

Metodologia da Gestão da Avaliação do Processo Ensino-Aprendizagem:

Psicometria e Estatística, Aproximações Possíveis e Necessárias

Regina Albanese Pose - Faculdade de Medicina USP (albanese@usp.br)

Rosângela Molini - Faculdade de Medicina USP (rosangela.molini@gmail.com)

Joaquim Edson Vieira - Faculdade de Medicina USP (joaquimev@usp.br)

Marcelo Ventura Freire - Escola de Artes, Ciências e Humanidades
e Faculdade de Medicina USP (mvf@usp.br)

Resumo

Este estudo descreve, de maneira simplificada e resumida, uma metodologia com técnicas de análise psicométricas e estatísticas aplicadas à gestão da avaliação do processo de Ensino-Aprendizagem que vem sendo desenvolvida em um curso de graduação em Medicina de uma universidade pública estadual de São Paulo. É desejado um instrumento de validação da proficiência médica construído em termos de habilidades, competências, conteúdos e cenários ilustrativos do perfil esperado para um profissional médico; qual seja, uma matriz de referências. Essa instituição aprovou, recentemente, propostas de reestruturação curricular, que considera exposição prática tutorada como metodologia complementar às tradicionais exposições teóricas. Uma reestruturação fundamentada nas novas Diretrizes Curriculares Nacionais, que, entre tantas modificações pretende direcionar os estudantes na busca de conhecimentos nas áreas de formação em medicina por meio de contextualização de condições de vida e saúde reais. A construção de uma matriz de referência pautada na matriz curricular do curso de Medicina deve estar associada às novas DCN. Os autores deste trabalho propõem uma analogia ao trinômio *Real–Possível–Necessário* de Piaget para a comissão de avaliação. Existem estudos discutindo a unidimensionalidade dos itens, porém advogamos o uso de um modelo de TRI unidimensional aproximação razoável para algumas situações em que a probabilidade de acerto é função de mais de uma habilidade, o que exigiria um modelo de TRI multidimensional para a análise das avaliações em Medicina. No caso dos estudantes de Medicina, é plausível assumir que o processo seletivo de ingresso no curso nivela os alunos em relação às proficiências

básicas não médicas, enquanto também é defensável a independência dos domínios do aprendizado da Medicina.

Palavras-chave: Teoria da Resposta ao Item, Teoria Clássica da Medida, Avaliação Educacional, Educação Médica.

Metodologia da Gestão da Avaliação do Processo Ensino-Aprendizagem

O desenvolvimento de um processo de gestão da avaliação do processo Ensino-Aprendizagem em Medicina, deve contar com um instrumento de validação da proficiência médica, instrumento esse construído em termos de habilidades, competências, conteúdos e cenários ilustrativos do perfil esperado para um profissional médico; qual seja, uma matriz de referências (Bloom, 1983). Para o curso de Medicina, o Ministério da Educação reformulou em junho de 2014 as Diretrizes Curriculares Nacionais (DCN) criadas em 2001, que definem competências, conteúdos essenciais e habilidades para o esse curso superior regulamentado no Brasil, definindo assim uma nova matriz curricular para o curso, cujo dever é apontar, de acordo com Marcondes (1998), para o resultado final de um longo processo educacional cada vez mais amplo e complexo para atender às expectativas sociais referentes a essa profissão. A construção de uma matriz de referência pautada na matriz curricular do curso de Medicina deve estar associada às novas DCN, como ilustra a Figura 1.

O pleno desenvolvimento da formação médica requer a construção de um processo educacional conduzido por meio de conceitos e metodologias de gestão (Harden, 1986) capazes de promover o planejamento e acompanhamento das várias ações e atividades de trabalho envolvidas, assim como a avaliação do processo como um todo.

Uma proposta de gestão da Avaliação do Processo Ensino-Aprendizagem necessitar ser pautada por pelo menos três princípios (Song, 2011):

- responsabilidade na obtenção dos dados de avaliação de aproveitamento e ambiente de ensino;
- liderança e colaboração de comissões formadas por docentes, discentes e instituições; e
- desenvolvimento docente (*expertise* na área específica e em técnicas de ensino-aprendizagem),

aos quais os autores urgem acrescentar o desenvolvimento técnico específico na área de Psicometria e Estatística, com profissionais que tenham formação específica para a função e qualificação técnica adequada. É necessário desenvolver e manter um sistema de gestão de avaliação que efetivamente mesure conhecimento, habilidades e atitudes, e ainda apresentem atributos qualitativos referentes a eficiências e deficiências, em tempo de acelerar ou recuperar o processo.

Os cursos de graduação em Medicina necessitam de avaliação como parte de seus procedimentos de gestão para a garantia da qualidade da formação do egresso como um profissional capacitado. Mas o valor da avaliação é muito maior do que a simples prestação de auditoria sobre as informações oferecidas. Enquanto parte integrante do processo de gestão, a avaliação na educação médica deve promover o desenvolvimento contínuo de instrumentos de medição confiáveis sobre o desempenho dos alunos.

Avaliação é parte essencial do processo educativo e conceitualmente ligada à aprendizagem, de tal forma que a quantidade e a qualidade devem ser mensuradas por razões de controle do processo, segurança da aplicabilidade dos conteúdos e habilidades e também como meio de realimentação do processo, promovendo alterações. Pode envolver medidas subjetivas e objetivas em abordagens qualitativas e quantitativas. Entrevistas, questionários, resultados de provas e testes permitem a obtenção de informações, a respeito dos respondentes e do processo envolvido, neste caso, de ensino-aprendizagem.

Avaliações das proficiências médicas devem ser elaboradas considerando os fundamentos da área, a validade e confiabilidade das avaliações e a clareza sobre sua função formativa (que permite retorno e utilização como aprendizado) ou somativa (que pode ser utilizada em processos de classificação), ilustradas na Figura 2. Em avaliações somativas (classificatórias), os testes de conhecimento devem ser fundamentados na matriz de referência, devem ser elaboradas, aplicadas e analisadas adequadamente e de forma longitudinal, de modo a validar o estágio do desenvolvimento dos estudantes no processo ensino-aprendizagem.

Avaliação do Processo Ensino-Aprendizagem

A Faculdade de Medicina da Universidade de São Paulo (FMUSP) (o curso de graduação em Medicina da universidade pública estadual de São Paulo supracitada) aprovou, recentemente, propostas de reestruturação curricular. A nova proposta

considera exposição prática tutorada como metodologia complementar às já tradicionais exposições teóricas de excelência centenárias. Uma reestruturação fundamentada nas novas Diretrizes Curriculares Nacionais, que, entre tantas modificações pretende direcionar os estudantes na busca de conhecimentos nas áreas de formação em medicina por meio de contextualização de condições de vida e saúde reais.

Neste processo, a gestão da avaliação do processo de Ensino-Aprendizagem constitui meio essencial nos três aspectos de avaliação (Bloom, 1983 – Figura 2). No aspecto formativo, que posiciona o estudante sobre suas excelências e eventuais deficiências, reforçando as atividades do semestre (para as avaliações ao final do semestre com equalização por itens comuns entre as provas de anos de curso). No aspecto somativo, que confere à instituição o desenvolvimento (longitudinal, inclusive, a partir do segundo ano de avaliação, ou seja, da terceira edição da prova) do processo ensino-aprendizagem da proficiência em medicina (assim, analisa o binômio docente–discente de forma ampla, de um ano calendário para outro, e ainda, comparando turmas distintas, com equalização por itens comuns entre as provas do mesmo ano), garantindo a excelência da FMUSP, priorizando aspectos de ética e humanidades. No aspecto diagnóstico, que confere ao docente informações necessárias para o planejamento do curso, como por exemplo, o professor do 4º ano, pode elaborar o planejamento de acordo com o desempenho dos estudantes que terminaram o 3º ano, assim, para uma turma mais avançada, é possível realizar um curso mais avançado.

A matriz curricular está sendo desenvolvida por comissões específicas, cujos membros são docentes da faculdade, com *expertise* em currículo Médico, com a expectativa de término ao longo de 2015. Desta feita, para primeira edição da prova será utilizado um esquema do tipo *blueprint* (Figura 3).

Além disso, na matriz curricular do curso de graduação em Medicina, os conteúdos são divididos em grandes áreas (Saúde Mental, Pediatria, Clínica Cirúrgica, entre outras) que não são hierarquicamente acumulativas entre si, como ocorre, por exemplo, entre as grandes áreas do curso de graduação em Matemática (Lógica, Análise, Álgebra, entre outras). Além disso, os conteúdos dessas diferentes áreas são suficientemente diferentes entre si ao ponto de, juntamente com a não-acumulatividade, permitirem aos especialistas em Educação Médica avaliar como plausível a independência estatística (no mesmo sentido em que são independentes os fatores latentes da Análise Fatorial Exploratória) entre as proficiências nas grandes áreas.

Deve-se ter muito claro que a matriz de referência para uma avaliação somativa deve ser construída de tal sorte que as competências sejam compostas por domínios do aprendizado, domínios estes descritos por várias competências que se combinam no desenvolvimento de uma competência mais ampla. Tal afirmação está pautada em Bloom (1983), “Talvez o nível de generalização seja o fator que diferencie mais nitidamente a avaliação somativa da formativa”.

Os autores sugerem que, para tanto, esse processo deva ser construído após o final das avaliações formativas e antes da avaliação somativa. Assim, essas competências complexas contudo independentes umas das outras podem ser estimadas por modelos unidimensionais da Teoria da Resposta ao Item (v. Seção *Por que TRI e no TRIM? a seguir*).

Psicometria e Estatística: aproximações possíveis e necessárias

Piaget (1985) descreve o processo de construção do *Real* (relacionado à construção do aprendizado do sujeito em interação com o objeto) expresso pela coordenação do *Possível* (que exprime liberdade de procedimento e que é um fator externo ao sujeito) e do *Necessário* (que exprime a auto-regulagem e fechamento das composições e que é um fator interno ao sujeito).

Os autores deste trabalho propõem uma analogia ao trinômio *Real–Possível–Necessário* de Piaget para a comissão de avaliação (o sujeito), em que a construção de uma “ilustração ou imagem” do desenvolvimento da competência médica é o *Real* da comissão de avaliação; em que a proficiência nos conteúdos da matriz de referência é o *Possível* da comissão de avaliação (estimável através da Teoria de Resposta ao Item), ou seja, um fator externo à própria comissão de avaliação; e em que a resposta ao item da prova é o *Necessário* da comissão de avaliação, o seu fator auto-regulador e interno.

Segundo Andrade (2000), a Teoria de Resposta ao Item (TRI) é um conjunto de modelos matemáticos em que a probabilidade de um sujeito responder corretamente determinado item é uma função dos parâmetros desse item (nível de dificuldade, capacidade de discriminação, entre outros possíveis) e da proficiência desse respondente, de modo que, quanto maior essa proficiência, maior a probabilidade de o respondente acertar o item. A partir dos Anos 30, começam a surgir os primeiros

trabalhos de TRI na solução de problemas relacionados à Educação e à Psicometria (Andrade, 2000).

Os parâmetros dos itens e as proficiências são estimados por meio de métodos estatísticos a partir das respostas dos avaliados e do modelo proposto. Com essas estimativas, os itens podem ser posicionados em uma escala psicométrica, permitindo atribuir-lhe uma interpretação pedagógica e validar a matriz de referência utilizada.

Há duas suposições básicas para a aplicação dos modelos usuais da TRI: unidimensionalidade e independência local. Por unidimensionalidade, entende-se que a prova esteja medindo um traço latente único, que pode representar uma proficiência ou uma composição de habilidades e proficiências dos avaliados. Por independência local, entende-se que a dependência entre os itens é perfeitamente explicada apenas pelo traço latente dos avaliados. Nesse contexto, a aplicação de modelos TRI torna-se bastante útil.

Assim, um item precisa estar relacionado a apenas uma única proficiência (traço latente), mesmo que, no conjunto da prova, várias proficiências diferentes possam ser avaliadas, cada qual com seus itens específicos.

Há vários modelos de TRI, de acordo com o tipo de item, com o tipo de resposta e com a escala de mensuração. Um dos mais simples é o modelo de Rasch (1960), que estima como parâmetro para cada item apenas o seu nível de dificuldade e que exige um tamanho de amostra menor que o dos outros modelos com mais parâmetros, desde que se verifique a premissa de que todos os itens para uma proficiência tenham o mesmo poder de discriminação.

Por que TRI e não TRIM? Unidimensionalidade *versus* multidimensionalidade

Existem estudos discutindo a unidimensionalidade dos itens (Childs 2000; Reckase, 2009), porém o uso de um modelo de TRI unidimensional é defensável como aproximação razoável para algumas situações em que a probabilidade de acerto é função de mais de uma habilidade, cenário que poderia exigir o emprego de um modelo de TRI multidimensional (TRIM), como, por exemplo, o modelo logístico não-compensatório de três parâmetros (Sympson, 1977; Whitely, 1980), que poderia ser empregado com avaliações em Medicina.

Apenas para fins de ilustração, consideremos o caso em que a probabilidade de acerto seja influenciada por duas proficiências θ_1 e θ_2 :

$$P(X = 1|\theta_1, \theta_2) = c + (1 - c) \cdot \frac{1}{1 + \exp\{a_1(\theta_1 - b_1)\}} \cdot \frac{1}{1 + \exp\{a_2(\theta_2 - b_2)\}}$$

como pode ser visto na Figura 4, com $c = 0$, $a_1 = a_2 = 1$ e $b_1 = b_2 = 0$. Cabe destacar que, para θ_1 em $[-5, 5]$ e θ_2 em $[-5, 5]$, a superfície característica do item não pode ser aproximada por uma curva característica de item (função de uma única proficiência), mas, à medida em que deslocamos somente θ_2 para uma região distante de b_2 (como mostra a segunda imagem da Figura 5), o comportamento da superfície característica (multidimensional) pode ser aproximado por uma curva característica (unidimensional), já que a influência de θ_2 passa a ser desprezível. A Figura 6 simplesmente apresenta a Figura 5b de pontos de vista mais frontais.

A interpretação dessa aproximação matemática em termos pedagógicos é a seguinte: à medida que um respondente desenvolva uma proficiência θ_2 muito acima da dificuldade b_2 , a probabilidade de acerto passa a ser cada vez menos sensível a aumentos de θ_2 , a ponto de depender praticamente apenas de θ_1 . Para ilustrar com um exemplo próximo ao apresentado por Barbetta (2014), um enunciado de um problema matemático difícil (como um típico problema do ENEM) porém com um enunciado bem redigido, em linguagem clara e acessível terá uma dificuldade b_2 muito baixa para a proficiência θ_2 em interpretação de texto enquanto terá uma dificuldade b_1 alta para a proficiência θ_1 em Geometria Espacial (por exemplo) se o item foi corretamente elaborado para o nível educacional alvo, de modo que, apesar de ser por definição uma função das duas proficiências, a probabilidade de acerto será função apenas de θ_1 para todos os fins práticos.

No caso dos estudantes de Medicina, é plausível assumir que o processo seletivo de ingresso no curso nivela os alunos em relação às proficiências básicas não médicas, enquanto também é defensável a independência dos domínios do aprendizado (Bloom, 1983) da Medicina, conforme mencionado no penúltimo parágrafo da *Seção Avaliação do Processo Ensino-Aprendizagem*.

Referências Bibliográficas

- ANDRADE, D.F.; TAVARES, H.R.; VALLE, R.C. Teoria da Resposta ao Item: Conceitos e Aplicações. São Paulo: SINAPE, 2000.
- BARBETTA, P.A.; TREVISAN, L.M.V.; TAVARES, H.; AZEVEDO, T.C.A.M. Aplicação da Teoria da Teoria da Resposta ao Item uni e multidimensional, Est. Aval. Educ., São Paulo, v. 25, n. 57, p. 280-302, jan./abr. 2014.
- BLOOM, B.; HASTINGS, J.T.; MADDAUS, G.F. Manual de Avaliação Formativa e Somativa do Aprendizado Escolar. Trad. Lílian Rochlitz Quintão. São Paulo: Livraria Pioneira Editor, 1983.
- BRASIL. Diretrizes Curriculares do Curso de Medicina. Brasília. 2014.
- CHILDS, R. A.; OPPLER, S. H. Implication of test dimensionality for unidimensional IRT scoring: an investigation of a High-Stake Testing Program. Education and Psychological Measurement, v. 60, p. 939-955, 2000. HARDEM, R.M. – Ten questions to ask when planning a course or curriculum, Med. Educ. 20(4): 356-65. 1986.
- MARCONDES, E.; GONÇALVES, E.L. Educação Médica. São Paulo: Editora Sarvier. 1998.
- PIAGET, J. *O possível e o necessário*. Vol. 1: Evolução dos possíveis na criança. Porto Alegre: Artesmédicas. 1985.
- RASCH, G. Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press. 1960.
- RECKASE, M. Multidimensional Item Response Theory. USA: Springer. 2009.
- SONG, Z.; SAFRAN, D.G.; LANDON, B.E.; HE, Y.; ELLIS, R. P.; MECHANIC, R. E.; MATTHEW, P.D.; CHERNEW, M. E. Health Care Spending and Quality in Year 1 of the Alternative Quality Contract; N Engl J Med. 2011; 365(10):909
- SYMPSON, J. A model for testing with multidimensional items. Proc. of the 1977 Computerized Adaptive Testing Conference. 1977.
- WHITELY, S. Measuring aptitude processes with multicomponent latent trait models. Technical Report. Lawrence: University of Kansas. 1980.

ANEXOS

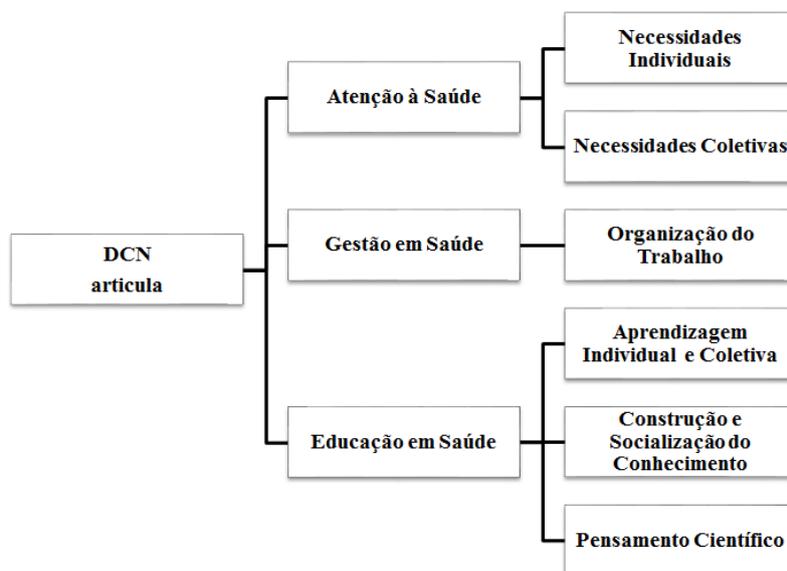


Figura 1: Diretrizes Curriculares Nacionais de Medicina de 2014

Semelhanças e Diferenças	Tipo de Avaliação		
	Diagnóstica	Formativa	Somativa
FUNÇÃO	Determina: a presença/ausência de competências subjacentes; o nível de domínio atual do aluno; causas subjacentes à dificuldades no processo ensino-aprendizagem	Determina: feedback ao aluno/professor em relação ao progresso do processo ensino-aprendizagem ao longo de uma unidade; localização de erros em termos de estrutura de uma unidade, a fim de indicar metodologia para a revisão do processo	Graduação ou atribuição de notas ao final de uma unidade, curso; Classificação em concursos; Certificações
ÉPOCA	Início de uma unidade/processo ensino-aprendizagem de cada competência/domínio	Ao longo da unidade/processo ensino-aprendizagem de cada competência/domínio	Ao final da unidade/processo ensino-aprendizagem de cada competência/domínio
ÊNFASE de AVALIAÇÃO	Escrita: conhecimentos e descrição de habilidades e atitudes Prática: Observação de habilidades e atitudes		
TIPO de INSTRUMENTO	Formativos e somativos para pré-testes; testes padronizados de desempenho ou de diagnóstico; observações e roteiros	Formativos	Finais/Somativos
AMOSTRA dos OBJETIVOS da AVALIAÇÃO	Amostras específicas para cada competência, objetivo, metodologia, instrumento que serão trabalhados no curso; de forma que seja possível considerar um perfil de habilidades e atitudes dos alunos (ideal = todos)	Amostra específica para as competências da MATRIZ CURRICULAR	Amostra específica para as competências da MATRIZ DE REFERÊNCIA (BLUEPRINT)
DIFICULDADE do ITEM	50% FÁCEIS, 25% DIFÍCEIS E 25% MÉDIOS	25% FÁCEIS, 50% MÉDIOS E 25% DIFÍCEIS	25% FÁCEIS, 50% MÉDIOS E 25% DIFÍCEIS
ATRIBUIÇÃO de PONTOS	norma e critério	critério	norma
MÉTODO de RELATO dos ESCORES	Perfil individual de competências subjacentes	Padrão individual de escores relativos a determinados erros em cada etapa do processo ensino-aprendizagem	Escore total da proficiência com os subescores de cada domínio especificados

Figura 2: Tipos de avaliação (adaptado de Bloom, 1983)

ÁREAS	CONHECIMENTOS	ITENS
		1 ANO
CLÍNICA MÉDICA	Conhecimento do processo normal e alterado	3 FÁCEIS; 5 MÉDIOS; 2 DIFÍCEIS
	Compreensão e domínio da propedêutica médica: história clínica, exame físico, conhecimento fisiopatológico	
	Diagnóstico, prognóstico e conduta terapêutica nas doenças	
CLÍNICA CIRÚRGICA	Conhecimento do processo normal e alterado	3 FÁCEIS; 5 MÉDIOS; 2 DIFÍCEIS
	Compreensão e domínio da propedêutica médica: história clínica, exame físico, conhecimento fisiopatológico	
	Diagnóstico, prognóstico e conduta terapêutica nas doenças	
GINECOLOGIA E OBSTETRÍCIA	Conhecimento do processo normal e alterado	3 FÁCEIS; 5 MÉDIOS; 2 DIFÍCEIS
	Compreensão e domínio da propedêutica médica: história clínica, exame físico, conhecimento fisiopatológico	
	Diagnóstico, prognóstico e conduta terapêutica nas doenças	
PEDIATRIA	Conhecimento do processo normal e alterado	3 FÁCEIS; 5 MÉDIOS; 2 DIFÍCEIS
	Compreensão e domínio da propedêutica médica: história clínica, exame físico, conhecimento fisiopatológico	
	Diagnóstico, prognóstico e conduta terapêutica nas doenças	
MENTAL	Conhecimento do processo normal e alterado	3 FÁCEIS; 5 MÉDIOS; 2 DIFÍCEIS
	Compreensão e domínio da propedêutica médica: história clínica, exame físico, conhecimento fisiopatológico	
	Diagnóstico, prognóstico e conduta terapêutica nas doenças	
MEDICINA PREVENTIVA	Conhecimento do processo normal e alterado	3 FÁCEIS; 5 MÉDIOS; 2 DIFÍCEIS
	Compreensão e domínio da propedêutica médica: história clínica, exame físico, conhecimento fisiopatológico	
	Diagnóstico, prognóstico e conduta terapêutica nas doenças	

Figura 3: Sugestão de *Blueprint*

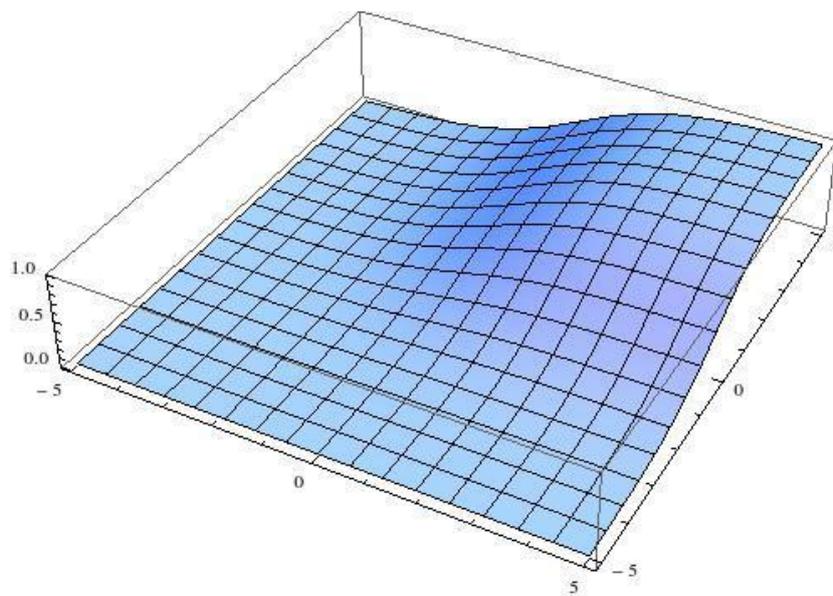


Figura 4: Modelo não-compensatório para θ_1 em $[-5, 5]$ e θ_2 em $[-5, 5]$

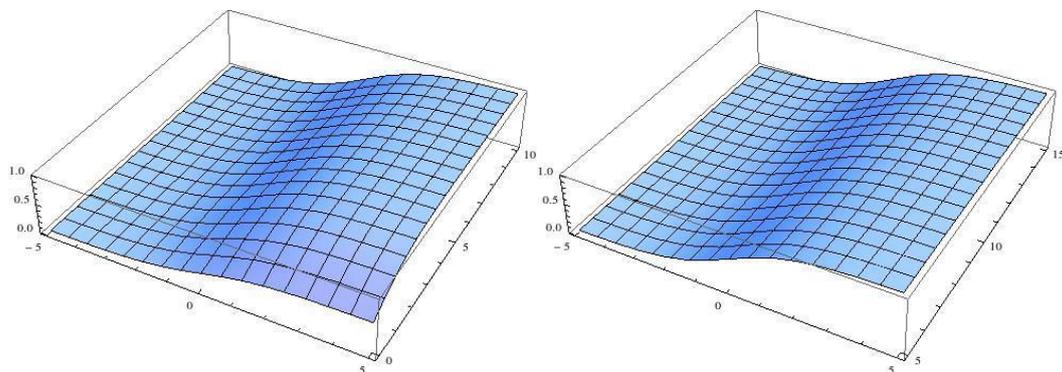


Figura 5: Modelo não-compensatório para (a) θ_2 em $[0, 10]$ e (b) θ_2 em $[0, 10]$

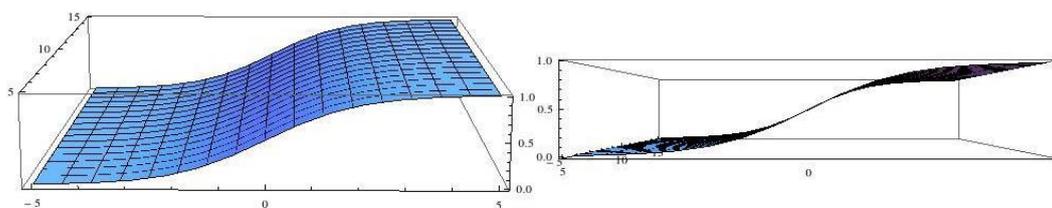


Figura 6: Figura 5b de pontos de vista mais frontais