4,5,6,7	1,2,3,4,5	1,2,3,4,5
LITERATURA	6,7,8,9,10	4,5,6,8,10	5,6,7,8,9,10	8,9,10	6,7,8,9,10	6,7,8,9,10
E – BLOOM						
CONHECIMENTO	4	2,3,4	2,3	5,6	4,8	4
COMPREENSÃO e/ou APLICAÇÃO	1,2,3,5,6,7,8,9,10	1,5,6,7,8,9,10	1,4,5,6,7,8,9,10	1,2,3,4,7,8,9,10	1,2,3,5,6,7,9,10	1,2,3,5,6,7,8,9,10

Para análise dessas provas utilizou-se a teoria clássica de itens aplicada ao conjunto de questões classificadas como sendo de mesmo tipo na matriz de referência (Quadro 1). Nesta teoria são definidos três índices básicos: (i) nível de desempenho dado aqui pela média dos pontos obtidos no conjunto de questões; (ii) o índice de consistência interna definido pelo alfa de Cronbach [Cronbach,1951] entre as questões que formam um dado conjunto; e (iii) o índice de discriminação de um dado conjunto de questões com o todo formado pela prova, definido aqui pela correlação Pearson entre os pontos obtidos pelos alunos em cada conjunto de questões e os pontos obtidos na prova.

As avaliações segundo o modelo ACE para as fases “Application”, “Creation” e “Evaluation” foram planejadas em nossa pesquisa nos mesmos moldes propostos por Borkulo (2009) e foram aplicadas durante os três períodos (tipo “A”) em que foi utilizada a modelagem computacional como recurso didático, versando respectivamente sobre os seguintes temas: dilatômetro, efeito estufa e motor de 4 tempos. O quadro 2 ilustra as três atividades ACE propostas com relação à habilidade “criar-complexo”.

Além das provas e das atividades ACE, foi utilizado também como instrumento de avaliação um inventário [Yeo e Zadnik, 2001] sobre conceitos científicos e sobre concepções alternativas a respeito do conteúdo programático abordado no curso: estudos sobre temperatura e calor. Este inventário foi aplicado no início e no final do período letivo a todos os 151 alunos e, para fins da presente análise, foram considerados apenas os ganhos em conceitualização científica, obtidos por esses alunos. Os detalhes desse estudo transcendem o escopo do presente trabalho e podem ser encontradas em [Louzada, 2011].

Quadro 2 - Atividades ACE propostas com relação à habilidade “criar-complexo”

Atividade 1 – Dilatômetro	
7.Com base no texto de nosso livro de física construa um modelo sobre a lâmina bimetálica (<u>Ver o texto do livro de física na página 26</u>).	
Atividade 2 - Efeito estufa	
6.Leia as notícias abaixo e use estas informações para alterar o modelo:	
<p>Matéria do jornal O Globo publicada em 24/08/2010</p> <p>TSE libera publicidade institucional da Eletrobrás contra queimadas</p> <p>BRASÍLIA - O presidente do Tribunal Superior Eleitoral (TSE), Ricardo Lewandowski, autorizou a Eletrobras a veicular uma campanha contra as queimadas. Segundo a empresa estatal, as queimadas, frequentes nesta época do ano, desligam as linhas de transmissão elétrica e danificam equipamentos, causando prejuízos às empresas do setor. (...) Além de cartilhas distribuídas em escolas, sindicatos e associações de produtores rurais, a campanha contará com um filme para televisão e uma peça para o rádio.</p>	<p>Edição do Jornal Nacional de 5/9/2007</p> <p>Temporada de queimadas, clima seco e problemas respiratórios</p> <p>Mais de 17 mil pontos de queimadas foram registrados pelo Inpe nas últimas 24 horas em todas as regiões do Brasil. Conseqüência de um período de seca fora do comum. (...) O clima, considerado de deserto, multiplica os problemas respiratórios. Nos hospitais, o movimento dobrou.</p>
Atividade 3 - Motor quatro tempos	
6.Com base nas nossas aulas sobre termodinâmica, construa um modelo de uma máquina a vapor.	

Os resultados obtidos com a aplicação das provas tradicionais e o modelo ACE serão discutidos a seguir.

4. Resultados

Provas Tradicionais

Inspecionando a matriz de referência no quadro 1 pode-se dizer que as provas apresentaram o seguinte perfil geral: questões semi-abertas, sem foco definido, sendo preparadas tanto pelo professor da turma ou obtidas de livros didáticos, com conteúdo da matéria apresentado mais frequentemente em um contexto escolarizado e exigindo habilidades de compreensão e de aplicação.

São apresentados na figura 2 os resultados na forma tabular e gráfica do desempenho médio %, por conjuntos de questões obtidos a partir da matriz de referência do quadro 1 e separadamente para cada uma das 6 provas aplicadas na sequência “BP1AP2BP3AP4BP5AP6”.

Como se pode notar, em geral o desempenho dos alunos foi bom em todas as provas, mas estes valores oscilam com o pico positivo coincidindo, na grande maioria

das vezes, com as provas (índices pares) aplicadas logo após o período de ensino-aprendizagem em que se usou a modelagem computacional.



Figura 2- Desempenho médio % dos alunos por tipo de questão em diferentes provas

É importante notar também, através do gráfico 1, que alguns tipos de questão estão mais sujeitas a estas oscilações, inclusive, com tendência a aumentar o desempenho médio de prova para prova, como por exemplo acontece com as questões envolvendo modelagem e as questões envolvendo representação gráfica, enquanto que outras quase não oscilam, como é o caso das questões do tipo semi-aberta e aquelas

apresentadas com fórmulas ou em um contexto escolarizado. Há ainda aquelas que sequer oscilam, ou seja, não detectam a mudança do método, como seria o caso das questões envolvendo conhecimento ou aquelas preparadas pelo professor.

Esses resultados sugerem que, do ponto de vista do nível de desempenho (média) estas 6 provas tradicionais aplicadas, quando tomadas separadamente, foram capazes de detectar a mudança do método didático entre a modelagem computacional e o tradicional e que, portanto, poderiam ser instrumentos de avaliação adequados para uma situação de aprendizagem centrada na modelagem computacional, desde que cuidadosamente planejados, construídos e analisados pelos professores, como ocorreu no presente contexto de uma pesquisa. Todavia, observa-se também que não é qualquer tipo de questão que seria apropriado.

Quando tomadas em conjunto, as 6 provas não refletem uma escala unidimensional, pois a consistência interna entre elas é muito baixa, apresentando um valor para o alfa de Cronbach estandardizado de apenas $\alpha=0,53$. Além disso, todas as provas apresentam um índice de discriminação muito baixo ($\approx 0,30$) com o conjunto das 6 provas tomadas como um todo, sendo que para a prova 4 este índice é praticamente nulo ($\approx 0,09$), portanto, destoando muito das demais. Isto, no mínimo, reflete quão injusta é a prática usual de representar o desempenho dos alunos por uma única nota final tomada pela média das provas de período.

Atividades ACE

A tabela 1 mostra separadamente, para as dimensões (a)-simples e (b)-complexa, os valores médios percentuais da avaliação ACE obtidos nas 3 atividades aplicadas para cada habilidade cognitiva testada (Application, Creation e Evaluation). São mostradas também a média geral (%) e as médias marginais (%) por atividade e por habilidade, como também os desvios destas médias marginais em relação à média geral, os quais representam, respectivamente, os efeitos principais das 3 atividades A_i ($i=1,2,3$) e das 3 habilidades (ACE)_j ($j=A,C,E$).

Analisando os resultados verifica-se que houve um efeito importante das atividades sobre a média geral seguindo o mesmo padrão, tanto na dimensão simples quanto na complexa: A1 abaixo, A2 próximo e A3 acima da média; enquanto que o efeito das habilidades ACE só parece ser mais importante na dimensão complexa assumindo um valor bem abaixo da média geral. Além disso, para a nossa amostra de alunos, a habilidade criar (“Create”) parece que foi bem mais difícil que a habilidade avaliar (“Evaluate”), uma contradição com o trabalho de Borkulo (2009).

Com base na descrição das três atividades “criar-complexo” apresentadas anteriormente (Quadro 2), podemos especular que o baixo percentual de aproveitamento obtido na atividade 1 relacionada ao tema “dilatômetro” pode ser explicada, em parte, pelo fato de ser baseada em uma situação mais escolarizada quando comparada com as outras duas nas quais os alunos costumam ter mais familiaridade. Por outro lado pode se considerar também o fato de que a atividade 1 foi provavelmente a primeira oportunidade de aprendizagem em que os alunos foram solicitados a transferirem conhecimento aprendido em sala de aula para um novo contexto.

Como era de se esperar, posto que as atividades ACE envolvem diferentes raciocínios (Application, Creation e Evaluation), graus de complexidade (Simple ou Complexo) e diferentes domínios (Dilatação, Aquecimento global e Motor à explosão) é

muito baixa a consistência entre cada atividade ACE (A1, A2, A3), variando de 0 a 0,39 quando tomadas individualmente e alcançando o máximo de 0,46 quando consideradas todas as 18 observações, sugerindo que, a exemplo das provas tradicionais, cada uma delas também representa uma unidade de observação diferenciada das demais.

Tabela 1. Valores médios percentuais das avaliações ACE

Dimensão A – simples					
Questões – ACE	Aplicar - Simples	Criar - Simples	Avaliar – Simples	M por Atividade	Atividade
Atividade 1	82,22	62,96	77,78	74,32	-7,90
Atividade 2	85,29	79,41	83,09	82,60	0,38
Atividade 3	89,23	85,38	94,62	89,74	7,52
M por Habilidade	85,58	75,92	85,16	M Geral 82,22	
ACE	3,4	-6,3	2,9		
Dimensão B – complexo					
Questões – ACE	Aplicar -Complexo	Criar -Complexo	Avaliar -Complexo	M por Atividade	Atividade
Atividade 1	70,37	28,15	74,44	57,65	-8,08
Atividade 2	72,06	46,32	77,94	65,44	-0,29
Atividade 3	76,92	60	85,38	74,10	8,37
M por Habilidade	73,12	44,82	79,25	Média Geral 65,73	
ACE	7,4	-20,9	13,5		

Análise comparativa: Provas x Habilidades ACE x Ganho nos Conceitos Científicos

As análises anteriores envolvendo respectivamente o desempenho dos alunos nas provas tradicionais, nas atividades de avaliação ACE e no nível de respostas científicas mostraram que os três instrumentos apresentam boas características técnico-operacionais desejáveis para um instrumento de avaliação, ou seja, todos passam no teste de confiabilidade para avaliar o desempenho dos alunos na matéria de física térmica.

Contudo, como se sabe a confiabilidade é uma condição necessária mas não suficiente para aceitarmos um dado instrumento de avaliação, é preciso também que ele seja válido, ou seja, que ele avalie o que desejamos que seja avaliado.

Particularmente, a hipótese de pesquisa colocada neste trabalho é verificar se o modelo de avaliação ACE proposto por Borkulo (2009) é mais adequado para avaliar o desempenho dos alunos em um curso em que se usa a modelagem computacional, do que as provas tradicionais.

No presente trabalho, esta hipótese foi testada de forma indireta, primeiramente, assumindo como critério de verdade o aprendizado de conceitos científicos. Em segundo lugar, fazendo uma comparação do ganho no aprendizado de conceitos científicos ocorrido ao longo do curso pelos alunos com as medidas de desempenho obtidas, respectivamente, por meio das provas tradicionais e do modelo ACE.

Com este objetivo, foi definido o ganho relativo em conceitualização científica ao longo do curso pela diferença entre o resultado do pós-teste e do pré-teste do inventário

aplicado aos alunos (descrito no final da seção 3) em relação ao resultado pré-teste, expressando este valor em percentagem:

$$GCCIEEN_r = \frac{(Pós_T - Pré_T)}{Pré_T} \times 100 (\%)$$

Em seguida, os valores dos ganhos GCCIEEN_r foram divididos em dois grupos: inferior (=1) e superior (=2); assumindo três diferentes pontos de corte no ganho relativo (40%, 60% e 80%) para verificar se uma eventual associação dependeria do nível do ganho.

Por fim, foi aplicada a técnica multivariada de análise de discriminância para discriminar os dois grupos, inferior e superior para cada ponto de corte, a partir de uma única função discriminante formada por uma combinação linear, de um lado (i) pelas 6 provas tradicionais e, de outro (ii) pelas 18 avaliações ACE. A Tabela 2 mostra os resultados das classificações obtidas em termos do % de alunos em dado grupo (1 ou 2) que foram corretamente classificados naquele grupo original (G1»1 ou G2»2) bem como, as três primeiras variáveis (instrumentos) mais discriminantes e seus respectivos coeficientes de correlação com a função discriminante.

É possível perceber através do percentual de acertos de classificação da tabela 2 que o conjunto formado pelas avaliações ACE é mais efetivo que o conjunto das provas tradicionais para discriminar o ganho em conceituação científica entre os alunos do grupo superior e inferior quando o nível de exigência é muito alto (ponto de corte com ganho superior a 80%), seja em termos do total de classificação corretas ou seja para classificar corretamente os melhores G2»2. E também é o mais efetivo quando o nível de exigência é muito baixo (ponto de corte com ganho superior apenas a 40%) para classificar corretamente os piores G1»1. Vê-se ainda que no meio da escala (ponto de corte com ganho superior a 60%) os dois sistemas de avaliação são igualmente efetivos para avaliar os alunos.

Tabela 2- Resultados da análise de discriminância: provas tradicionais vs. Avaliações ACE para diferentes pontos de corte (40%, 60% e 80%)

Variáveis = Provas Tradicionais						
Grupos GCCien 1 e 2	% de Acertos Classif			Principais Provas (coeficiente de correlação)		
	G1»1	G2»2	Total			
Variáveis = Provas Tradicionais						
(1<40%; 2>40%)	50	63,8	60,2	Prova 6 (0,623)	Prova 3 (0,555)	Prova 4 (0,533)
(1<60%; 2>60%)	67,3	51,3	57,8	Prova 3 (0,860)	Prova 6 (0,566)	Prova 5 (0,308)
(1<80%; 2>80%)	60,9	55,9	58,6	Prova 3 (0,779)	Prova 6 (0,453)	Prova 2 (0,319)
Variáveis = Avaliações ACE						
(1<40%; 2>40%)	81,5	63,9	68,2	ACE 3 EC (-0,406)	ACE2 EC (0,394)	ACE3 CC (0,358)
(1<60%; 2>60%)	63,4	63,8	63,6	ACE 3 AS (-0,485)	ACE 2 EC (0,474)	ACE 3 CC (0,325)
(1<80%; 2>80%)	64,3	72,2	68,2	ACE 2 EC (0,579)	ACE 2 CC (0,511)	ACE 3 AS (-0,307)

E inspecionando os coeficientes de correlação dos instrumentos ACE com a função discriminante, nota-se que a atividade 1 não ajudou em nada na discriminação dos alunos e a atividade 3 foi a mais discriminativa aparecendo 5 vezes dentre as 9 melhores. Analogamente, o tipo de raciocínio “evaluation” (4 em 9) e a dimensão

“complexa” (7 em 9) são de longe mais discriminativas que a dimensão “simples”. Com relação aos instrumentos tradicionais, as provas mais discriminativas foram as prova 3 e prova 6.

5. Considerações Finais

Este trabalho comparou o modelo de avaliação ACE proposto por Borkulo (2009) para avaliar estratégia didáticas baseadas no uso da modelagem computacional e os modelos tradicionais de avaliação baseados em provas, tendo como critério de referência o ganho dos alunos em conceituação científica. Os resultados desta comparação mostram que o primeiro é mais efetivo para identificar alunos nos extremos superior e inferior dessa escala de habilidade, enquanto que ambas se equivalem no meio da mesma.

Esses resultados trouxeram também fortes indícios (resumidos na figura 2) que medidas educacionais obtidas por meio de provas são adequadas, desde que as questões sejam construídas atendendo os requisitos técnicos e operacionais necessários.

Este trabalho fez parte de uma pesquisa de dissertação de mestrado [Louzada, 2011], mas sua proposta continua em andamento como uma inovação institucional na escola pública em que foi aplicada na modalidade de uma pesquisa-ação, na disciplina de física.

6. Agradecimentos

Agradecemos à Direção da Escola Técnica Estadual Oscar Tenório pelo apoio oferecido para a realização desta pesquisa.

Referências

- Borkulo, S. P. van (2009). The assessment of learning outcomes of computer modelling in secondary science education. Tese de D.Sc, University of Twente, Enschede, 2009.
- Cohen, L., Manion, L. & Morrison, K. Research Methods in Education. Taylor & Francis Group. London and New York, 2000.
- Cronbach, L. J., 1951. Coefficient alpha and the internal structure of test. *Psychometrika*, 16:297-334.
- De Jong, T., & Van Joolingen, W. R. (2007). Model-facilitated learning. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer & M. P. Driscoll (Eds.), Handbook of research on educational communication and technology (3rd ed., pp. 457-468): Lawrence Erlbaum.
- Louzada A.N. Uso da Modelagem Computacional Como Proposta de Inovação Curricular Para o Ensino de Física no Ensino Médio. Rio de Janeiro. Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Núcleo de Computação Eletrônica, 2011.
- Oliveira, C. B. P. de ; Sampaio, F.F. ; Elia, M. F. . Utilização da Ferramenta de Modelagem Dinâmica JLinkIt no Processo de Ensino-Aprendizado de Física no Ensino Médio. *Revista Brasileira de Informação na Educação*, v. 18, p. 55-69, 2010.
- Yeo, S.;Zadnik, M.Introductory Thermal Concept Evaluation: Assessing student’s understanding, *Physics Teacher*, V. 39, pp.496-504, 2001.