

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE EDUCAÇÃO

Thiago Fernando Ferreira Costa

**Testes Adaptativos Informatizados (TAI) e desafios da avaliação da  
proficiência em leitura nos anos iniciais do Ensino Fundamental: um  
modelo de análise de alternativas de itens de múltipla escolha como  
contribuição para o sucesso escolar**

São Paulo

2023

**Thiago Fernando Ferreira Costa**

**Testes Adaptativos Informatizados (TAI) e desafios da avaliação da  
proficiência em leitura nos anos iniciais do Ensino Fundamental: um  
modelo de análise de alternativas de itens de múltipla escolha como  
contribuição para o sucesso escolar**

**Versão revisada**

Dissertação apresentada como exigência para obtenção do título de Mestre em Educação junto ao Programa de Pós-Graduação em Educação da Faculdade de Educação da Universidade de São Paulo (Feusp).

Área de Concentração: Estado, Sociedade e Educação.

Orientador: Prof. Ocimar Munhoz Alavarse

São Paulo

2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo da Publicação

Ficha elaborada pelo Sistema de Geração Automática a partir de dados fornecidos pelo(a) autor(a)  
Bibliotecária da FE/USP: Nicolly Soares Leite - CRB-8/8204

CC837t COSTA, Thiago Fernando Ferreira  
Testes Adaptativos Informatizados (TAI) e desafios da avaliação da proficiência em leitura nos anos iniciais do Ensino Fundamental: - um modelo de análise de alternativas de itens de múltipla escolha como contribuição para o sucesso escolar / Thiago Fernando Ferreira COSTA; orientador Ocimar Munhoz ALAVARSE. -- São Paulo, 2023.  
135 p.

Dissertação (Mestrado - Programa de Pós-Graduação Estado, Sociedade e Educação) -- Faculdade de Educação, Universidade de São Paulo, 2023.

1. Políticas de avaliação. 2. Avaliação da aprendizagem. 3. Teste Adaptativo Informatizado. 4. Teoria da Resposta ao Item. 5. Modelo Aninhado. I. ALAVARSE, Ocimar Munhoz, orient. II. Título.

COSTA, Thiago Fernando Ferreira. **Testes Adaptativos Informatizados (TAI) e desafios da avaliação da proficiência em leitura nos anos iniciais do Ensino Fundamental**: um modelo de análise de alternativas de itens de múltipla escolha como contribuição para o sucesso escolar. 2023. 262 f. Dissertação (Mestrado em Educação) – Programa de Pós-Graduação em Educação, Faculdade de Educação da Universidade de São Paulo. São Paulo, 2023.

Aprovado em:

**Banca Examinadora**

Prof. Dr. Héilton Ribeiro Tavares

Instituição: Universidade Federal do Pará

Julgamento: Aprovado

Prof. Dr<sup>a</sup>. Érica Maria Toledo Catalani

Instituição: Cenpec

Julgamento: Aprovado

Prof. Dr. Ocimar Munhoz Alavarse

Instituição: Faculdade de Educação da USP

Julgamento: Aprovado

## **Dedicatória**

À minha querida companheira Juliana, à minha filha Beatriz e aos meus filhos Theo e Gael, por terem respeitado minha ausência, principalmente, nos diversos momentos de confraternização que não pude estar presente, pelo apoio incondicional e pelo incentivo sem contrapartidas em toda essa caminhada.

## **Agradecimentos**

Ao Prof. Ocimar Munhoz Alavarse, por ter acreditado na minha capacidade como futuro pesquisador, com a sua generosidade e paciência, me acolheu nos momentos difíceis, pela sua humildade, em compartilhar as suas experiências de vida e o seu vasto conhecimento em diversas áreas e contribuído com o meu desenvolvimento como pessoa, profissional e na vida acadêmica.

Aos colegas e às colegas do Gepave, que mesmo com a distância imposta pela tragédia que foi a pandemia de Covid-19, conseguiram estar presentes em momentos virtuais de muito aprendizado sobre sistema digital, medida e avaliação educacional; em especial, ao Eduardo Tomita e ao Douglas de Rizzo Meneghetti, por contribuírem, respectivamente, com o desenvolvimento do sistema, incluindo os inúmeros testes e diversas conversas sobre os algoritmos do TAI – Leitura, fundamentais para este trabalho.

Ao Prof. Dr. Heliton Ribeiro Tavares, Dra. Érica Maria Toledo Catalani e Dr. Wolney Candido de Melo, pelas contribuições no Exame de Qualificação.

Às equipes gestoras, aos docentes e discentes das 5 escolas que participaram da pesquisa, sempre com muita paciência e envolvimento nas diversas atividades, sempre muita dedicação e franqueza, na busca por novos caminhos para a avaliação educacional.

## Resumo

COSTA, Thiago Fernando Ferreira. **Testes Adaptativos Informatizados (TAI) e desafios da avaliação da proficiência em leitura nos anos iniciais do Ensino Fundamental**: um modelo de análise de alternativas de itens de múltipla escolha como contribuição para o sucesso escolar. 2023. 262 f. Dissertação (Mestrado em Educação) – Programa de Pós-Graduação em Educação, Faculdade de Educação da Universidade de São Paulo. São Paulo, 2023.

Esta Dissertação resulta de uma pesquisa cujo objetivo principal foi o desenvolvimento de um Teste Adaptativo Informatizado (TAI), com base nos itens da Provinha Brasil (PB), para estimar a proficiência em leitura nos anos iniciais do Ensino Fundamental, considerando, em seus algoritmos, a dificuldade de cada uma das alternativas dos itens, visando favorecer as análises de resultados e os potenciais desdobramentos no ensino da leitura. Como um dos grandes desafios das políticas educacionais brasileiras, ainda encontramos amplos contingentes de alunos, especialmente nos anos iniciais do Ensino Fundamental, com profundas limitações em sua competência leitora, algo que compromete a própria escolarização, ademais de graves consequências sociais. Das várias possibilidades de superar esse desafio, considera-se a necessidade de avaliar a aprendizagem em leitura, em patamares consistentes, com um instrumento que seja um ponto de apoio para processos de ensino. Sendo assim, consideramos o potencial dos TAI, que se colocam como uma ferramenta extremamente útil para o trabalho docente, sobretudo, porque, ao apoiar a prática avaliativa, podem favorecer o desenvolvimento do ensino mais ajustado aos patamares de aprendizado dos alunos e se constituem num expediente político, em face do quadro nacional de domínio da leitura. Em um TAI os itens são selecionados e apresentados aos respondentes de acordo com as respostas dadas anteriormente e, nesse sentido, procuramos a relação que possa existir entre a escolha de cada alternativa e a aproximação mais rápida da estimativa da proficiência, uma das condições de finalização de uma sessão de aplicação do teste, além de uma interpretação pedagógica mais consistente, pois considera o domínio da leitura mesmo diante de eventuais erros, o que torna o teste e seus resultados um ferramenta mais sólida para dar suporte às ações realizadas na sala de aula.

**Palavras-chave:** avaliação da aprendizagem; leitura; proficiência em leitura; Teste Adaptativo Informatizado (TAI); Teoria da Resposta ao Item (TRI); Modelo Aninhado; Teoria Clássica dos Testes (TCT)

## Abstract

COSTA, Thiago Fernando Ferreira. **Computerized Adaptive Testing (CAT) and challenges in assessing reading proficiency in the early years of Elementary School**: an analysis model of multiple-choice item alternatives as a contribution to academic success. 2023. 130 p. Dissertation (Master's in Education) – Graduate Program in Education, Faculty of Education, University of São Paulo. São Paulo, 2023.

This dissertation is the result of a research whose main objective was the development of a Computerized Adaptive Test (CAT), based on items from the Provinha Brasil (PB), to estimate reading proficiency in the early years of Elementary School, considering, in its algorithms, the difficulty of each alternative of the items for their analyses and implications in reading instruction. As one of the great challenges of Brazilian educational policies, there are still large contingents of students, especially in the early years of Elementary School, with deep limitations in their reading competence, something that compromises their schooling itself, besides having serious social consequences. Among the various possibilities to overcome this challenge, it is considered necessary to evaluate reading learning, at consistent levels, with an instrument that can support teaching processes. Therefore, we consider the potential of CATs, which are an extremely useful tool for teaching work, especially because, by supporting evaluative practice, they can favor the development of teaching that is more tailored to students and constitute a political resource, given the national framework of reading proficiency. In a CAT, items are selected and presented to respondents according to the answers given previously, and, in this sense, we seek the relationship that may exist between the choice of each alternative and the quicker estimation of proficiency, one of the conditions for completing a test session, in addition to a more consistent pedagogical interpretation, as it considers reading proficiency even in the face of possible errors, making the test and its results a more solid tool to support actions taken in the classroom.

**Keywords:** learning assessment; reading; reading proficiency; Computerized Adaptive Test (CAT); Item Response Theory (IRT); Nested Model; Classical Test Theory (CTT)



## Lista de figuras

Figura 1 – Provinha Brasil. Questão 5. Teste 1. 2013 .....	31
Figura 2 – Provinha Brasil. Questão 4. Teste 2. 2016 .....	33
Figura 3 – Provinha Brasil. Questão 11. Teste 2. 2015.....	34
Figura 4 – Provinha Brasil. Quantidade de acertos e níveis de desempenho. 2016.....	36
Figura 5 – Provinha Brasil. Interpretação pedagógica dos Níveis 1, 2 e 3. Teste 2. 2016 ....	37
Figura 6 – Distribuição dos desempenhos dos alunos da RME-SP nos Testes 1 e 2. 2013 ...	41
Figura 7 – Distribuição dos desempenhos dos alunos da RME-SP. Testes 1 e 2. 2015 .....	42
Figura 8 – Curva Informação do Item em conjunto com a Curva Característica do Item ....	65
Figura 9 – Esquema para representar a relação entre quantidade de grupos e provas .....	67
Figura 10 – Comparação das Curvas de Informação do Teste (CIT). Modelos Rasch, 2PL, Nominal em comparação com o Aninhado.....	70
Figura 11 – Curva Característica do Item para o modelo Aninhado .....	74
Figura 12 – Diagrama do fluxo de um Teste Adaptativo.....	79
Figura 13 – Provinha Brasil – Leitura. Interpretação pedagógica do Nível 2.....	132

## Lista de tabelas

Tabela 1 – Provinha Brasil. Quantidade de alunos em cada aplicação na RME-SP. 2012-2015.....	39
Tabela 2 – Modelos logísticos da TRI.....	58
Tabela 3 – Modelos Nominal e Aninhado da TRI para itens politômicos.....	61
Tabela 4 – Modelos politômicos da TRI .....	63
Tabela 5 – Informações gerais sobre a análise estatística da base simulada.....	85
Tabela 6 – Informações gerais sobre a simulação com o TAI-Leitura 3.0.....	86
Tabela 7 – Informações sobre a simulação do TAI-Leitura 3.0 pseudoB.....	87
Tabela 8 – Distribuição da quantidade de itens por nível em ambas as simulações.....	88
Tabela 9 – TAI-Leitura 3.0. Aplicação na Escola Zimbábue. Março 2023.....	90
Tabela 10 – TAI-Leitura 3.0. Média de itens respondidos por nível. Escola Zimbábue. Março 2023.....	90
Tabela 11 – TAI-Leitura pseudoB. Aplicação em Escolas parceiras. Março 2023.....	91
Tabela 12 – TAI-Leitura pseudoB. Média dos itens respondidos. Escolas parceiras. Março 2023.....	92
Tabela 13 – TAI-Leitura 3.3. Quantidade e percentual de alunos participantes. Escolas Mauritània, Senegal e Zimbábue. Abril 2023.....	102
Tabela 14 – TAI-Leitura 3.0. Distribuição dos alunos por escola em cada nível. Abril 2023.....	103
Tabela 15 – TAI-Leitura 3.0. Distribuição dos alunos nos dois algoritmos. Abril 2023.....	104
Tabela 16 – TAI-Leitura 3.0. Distribuição da média de itens aplicados por nível. Abril 2023.....	105
Tabela 17 – TAI-Leitura 3.0. Quantidade de alunos no Nível 1, médias de proficiência e erro padrão por ano escolar. Abril 2023.....	106
Tabela 18 – TAI-Leitura 3.0. Quantidade de alunos no Nível 2, médias de proficiência e erro padrão por ano escolar. Abril 2023.....	107

Tabela 19 – TAI-Leitura 3.0. Quantidade de alunos no Nível 3, médias de proficiência e erro padrão por ano escolar. Abril 2023.....	108
Tabela 20 – TAI-Leitura 3.0. Quantidade de alunos no Nível 4, médias de proficiência e erro padrão por ano escolar. Abril 2023.....	109
Tabela 21 – TAI-Leitura 3.0. Quantidade de alunos no Nível 5, médias de proficiência e erro padrão por ano escolar. Abril 2023.....	110
Tabela 22 – TAI-Leitura 3.0. Informações pedagógicas dos 5 primeiros itens dos testes dos alunos A5429 e A3809. Abril 2023 .....	134

## Lista de Gráficos

Gráfico 1 – Provinha Brasil. Distribuição do desempenho dos alunos da RME-SP por níveis. Teste 2. 2012.....	40
Gráfico 2 – Distribuição dos desempenhos dos alunos da RME-SP no Teste 2 de 2014.....	41
Gráfico 3 – Distribuição do percentual das alternativas mais escolhidas por nível e o percentual de acerto por item. Teste 2. 2012.....	43
Gráfico 4 – Distribuição do percentual das alternativas mais escolhidas por nível e o percentual de acerto por item no Teste 2 de 2013.....	45
Gráfico 5 – Distribuição do percentual das alternativas mais escolhidas por nível e o percentual de acerto por item. Teste 2. 2014.....	46
Gráfico 6 – Distribuição do percentual das alternativas mais escolhidas por nível e o percentual de acerto por item. Teste 2. 2015.....	47
Gráfico 7 – Curva Característica de um Item (CCI) para o modelo logístico de 3 parâmetros.....	59
Gráfico 8 – Distribuição dos parâmetros b dos 120 itens da PB.....	73
Gráfico 9 – Distribuição das proficiências estimadas por meio do pacote “mirt” .....	84
Gráfico 10 – TAI-Leitura. Distribuição das proficiências estimadas .....	88
Gráfico 11 – TAI-Leitura 3.0. Distribuição das escolhas por alternativa no item exemplo.....	93
Gráfico 12 – TAI-Leitura 3.0. Distribuição das escolhas por alternativa no item 1.....	94
Gráfico 13 – TAI-Leitura 3.0. Distribuição no Nível 1 dos pseudoB das alternativas.....	96
Gráfico 14 – TAI-Leitura 3.0. Distribuição no Nível 2 dos pseudoB das alternativas.....	97
Gráfico 15 – TAI-Leitura 3.0. Distribuição no Nível 3 dos pseudoB das alternativas.....	98
Gráfico 16 – TAI-Leitura 3.0. Distribuição no Nível 4 dos pseudoB das alternativas.....	99
Gráfico 17 – TAI-Leitura 3.0. Distribuição no Nível 5 dos pseudoB das alternativas.....	100
Gráfico 18 – TAI-Leitura 3.0. Distribuição dos parâmetros b e pseudoB dos itens do Banco de Itens.....	111
Gráfico 19 – TAI-Leitura 3.0. Distribuição no Nível 1 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023.....	112
Gráfico 20 – TAI-Leitura 3.0. Distribuição no Nível 2 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023.....	113
Gráfico 21 – TAI-Leitura 3.0. Distribuição no Nível 3 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023.....	114

Gráfico 22 – TAI-Leitura 3.0. Distribuição no Nível 4 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023.....	115
Gráfico 23 – TAI-Leitura 3.0. Distribuição no Nível 5 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023.....	116
Gráfico 24 – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 1. Abril 2023.....	117
Gráfico 25 – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 1. Abril 2023 .....	118
Gráfico 26 – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 2. Abril 2023.....	120
Gráfico 27 – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 2. Abril 2023 .....	121
Gráfico 28 – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 3. 2023.....	122
Gráfico 29 – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 3. Abril 2023 .....	123
Gráfico 30 – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 4. Abril 2023 .....	124
Gráfico 31 – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 4. Abril 2023.....	125
Gráfico 32 – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 5. Abril 2023.....	126
Gráfico 33 – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 5. Abril 2023 .....	127
Gráfico 34 – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 1. Abril 2023.....	128
Gráfico 35 – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 2. Abril 2023.....	129
Gráfico 36 – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 3. Abril 2023.....	130
Gráfico 37 – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 4. Abril 2023.....	130

Gráfico 38 – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 5. Abril 2020 .....	131
Gráfico 39 – Função de Probabilidade – Item 01 .....	142
Gráfico 40 – Função de Probabilidade – Item 02 .....	143
Gráfico 41 – Função de Probabilidade – Item 03 .....	144
Gráfico 42 – Função de Probabilidade – Item 04 .....	145
Gráfico 43 – Função de Probabilidade – Item 05 .....	146
Gráfico 44 – Função de Probabilidade – Item 06 .....	147
Gráfico 45 – Função de Probabilidade – Item 07 .....	148
Gráfico 46 – Função de Probabilidade – Item 08 .....	149
Gráfico 47 – Função de Probabilidade – Item 09 .....	150
Gráfico 48 – Função de Probabilidade – Item 10 .....	151
Gráfico 49 – Função de Probabilidade – Item 11 .....	152
Gráfico 50 – Função de Probabilidade – Item 12 .....	153
Gráfico 51 – Função de Probabilidade – Item 13 .....	154
Gráfico 52 – Função de Probabilidade – Item 14 .....	155
Gráfico 53 – Função de Probabilidade – Item 15 .....	156
Gráfico 54 – Função de Probabilidade – Item 16 .....	157
Gráfico 55 – Função de Probabilidade – Item 17 .....	158
Gráfico 56 – Função de Probabilidade – Item 18 .....	159
Gráfico 57 – Função de Probabilidade – Item 19 .....	160
Gráfico 58 – Função de Probabilidade – Item 20 .....	161
Gráfico 59 – Função de Probabilidade – Item 21 .....	162
Gráfico 60 – Função de Probabilidade – Item 22 .....	163
Gráfico 61 – Função de Probabilidade – Item 23 .....	164
Gráfico 62 – Função de Probabilidade – Item 24 .....	165
Gráfico 63 – Função de Probabilidade – Item 25 .....	166
Gráfico 64 – Função de Probabilidade – Item 26 .....	167
Gráfico 65 – Função de Probabilidade – Item 27 .....	168
Gráfico 66 – Função de Probabilidade – Item 28 .....	169
Gráfico 67 – Função de Probabilidade – Item 29 .....	170
Gráfico 68 – Função de Probabilidade – Item 30 .....	171
Gráfico 69 – Função de Probabilidade – Item 31 .....	172

Gráfico 70 – Função de Probabilidade – Item 32 .....	173
Gráfico 71 – Função de Probabilidade – Item 33 .....	174
Gráfico 72 – Função de Probabilidade – Item 34 .....	175
Gráfico 73 – Função de Probabilidade – Item 35 .....	176
Gráfico 74 – Função de Probabilidade – Item 36 .....	177
Gráfico 75 – Função de Probabilidade – Item 37 .....	178
Gráfico 76 – Função de Probabilidade – Item 38 .....	179
Gráfico 77 – Função de Probabilidade – Item 39 .....	180
Gráfico 78 – Função de Probabilidade – Item 40 .....	181
Gráfico 79 – Função de Probabilidade – Item 41 .....	182
Gráfico 80 – Função de Probabilidade – Item 42 .....	183
Gráfico 81 – Função de Probabilidade – Item 43 .....	184
Gráfico 82 – Função de Probabilidade – Item 44 .....	185
Gráfico 83 – Função de Probabilidade – Item 45 .....	186
Gráfico 84 – Função de Probabilidade – Item 46 .....	187
Gráfico 85 – Função de Probabilidade – Item 47 .....	188
Gráfico 86 – Função de Probabilidade – Item 48 .....	189
Gráfico 87 – Função de Probabilidade – Item 49 .....	190
Gráfico 88 – Função de Probabilidade – Item 50 .....	191
Gráfico 89 – Função de Probabilidade – Item 51 .....	192
Gráfico 90 – Função de Probabilidade – Item 52 .....	193
Gráfico 91 – Função de Probabilidade – Item 53 .....	194
Gráfico 92 – Função de Probabilidade – Item 54 .....	195
Gráfico 93 – Função de Probabilidade – Item 55 .....	196
Gráfico 94 – Função de Probabilidade – Item 56 .....	197
Gráfico 95 – Função de Probabilidade – Item 57 .....	198
Gráfico 96 – Função de Probabilidade – Item 58 .....	199
Gráfico 97 – Função de Probabilidade – Item 59 .....	200
Gráfico 98 – Função de Probabilidade – Item 60 .....	201
Gráfico 99 – Função de Probabilidade – Item 61 .....	202
Gráfico 100 – Função de Probabilidade – Item 62 .....	203
Gráfico 101 – Função de Probabilidade – Item 63 .....	204
Gráfico 102 – Função de Probabilidade – Item 64 .....	205

Gráfico 103 – Função de Probabilidade – Item 65 .....	206
Gráfico 104 – Função de Probabilidade – Item 66 .....	207
Gráfico 105 – Função de Probabilidade – Item 67 .....	208
Gráfico 106 – Função de Probabilidade – Item 68 .....	209
Gráfico 107 – Função de Probabilidade – Item 69 .....	210
Gráfico 108 – Função de Probabilidade – Item 70 .....	211
Gráfico 109 – Função de Probabilidade – Item 71 .....	212
Gráfico 110 – Função de Probabilidade – Item 72 .....	213
Gráfico 111 – Função de Probabilidade – Item 73 .....	214
Gráfico 112 – Função de Probabilidade – Item 74 .....	215
Gráfico 113 – Função de Probabilidade – Item 75 .....	216
Gráfico 114 – Função de Probabilidade – Item 76 .....	217
Gráfico 115 – Função de Probabilidade – Item 77 .....	218
Gráfico 116 – Função de Probabilidade – Item 78 .....	219
Gráfico 117 – Função de Probabilidade – Item 79 .....	220
Gráfico 118 – Função de Probabilidade – Item 80 .....	221
Gráfico 119 – Função de Probabilidade – Item 81 .....	222
Gráfico 120 – Função de Probabilidade – Item 82 .....	223
Gráfico 121 – Função de Probabilidade – Item 83 .....	224
Gráfico 122 – Função de Probabilidade – Item 84 .....	225
Gráfico 123 – Função de Probabilidade – Item 85 .....	226
Gráfico 124 – Função de Probabilidade – Item 86 .....	227
Gráfico 125 – Função de Probabilidade – Item 87 .....	228
Gráfico 126 – Função de Probabilidade – Item 88 .....	229
Gráfico 127 – Função de Probabilidade – Item 89 .....	230
Gráfico 128 – Função de Probabilidade – Item 90 .....	231
Gráfico 129 – Função de Probabilidade – Item 91 .....	232
Gráfico 130 – Função de Probabilidade – Item 92 .....	233
Gráfico 131 – Função de Probabilidade – Item 93 .....	234
Gráfico 132 – Função de Probabilidade – Item 94 .....	235
Gráfico 133 – Função de Probabilidade – Item 95 .....	236
Gráfico 134 – Função de Probabilidade – Item 96 .....	237
Gráfico 135 – Função de Probabilidade – Item 97 .....	238



Gráfico 136 – Função de Probabilidade – Item 98 .....	239
Gráfico 137 – Função de Probabilidade – Item 99 .....	240
Gráfico 138 – Função de Probabilidade – Item 100 .....	241
Gráfico 149 – Função de Probabilidade – Item 101 .....	242
Gráfico 140 – Função de Probabilidade – Item 102 .....	243
Gráfico 141 – Função de Probabilidade – Item 103 .....	244
Gráfico 142 – Função de Probabilidade – Item 104 .....	245
Gráfico 143 – Função de Probabilidade – Item 105 .....	246
Gráfico 144 – Função de Probabilidade – Item 106 .....	247
Gráfico 145 – Função de Probabilidade – Item 107 .....	248
Gráfico 146 – Função de Probabilidade – Item 108 .....	249
Gráfico 147 – Função de Probabilidade – Item 109 .....	250
Gráfico 148 – Função de Probabilidade – Item 110 .....	251
Gráfico 149 – Função de Probabilidade – Item 111 .....	252
Gráfico 150 – Função de Probabilidade – Item 112 .....	253
Gráfico 160 – Função de Probabilidade – Item 113 .....	254
Gráfico 161 – Função de Probabilidade – Item 114 .....	255
Gráfico 162 – Função de Probabilidade – Item 115 .....	256
Gráfico 163 – Função de Probabilidade – Item 116 .....	257
Gráfico 164 – Função de Probabilidade – Item 117 .....	258
Gráfico 165 – Função de Probabilidade – Item 118 .....	259
Gráfico 166 – Função de Probabilidade – Item 119 .....	260
Gráfico 167 – Função de Probabilidade – Item 120 .....	261

## Sumário

Introdução .....	19
1 – Avaliação educacional: a Provinha Brasil como base para o TAI .....	27
1.1 – Provinha Brasil: estrutura, características e um breve histórico .....	29
1.2 – Os resultados da Provinha Brasil – Leitura na Rede Municipal de Educação de São Paulo (2012 a 2015) .....	38
2 – Medidas educacionais e os desafios para descobrir o que sabem os alunos: alguns delineamentos e resultados .....	49
2.1 – As duas teorias mais relevantes para as medidas educacionais: a TCT e a TRI.....	51
2.1.1 – Teoria Clássica dos Testes (TCT): algumas características .....	52
2.1.2 – Teoria da Resposta ao Item (TRI): algumas características .....	56
2.1.3 – O número de populações envolvidas .....	66
2.2 – Processamento estatístico para definição de um modelo da TRI: análise dos dados da Provinha Brasil de Leitura aplicada na Rede Municipal de Ensino de São Paulo .....	68
2.2.1 – Definição do modelo da TRI: processamento da Provinha Brasil primeira aplicação de 2015.....	68
2.2.2 – Análise com o modelo Aninhado: calibração dos itens que compõem o Banco de Itens do TAI-Leitura 3.0 .....	71
3 – TAI-Leitura na perspectiva da análise das alternativas.....	76
3.1 – Breve histórico sobre o TAI-Leitura .....	76
3.2 – As principais características do TAI-Leitura.....	77
3.3 – TAI-Leitura 3.0 e pseudoB: as principais diferenças desses dois testes .....	81
4 – Análise das simulações e dos pré-testes do TAI-Leitura 3.0 .....	83
4.1 – A estrutura da base simulada e resultados descritivos das aplicações .....	83
4.2 – O pré-teste nas escolas e os resultados descritivos das aplicações .....	89
4.3 – Análise comparativa das aplicações: o que podemos concluir?.....	92
5 – Análise dos resultados da aplicação do TAI-Leitura 3.0 .....	102
5.1 – Interpretação pedagógica: uma possibilidade de análise das alternativas em um TAI-Leitura .....	132
Considerações finais e perspectivas futuras .....	136
Referências .....	139
Anexo – Gráficos de Função de Probabilidade dos itens .....	142

## Introdução

A leitura é uma competência decisiva, social e academicamente, e sua aprendizagem é antes de tudo um desafio político da democratização da escola. Neste trabalho entendemos que os avanços na avaliação da aprendizagem da competência leitora podem contribuir para seu ensino. Neste sentido, melhorar processos de avaliação da aprendizagem, especialmente quanto aos instrumentos e procedimentos, é uma contribuição que se pretende aportar, como fruto do desenvolvimento da pesquisa da qual decorre esta Dissertação de Mestrado. Nesta perspectiva, o Teste Adaptativo Informatizado (TAI) em leitura – TAI-Leitura –, um instrumento avaliativo que se colocou como objeto de investigação, ainda que demande novos aperfeiçoamentos, é uma contribuição, especialmente em face do problema de pesquisa.

A construção e implementação de um TAI é processo intrinsecamente multidisciplinar, pois requer a participação de professoras e professores, que são os responsáveis pela ação pedagógica; de estatísticos e psicometristas, responsáveis pelo processamento e análise da medida; de desenvolvedores de softwares, que desenvolvem a plataforma utilizada para a aplicação do TAI; de responsáveis pelas redes de computadores, de suma importância quando se pretende aplicar um TAI.

Um TAI é um dispositivo eletrônico capaz de selecionar itens de um Banco de Itens (BI) e apresentá-los aos alunos que respondem ao teste levando em conta uma ou mais respostas dadas aos itens – usualmente denominados como “questões” em ambientes escolares – apresentados anteriormente ao respondente, podendo acarretar a seleção de próximo(s) item(ns) de maior ou menor complexidade, tornando o teste ajustado à proficiência – usualmente denominada de domínio quanto ao conteúdo do teste – do aluno. O primeiro item do teste pode ser selecionado aleatoriamente ou atender a uma condição e a finalização da aplicação do teste também atenderá a uma ou mais condição(ões).

Nestes termos, um TAI rompe com a característica dos testes convencionais, nos quais o conjunto de itens ou mesmo a sequência de apresentação se encontram definidos a priori, independentemente das respostas dos respondentes, pois o teste vai sendo construído de acordo com as respostas dadas. Assim, um TAI é um teste que se adapta ao respondente, administrado por algoritmos informatizados, a partir de um BI com itens que devem abranger a matriz de referência de referência relativa à competência do teste, com uma ampla gama de dificuldade. Desse modo, consegue chegar a resultados mais precisos, ou seja, com maior confiabilidade

contando, geralmente, com menor quantidade de itens apresentados e, conseqüentemente, com menor tempo de aplicação. Como consequência, um TAI deve facilitar e qualificar a tarefa avaliativa, por exemplo, de um professor.

Os TAI implementados têm mostrado maior flexibilidade, adaptabilidade, redução de tempo de teste, resultados imediatos e maior precisão em relação aos testes tradicionais, como destaca, entre outros autores, Olea (1999). Com base nesses diferenciais, os TAI se revelam como um avanço para o campo da avaliação de um modo geral e, especialmente, na educação. Ainda que não se vislumbre que um TAI vá superar, por si só, vários impasses na área da avaliação educacional, constatadas várias restrições nos processos avaliativos conduzidos por professores e em larga escala, considera-se que sua utilização, também, possa ser um ponto de apoio à reflexão sobre essas limitações de práticas avaliativas, inclusive, sobre os próprios instrumentos avaliativos, que muitas vezes acabam sendo fontes de imprecisões.

Assim, alinhados com o que apresenta Alavarse (2013, p. 105), “o emprego do TAI é uma forma de testar o conhecimento de um respondente que, durante o teste, recebe um determinado item cujo o nível de dificuldade é ajustado e dependente da resposta dada ao item anteriormente respondido”, atenuando-se vários efeitos de testes que são produzidos para muitos alunos em geral, mas que não atendem demandas de alunos específicos, algo que pode gerar um menor engajamento na realização do teste, condição necessária para, por exemplo, estimar sua proficiência em determinada competência.

Apoiado no princípio acima, para a construção de um TAI, alguns elementos básicos precisam ser garantidos e, segundo Olea e Ponsoda (2002, p. 5), são os seguintes:

- a) um banco de itens com propriedades psicométricas conhecidas, com os parâmetros estimados por um modelo da Teoria de Resposta ao Item; b) um procedimento que estabeleça a maneira de começar e finalizar a prova, assim como a forma de selecionar progressivamente os melhores itens; e c) um método estatístico de estimação dos níveis de proficiência. Esses elementos, além de definirem o cerne do TAI, trazem consigo a complexidade que se apresenta ao conceber essa ferramenta como uma realidade no dia a dia das avaliações.

Outrossim, a avaliação educacional necessita da delimitação a mais precisa possível de seu objeto, aquilo que se pretende avaliar, que, no caso do TAI-Leitura trata-se da proficiência em leitura. Isso, por sua vez, destaca a extrema relevância do BI. A leitura tem fundamental importância para a Educação Básica, principalmente, para os anos de escolaridade que compõem o início da alfabetização, por ser um momento crucial na formação dos alunos. A

universalização do acesso à escola tem se mostrado eficaz, porém, com esse aumento no atendimento, surgiram novos desafios, como, por exemplo, o nível do aprendizado dos alunos ao longo da escolarização, notadamente quanto à capacidade de leitura, fazendo com que a alfabetização continue sendo um desafio ainda a ser enfrentado. Para Mernarbini e Gomes (2019, p. 7), o fracasso escolar, por muito tempo, deu-se devido ao grande número de crianças fora da escola, no entanto, com as políticas de universalização do ensino,

[...] constata-se uma nova realidade em que os alunos permanecem no sistema educacional por vários anos sem aprender a ler e escrever, ou adquirem conhecimentos e competências de leitura e escrita em nível instrumental, na condição de analfabeto funcional. Esta condição pode, muitas vezes, auxiliar nas necessidades diárias ou no mundo do trabalho, mas, inevitavelmente, implica na exclusão da participação de outros campos de poder e, sobretudo, impede o exercício consciente e crítico da cidadania. Esse aspecto demonstra o quanto é desafiador a democratização da escola, pois não é suficiente a garantia do acesso como prerrogativa para melhoria da qualidade do ensino. Nessa perspectiva, apresentamos a leitura como uma competência decisiva, capaz de influenciar em toda trajetória social e acadêmica dos alunos. Argumentamos, ainda, que uma melhor avaliação da aprendizagem em leitura pode favorecer processos de ensino que visem o sucesso de todos e considerando, ainda, que muitos instrumentos utilizados em sala de aula e mesmo em larga escala não atendem todas as condições para considerá-los à altura de fornecer informações que facilitem a função formativa da avaliação da aprendizagem.

O TAI-Leitura que consideramos como ponto de partida em nossa pesquisa foi desenvolvido com base no instrumento a Provinha Brasil (PB), focada, centralmente, em alunos do 2º ano do Ensino Fundamental, mas que apresenta, por ser um instrumento para aferir a competência leitora, potencial para todo o período dos anos iniciais e, até mesmo, finais dessa etapa da Educação Básica. A PB foi lançada em 2008 pelo Instituto Nacional de Estudos e Pesquisa Educacionais Anísio Teixeira (Inep), autarquia do Ministério da Educação (MEC) e era oferecida às escolas semestralmente até o segundo semestre de 2016, quando houve a sua descontinuação.

O cerne da PB era propiciar o julgamento rápido do que o aluno aprendeu, ou deveria ter aprendido, conforme colocam Alavarse, Catalani, Meneghetti e Travitzki (2017, p. 69), para quem,

O principal objetivo da PB é permitir o diagnóstico rápido e mais preciso dos problemas no processo de aprendizagem e possibilitar instrumento padronizado que auxilie em intervenções potencializadoras para que o

processo inicial de alfabetização seja concretizado até o final do terceiro ano do ensino fundamental.

Assim, a preferência pela PB de Língua Portuguesa se justifica por ter como foco a avaliação de resultados da alfabetização e do letramento em Língua Portuguesa, ser voltada para o 2º ano e, especialmente, por : estar pautada em uma matriz de avaliação, que expressa as características gerais da competência leitora e apoia toda a elaboração dos itens da prova; apresentar uma escala, que permite obter o resultado da avaliação e verificar o progresso desse resultado ao longo do tempo; ter um conjunto expressivo de instrumentos, com itens que cobrem a matriz de avaliação; apresentar um critério para o processo de julgamento e; uma escala interpretada pedagogicamente, possibilitando análise de resultados para além do valor numérico e que tenham um sentido pedagógico para as professoras e professores.

Para Mortatti (2013, p. 28), a PB apresenta implicitamente a “marcha sintética”<sup>1</sup> como método de alfabetização, mesmo que alfabetização e letramento sejam inseparáveis, e que um aluno será considerado alfabetizado quando

[...] tiver alcançado o “Nível 5”, ou seja, que tiver se “apropriado do sistema de escrita”, conforme a teoria da psicogênese da língua escrita. E, do professor, demanda-se trabalho pedagógico com objetivo de “expandir as capacidades relativas ao letramento, que envolvem a compreensão e o uso de textos variados, com estrutura mais complexa, com temas diversificados, e que circulem em diferentes esferas sociais.

A autora apresenta questionamentos ao modelo implementado por diversos governos (Federal, Estadual e Municipal) no Brasil em relação à avaliação externa, pois em algumas delas há uma lógica, com interferências nas relações intraescolares por conta de excessiva responsabilização dos professores, principalmente, quando esteadas na bonificação. Mortatti (2013) apresenta a inevitabilidade de se discutir, no século XXI, o conteúdo do que ela chama de “caixa-preta da alfabetização”, que por sua vez, incorpora essa lógica das avaliações externas. Dessa maneira, aprofundar os estudos sobre avaliação da alfabetização particularmente da leitura está na esteira para a introdução do TAI no dia a dia da escola.

---

<sup>1</sup> A autora apresenta os descritores da PB: reconhecimento de letras, das sílabas e estabelecimento de relações entre fonemas e grafemas; leitura de palavras e frases, localização de informação explícita em textos, reconhecimento de assunto de um texto, identificação da finalidade do texto, estabelecimento de relação entre parte do texto e inferência de informação, como parte do método de marcha sintética, em que se vai das partes para o todo.

Com base nessa premissa, solicitamos para o Inep as bases de dados dos pré-testes de itens da PB, entre os anos de 2009 e 2015, os dados coletados pela Rede Municipal de Educação de São Paulo (RME-SP) no período de 2012 a 2015 e a base de dados do sistema que o Inep criou para coletar, pela primeira vez, os dados da PB em todo o país. Dados que, recebidos, foram utilizados como parte das fontes da pesquisa, além dos dados provenientes da aplicação de um pré-teste e de uma aplicação em três escolas parceiras do Grupo de Estudos e Pesquisa em Avaliação Educacional (Gepave).

O desenvolvimento do TAI-Leitura, focado neste trabalho, teve como relevante referência a pesquisa de Catalani (2019), que abordou a construção de um TAI focado na PB e submetido a mais de 2 mil alunos do 1º e 2º anos de 15 escolas da RME-SP, em 2016, no qual as respostas aos itens eram consideradas corretas ou incorretas, isto é, se foram assinaladas as alternativas-gabarito, para a primeira consideração, ou as demais, para a outra. No transcurso da pesquisa esse foi denominado de TAI-Leitura 1.0. Posteriormente, com o aperfeiçoamento do algoritmo, no âmbito da pesquisa chegou-se ao TAI-Leitura 2.0, porém mantido o tratamento dicotômico das respostas.

O diferencial do presente trabalho está na construção de um TAI que estamos intitulado de TAI-Leitura 3.0 no qual são consideradas as respostas a todas as alternativas do item, sendo agora estas tratadas de modo a estimar a proficiência do respondente conforme a alternativa assinalada. Com efeito, cada alternativa passou a ter um parâmetro de dificuldade, conforme se apresentará nesta Dissertação, e com isso criou-se a possibilidade de diferenciar os erros, que no modelo dicotômico não ocorre. O estudo das respostas às alternativas corretas e incorretas busca aproveitar, na análise estatística, o sentido pedagógico para o processo que o aluno constituiu ao longo da realização da prova e, adicionalmente, considerando o trabalho do elaborador de itens em sua totalidade, visto que nessa elaboração são indicadas as justificativas para as respostas consideradas incorretas.

A metodologia de tratamento estatístico implementada no Brasil, sobretudo com o Saeb, na década de 1990, e que reverbera até os dias de hoje, desconsidera as alternativas que não sejam corretas, atribuindo 0 para elas no processamento estatístico, dicotomizando as respostas em corretas e incorretas, sem que pudesse haver nuances de domínio entre os respondentes de acordo com a alternativa incorreta assinalada.

De certa monta, levar em consideração as marcações das alternativas do item, para além da dicotomia 0 ou 1, traz a necessidade de se realizar pelo menos três mudanças nos procedimentos de tratamento estatístico e psicométrico, sendo elas: a utilização de um novo modelo da Teoria da Resposta ao Item (TRI) para obtenção dos parâmetros dos itens e das alternativas; uma nova maneira de ancorar os itens, levando em consideração as alternativas, uma vez que cada item estaria ancorado na escala de proficiência conforme a quantidade de alternativas; e a construção de uma escala de proficiência com interpretação pedagógica contemplando todas as alternativas dos itens. Neste trabalho não serão abordadas todas essas necessidades, mas importa destacá-las como investigações futuras que se abrem com o TAI-Leitura 3.0.

A utilização de um novo modelo da TRI se faz necessária para que as alternativas recebam parâmetros, algo que não ocorre na modelagem que o Inep utiliza para processar estatisticamente respostas a itens de múltipla escolha como, por exemplo, no Saeb e no Enem. A TRI é uma teoria que reúne diversos modelos matemáticos que possibilitam obter a medida de proficiência do aluno em relação ao traço latente e, de acordo com Andrade, Tavares e Valle (2000, p. 3), essa teoria

[...] propõe modelos para os traços latentes, ou seja, características do indivíduo que não podem ser observadas diretamente. Esse tipo de variável deve ser inferida a partir da observação de variáveis secundárias que estejam relacionadas a ela. O que esta metodologia sugere são formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item e seus traços latentes, proficiências ou habilidades na área de conhecimento avaliada.

Por relacionar a probabilidade de acertar um item ao traço latente do aluno, a TRI possibilita algo que a Teoria Clássica dos Testes (TCT) não pode garantir: a troca de itens do instrumento de maneira comparável. Como a TRI está especificamente voltada para o item e o tem como elemento central, viabiliza a comparação entre populações diferentes, mas que fazem provas com alguns itens iguais ou, uma mesma população submetida às provas diferentes. Logo, a TRI viabiliza para uma prova a modificação do instrumento, não afetando as comparações entre alunos de anos/séries diferentes, no mesmo ano letivo, como também, ao longo dos anos letivos.

Nos modelos adotados pelo Inep, contudo, na análise estatística, a alternativa considerada correta para o item recebe o valor 1 e as demais alternativas, consideradas incorretas, e recebem o valor 0. Essa forma de tratamento faz com que o potencial pedagógico de análise das respostas dadas aos itens, observando-se erros, perca a possibilidade de diferenciar o “nível do erro”, ou



seja, iguala-se o que, tudo indica, não é igual, pois os erros não só se apoiam em algum conhecimento com, sobretudo, em conhecimentos diferentes para erros diferentes. No entanto, entendemos que a análise que professores fazem das respostas de alunos consideram os erros, fazendo com que o tratamento estatístico se aproxime da prática cotidiana de professoras e de professores.

Dessa forma, propomos problematizar a utilização de um TAI para a avaliação da proficiência em leitura, por parte das professoras e professores, contribuindo com a reflexão sobre os processos cognitivos realizados ao longo de sua aplicação, além de fornecer a proficiência leitora dos alunos, utilizando as alternativas dos itens nesse procedimento. O que pode potencializar o uso dos resultados dos instrumentos de coletas de informações relacionadas a competência leitora, como também, contribuir com as discussões sobre as diferentes formas de acompanhar o avanço do aprendizado dos alunos.

Sendo assim, esta Dissertação tem como objetivo geral apresentar um teste desenvolvido – o TAI-Leitura 3.0 – para verificar a proficiência em leitura, utilizando itens da Provinha Brasil de Língua Portuguesa, com a calibração das alternativas incorretas para serem consideradas nas estimativas das proficiências e nos algoritmos que constituem o TAI. Ademais, apresentamos a organização de análises estatísticas que possibilitaram a definição do modelo da TRI mais ajustado para o objetivo, as simulações com os algoritmos do TAI-Leitura 3.0 e as aplicações, para alunos do Ensino Fundamental, de um pré-teste em uma das escolas parceiras do projeto e de outra, para alunos de três escolas que colaboram junto ao Gepave nos estudos sobre avaliação educacional.

A pesquisa, de abordagem qualitativa e quantitativa, teve como objetivo, em seu cerne, estabelecer outro sentido para o “erro” na avaliação da aprendizagem de leitura, algo que deve ser abordado e explorado ulteriormente, pois no âmbito da avaliação externa em larga escala não é comum esse tipo de análise dos “erros” para a aferição da proficiência dos alunos, e mesmo nas avaliações internas, conduzidas pelos professores, ainda que aí isso possa ser mais recorrente.

Para desenvolvimento da pesquisa, inicialmente, realizamos um levantamento bibliográfico, seguido de seu tratamento, sobre: as pesquisas sobre TAI; a Teoria de Resposta ao Item (TRI), utilizada no tratamento estatístico, principalmente sobre a TRI que considera as alternativas no

seu modelo matemático; os processos de alfabetização e letramento; a avaliação da aprendizagem de leitura; a elaboração de itens; e a interpretação pedagógica.

A definição do modelo da TRI, a análise estatística para definição do modelo e os testes com simulações foram de suma importância para este trabalho, pois a análise dos “erros” está diretamente relacionada com a definição de um modelo para o tratamento estatísticos dos mesmos. Este foi um aspecto candente da pesquisa, justamente para trazer o sentido pedagógico na análise do “erro” na avaliação, com o intuito de problematizar a trajetória do pensamento do aluno ao responder aos itens, explorando os saberes e conhecimentos que em relação à competência cuja proficiência se torna o objeto que está sendo avaliado, culminando com a apresentação de uma interpretação pedagógica para os resultados obtidos.

Neste sentido, a Dissertação tem como uma de suas principais justificativas, buscando aportar contribuições relevantes para o campo da educação, notadamente para a avaliação educacional e o sucesso escolar, trazer melhorias para o processo avaliativo, o que impacta, mesmo que indiretamente, as decisões tomadas na busca da melhoria do aprendizado e, por consequência, o sucesso escolar. A avaliação, por si só, não produz avanço no aprendizado, mas sem ela o ensino não dispõe de melhores informações.

A Dissertação, nos capítulos que se seguem, apresenta, no primeiro, um histórico da Provinha Brasil e algumas das características desse instrumento, além de uma análise descritiva dos resultados das aplicações da PB na RME-SP entre os anos de 2012 e 2015; no segundo, um panorama sobre as medidas educacionais, privilegiando as principais teorias utilizadas, a TCT e a TRI, ademais as análises das bases de dados da RME-SP por meio da TRI e definição do modelo que será utilizado no TAI-Leitura 3.0; no terceiro, um breve histórico sobre a implementação do TAI-Leitura e as principais característica de um TAI; para o quarto, temos a apresentação das simulações com os algoritmos do TAI-Leitura 3.0 e os resultados do pré-teste em uma das escolas parceiras; no quinto, os resultados da aplicação geral do TAI-Leitura 3.0 nas três escolas que compõem o projeto e o início de uma discussão sobre as possibilidades de interpretação pedagógicas com este novo modelo de TAI; finalizando, na Considerações finais, com uma sistematização dos resultados da pesquisa e indicações de possíveis desdobramentos.

## **1 – Avaliação educacional: a Provinha Brasil como base para o TAI**

A palavra avaliação, sempre que é mencionada no ambiente escolar ou em qualquer argumentação sobre Educação, acaba ficando no centro de uma discussão há muito difundida no meio educacional, principalmente quanto ao significado dessa palavra e seu alcance na prática. Ademais, com o avanço das avaliações externas e em larga escala, esse debate ganhou contornos “separatistas”, no qual temos aqueles que são completamente contrários a esse tipo de avaliação e aqueles que defendem como que “cegamente” seus resultados. Obviamente, essa querela entre ambas as partes têm, como pano de fundo, políticas educacionais, igualmente controversas, que permeiam as iniciativas de avaliação, mas que, a despeito de sua importância, não são abordadas nesta Dissertação.

De qualquer maneira, é importante ressaltar que historicamente essas discussões sobre avaliação não se restringem ao advento das avaliações externas; pelo contrário, existiram diversas modificações e transformações há muito tempo antes no campo da avaliação educacional. Para Guba e Lincoln (2011, p. 13), existiram, pelo menos, três gerações “centradas na mensuração, na descrição e no juízo de valor”<sup>2</sup>. De todo modo, com a ampliação das mais diversas avaliações externas na educação, principalmente, a partir da década de 1990 aqui no Brasil, os instrumentos padronizados começaram a envolver mais amplamente docentes da Educação Básica, ainda que isso não necessariamente tenha sido acompanhado de um entendimento deles, menos ainda com grande adesão, gerando inclusive fortes resistências às avaliações externas em larga escala.

Com base nesse avanço das avaliações em larga escala, o que poderia ser uma oportunidade para se ampliar o debate sobre avaliação, deixou, em nossa opinião, a discussão mais focada nos usos dos resultados de desempenho nesses instrumentos, sem abordar as centralidades que o tema pode nos proporcionar, como, por exemplo, a diferença entre “medir” e “avaliar”, fazendo com que mesmo nos dias atuais existam profissionais da educação que ainda confundem o ato de medir com o de avaliar, considerando como ações sinônimas. Ao mesmo tempo, os atores responsáveis pela implementação das avaliações externas agem, ou pelo menos são coniventes, para com os usos indevidos e contraditórios dos resultados dos desempenhos nos instrumentos.

---

<sup>2</sup> Os autores compreendem existir uma quarta geração de avaliação, focada no construtivismo, mas que não vamos abordar neste trabalho, pois não é o nosso foco.

Para se apreender esse processo histórico, encontramos em Bonamino e Sousa (2012, p. 375) uma abordagem que considera que as avaliações externas no Brasil podem ser classificadas em três gerações que apresentam características diferentes, a saber:

- diagnóstico da qualidade da educação ofertada no Brasil, sem atribuição de consequências diretas para as escolas e para o currículo;
- contemplam, além da divulgação pública, a devolução dos resultados para as escolas, sem estabelecer consequências materiais;
- referenciam políticas de responsabilização forte ou *high stakes*, contemplando sanções ou recompensas em decorrência dos resultados de alunos e escolas;

Em certa monta, as ações explicitadas da terceira geração foram implementadas pelos estados e municípios, vinculando avaliação com políticas de bonificação para os funcionários da educação e repasses de verbas para as escolas. Já o Sistema de Avaliação da Educação Básica (Saeb), principalmente a partir de 2005, quando a aplicação do instrumento passou a ser censitária para os alunos do 5º e 9º ano do Ensino Fundamental e 3ª série do Ensino Médio, se enquadra na segunda geração, pois existe a divulgação dos resultados, mas sem vinculação com bônus ou punições.

Entretanto, na perspectiva concreta de que as avaliações internas e externas existem no dia a dia das escolas, das professoras e dos professores e de alunas e alunos no Brasil, é importante termos uma posição crítica, porém sem desconsiderar a conjuntura atual, pois isso seria uma tentativa de encaixar o que pensamos na realidade concreta ao invés de extrairmos dela. Dessa maneira, estabelecer um diálogo profícuo entre ambas as avaliações nos parece mais adequado e, principalmente, mais proveitoso para o aprendizado dos estudantes. Para Nevo (1997, p. 96), a avaliação deve ter a escola como centro, da mesma forma

[é] preciso que as escolas tenham a avaliação em mente, se a intenção for fazer com que a avaliação tenha qualquer papel significativo para melhorar a educação. Mas as escolas são formadas por alunos, professores e por administradores, e somente eles poderão fazer dela uma escola que tenha a avaliação em mente. Alunos, professores e administradores escolares devem transformar-se em participantes ativos dos diálogos de avaliação, em vez de continuarem sendo recipientes passivos das descrições e dos julgamentos oriundos das avaliações.

Na perspectiva do autor, é possível que exista um mundo em que as avaliações internas e externas convivam em certa harmonia, porém, para tanto, é necessário que a escola se torne o centro da ação, de modo que ela não seja somente uma mera “utilizadora” de resultados e

indicadores. Dessa maneira, compreendemos que o projeto desenvolvido no bojo do Grupo de Estudo e Pesquisa em Avaliação Educacional (Gepave), vinculado à Faculdade de Educação da Universidade de São Paulo, em que os aspectos de formação em avaliação educacional, das características dos instrumentos de coleta de dados para a avaliação, além da própria aplicação do TAI e discussão dos seus resultados, podem contribuir com a implementação desse diálogo. Dessa forma, este trabalho busca apresentar um aspecto desse diálogo, a contribuição que o TAI pode trazer para a avaliação interna e, ao mesmo tempo, externa. Destarte, a base metodológica que do trabalho de Catalani (2019) nos proporcionou avançamos com as análises mais internas ao TAI, na busca de melhorar ou, pelo menos, ampliar a compreensão do desempenho dos estudantes nos testes. A seguir, vamos apresentar a estrutura, algumas características e um breve histórico da PB.

### **1.1 – Provinha Brasil: estrutura, características e um breve histórico**

A Provinha Brasil (PB) foi instituída por meio de portaria normativa Brasil (2007), com os objetivos de:

- a) avaliar o nível de alfabetização dos educandos nos anos iniciais do ensino fundamental;
- b) oferecer às redes de ensino um resultado da qualidade do ensino, prevenindo o diagnóstico tardio das dificuldades de aprendizagem; e
- c) concorrer para a melhoria da qualidade de ensino e redução das desigualdades, em consonância com as metas e políticas estabelecidas pelas diretrizes da educação nacional.

Esses objetivos foram estabelecidos para serem alcançados no 2º ano do Ensino Fundamental e para viabilizá-los a PB era distribuída às redes de ensino já impressa para aplicação no meio e no final do ano letivo. Essa metodologia proporcionava uma comparação entre os resultados de ambas as aplicações, o que permitia verificar se avanços ocorreram em relação ao desempenho dos alunos durante o ano letivo. Em cada uma dessas aplicações buscava-se medir a proficiência em Língua Portuguesa e Matemática, com foco, respectivamente, em Leitura/Letramento e Resolução de Problemas.

Os instrumentos da PB, entre os anos de 2008 a 2010, em ambas as aplicações, eram compostos por 24 itens<sup>3</sup>, que se dividiam entre itens lidos totalmente pelas professoras e pelos professores,

---

<sup>3</sup> Item (“questão”) consiste na unidade elementar de um instrumento de coleta de dados.

aqueles que eram parcialmente lidos e os que eram autônomos, com leitura completamente individual. A partir de 2011 houve uma redução nessa quantidade de questões, e os instrumentos passaram a ser compostos por 20 itens. A decisão para a redução, segundo o Inep (2011, p. 5), foi tomada “[c]om o objetivo de atender as sugestões de diminuir a extensão da prova, por ser cansativa, houve redução do número de questões”. Esse quantitativo se manteve até a PB ser descontinuada após a disponibilização do Teste 2 de 2016.

Para além da prova impressa, as escolas recebiam outros materiais, que compunham o “Kit da Provinha Brasil” e que foram distribuídos nacionalmente entre 2008 e o primeiro teste de 2016, sendo que no Teste 2 de 2016 o kit foi disponibilizado somente no site do Inep. Esse material continha um “Guia de aplicação”, no qual os itens que compunham a prova que deveria ser aplicada, incluindo instruções de quais questões, deveriam ser lidas. Ademais, tinha o “Guia de correção”, em que informações sobre a estrutura da prova, a tabela com as quantidades de acertos e os níveis e a interpretação pedagógica de cada nível e outro documento contendo “Reflexões Sobre a Prática”, com considerações sobre o processo de alfabetização. Os dois primeiros estiveram presentes em todos os testes, e o documento de reflexões foi descontinuado após o Teste 2 de 2011. A seguir, vamos apresentar mais detalhes de cada um desses materiais.

O “Guia de aplicação” era um documento que trazia informações sobre os objetivos das aplicações da PB, incluindo formulações mais gerais, como por exemplo, em Inep (2011, p. 5), “subsídios para formulação de políticas de alfabetização”. Esse guia também trazia os alunos matriculados no segundo ano do Ensino Fundamental como público-alvo da PB, as indicações de quais itens deveriam ser lidos parcialmente, completamente e individualmente por cada estudante. A Figura 1, a título de exemplo desses materiais, nos apresenta um dos itens que estava presente no “Guia de aplicação” do Teste 1 de 2013.

**Figura 1** – Provinha Brasil. Questão 5. Teste 1. 2013

 Marque um X no quadradinho da palavra CAMISETA.



- (A)  CAMISOLA
- (B)  CANETA
- (C)  CAMISETA
- (D)  CAMIONETA

*Comentário:* A questão avalia a habilidade de ler palavras associando imagem e fonemas à sua representação gráfica. As palavras apresentadas preservam a mesma sílaba inicial, com semelhança fonética e são de estrutura silábica trissílaba e polissílaba, além de significados diferentes, o que agrega certa complexidade para o aluno em processo de alfabetização. O aluno que opta pelo gabarito C demonstra estabelecer a associação entre significante e significado da palavra e provavelmente domínio de estrutura silábica. No caso de o aluno que opta pelas alternativas A e D, possivelmente lê apenas parte da palavra, uma vez que as palavras apresentadas nessas alternativas têm a mesma estrutura das duas primeiras sílabas do gabarito. O aluno que opta pela alternativa B pode ainda não dominar a leitura de palavras de maior número de sílabas (trissílabas e polissílabas).

**Fonte:** Inep (2013, p. 11).

O item que aparece na Figura 1 foi elaborado com base na habilidade de ler palavras associando imagem e fonemas à sua representação gráfica. Essa informação é apresentada no “Comentário” do item no “Guia de aplicação” e na matriz de avaliação no “Guia de correção”. Além da habilidade, temos o próprio item em si, em que podemos identificar todas as partes que o compõem<sup>4</sup>, como texto base – é a parte que traz o contexto do item, mas precisa ser realmente necessária para a realização da tarefa. Pode conter texto de diversos gêneros e imagens

<sup>4</sup> A Provinha Brasil contava com o documento “Guia de elaboração de itens: Provinha Brasil”, que referenciava a elaboração de itens. Disponível em: [Miolo Guia Elaboração de Itens Provinha Brasil.indd \(inep.gov.br\)](http://miolo.guia-elaboracao-de-itens-provinha-brasil.inep.gov.br). Acesso em: 24 abr. 2023.

(desenhos); enunciado – frase ou texto curto que apresenta a tarefa ou problema, com base na habilidade do item, que deve ser realizada por quem está respondendo ao item. Não pode trazer algo novo, sendo assim, toda informação necessária ao contexto do item deve estar no texto-base. É necessário que a pergunta não seja dúbia, pois isso pode gerar duplo gabarito ou, até mesmo, a falta dele; alternativas – se dividem em correta e incorretas<sup>5</sup>, em que a considerada correta precisa responder ou completar o que foi anunciado no enunciado. As que fazem parte das alternativas incorretas, necessitam apresentar uma plausibilidade, que segundo o Inep (2012, p. 37),

[S]ua elaboração não pode ocorrer de forma aleatória, mas devem ser formulados de modo que reflitam uma possibilidade de resolução ou um raciocínio possível para aquelas crianças que ainda não desenvolveram determinada habilidade. A utilização de erros comuns observados em situação de ensino-aprendizagem costuma aumentar a plausibilidade dos distratores.

Essa plausibilidade presente nos itens é utilizada como algo central para esse trabalho, pois é com base nela que vislumbramos a possibilidade de avançar nas pesquisas envolvendo o TAI-Leitura<sup>6</sup>, com o intuito de aproximar esse tipo de teste ao ambiente escolar, para, principalmente, ampliar as possibilidades de utilização pedagógica do item de múltipla escolha. É importante destacar que as versões dos itens presentes na prova entregue aos alunos e no “Guia de aplicação”, apresentavam diferenças importantes para o processo de coleta dos dados, pois na versão dos alunos era possível que não existisse o enunciado, somente o texto ou a imagem do texto-base. Já no guia, essas partes que precisavam ser lidas, tinham uma indicação prévia de um desenho de um megafone, assim, nas partes que o megafone aparecia, as professoras e os professores deveriam ler para todas as alunas e todos os alunos.

Antes de avançarmos para o “Guia de correção”, queremos salientar a importância que itens bem elaborados têm para qualquer instrumento, principalmente para um instrumento que busca analisar como cada alternativa pode estar associada ao desempenho de cada aluno, sobretudo para estimar sua proficiência com a realização do teste ao responder vários itens. Dessa maneira, adotarmos um modelo de análise estatística coerente com o nosso propósito é fundamental<sup>7</sup>. Ao

---

<sup>5</sup> É comum essas alternativas incorretas serem intituladas de “distratores”, porém, neste trabalho, compreendemos ser uma nomeação errônea, pois indica haver algo que distraia o respondente, inclusive como um ato deliberado do elaborador do item, quando não é isso que de fato essas possíveis respostas representam.

<sup>6</sup> Elaborado, inicialmente, em Catalani (2019).

<sup>7</sup> Iremos discutir sobre modelos estatísticos no capítulo 2.



mesmo tempo, a escolha pelos itens de Língua Portuguesa da PB, nos proporciona um limiar entre etapas da alfabetização e a própria competência leitora em si, pois existem itens que estão voltados para o Sistema de Escrita e outros que demandam uma compreensão de estruturas do texto. Nas Figuras 2 e 3, temos dois itens que compõem o Banco de Itens do TAI-Leitura 3.0 e que são apresentados para todas as alunas e todos os alunos que respondem ao TAI.

**Figura 2** – Provinha Brasil. Questão 4. Teste 2. 2016

 **Veja a figura.**



Disponível em: [www.smartkids.com.br](http://www.smartkids.com.br). Acesso em: 23 jan. 2015.

 **Marque um X no quadradinho em que aparece a palavra BALÃO.**

- (A) PILÃO
- (B) GALÃO
- (C) VILÃO
- (D) BALÃO

**Comentário:** Este item avalia a habilidade de reconhecer o valor sonoro de uma sílaba a partir de uma imagem, neste caso variando nas alternativas apenas a sílaba inicial das palavras. Para responder corretamente a questão, o(a) estudante precisará identificar a sílaba BA como a sílaba inicial do nome do objeto apresentado na imagem, alternativa (D). A opção pelas alternativas (A), (B) ou (C) pode indicar que o(a) estudante ainda não identifica o valor sonoro das letras B e A.

**Fonte:** Inep (2016, p. 10).

**Figura 3** – Provinha Brasil. Questão 11. Teste 2. 2015

 **Leia o texto silenciosamente. Depois que todos terminarem, eu vou dizer o que é para fazer.**


**QUER BRINCAR DE PIQUE-ESCONDE?**

O MACACO BRINCALHÃO,  
FAZENDO PALHAÇADA,  
QUER BRINCAR DE PIQUE-ESCONDE  
COM TODA BICHARADA.

O COELHO BEM LIGEIRO  
FOI PARA TRÁS DE UM COQUEIRO.  
MAS E AGORA?  
DEIXOU AS ORELHAS DE FORA!

A GIRafa ENCOLHIDINHA  
ATRÁS DO GALHO FICOU QUIETINHA.  
MAS E AGORA?  
DEIXOU O PESCOÇO DE FORA.

ARPANEDA, I.; BRAGANÇA, A. *Quer brincar de pique-esconde?* São Paulo: FTD, 2007 (fragmento).

 **Marque um X no quadradinho que indica o assunto desse texto.**

- (A) UM MACACO NA MATA.
- (B) UMA BRIGA DA BICHARADA.
- (C) UM COELHO NO COQUEIRO.
- (D) UMA BRINCADEIRA DE ESCONDER.

*Comentário: Este item avalia a habilidade de reconhecer o assunto de um texto com base em seu título. Ao assinalar a alternativa (D), o aluno demonstrou domínio da habilidade. Ao assinalar as demais alternativas, o aluno pode ter considerado expressões pontuais presentes no texto, mas que não correspondem ao assunto enunciado pelo título.*

**Fonte:** Inep (2015, p. 18).

O item da Figura 2 é apresentado como um item de familiarização dos alunos com os sistemas, dessa maneira, eles podem compreender o que devem fazer para conseguir escutar novamente a leitura do enunciado, além da parte que avisa o que deve ser feito antes do texto-base, como por exemplo, “veja a figura”. Esse item, assim como na PB, não conta para o escore da prova, sendo assim, as respostas proferidas pelos alunos não vão influenciar a medida da sua proficiência.

Na Figura 3 temos o primeiro item do TAI-Leitura 3.0, respondido por todos os alunos, escolhido porque a resposta a ele estabelece uma priori de proficiência para começar o teste, algo que será explicado no capítulo 3. Assim, dependendo de qual alternativa for escolhida, um novo item diferente será apresentado e, como pontuamos anteriormente, essa diferenciação com o próximo item tem muito a ver com possibilidades pedagógicas, pois, se um aluno, ao se deparar com o item da Figura 3 e apresentar dificuldades para realizar, de maneira autônoma, a leitura do fragmento de texto, pode escolher a alternativa (A) – UM MACACO NA MATA. Com base no que o comentário do item nos apresenta no Inep (2015, p. 18), “o aluno pode ter considerado as expressões pontuais presente no texto, mas que não correspondem ao assunto enunciado pelo título”. Em outras palavras, o aluno não soube reconhecer qual era o assunto do texto, porém, ele consegue localizar informação explícita no texto, mesmo que não tenha relação direta com a tarefa solicitada.

Desse jeito, não faz sentido que esse aluno que escolhe MACACO como resposta receba, como próximo item, o mesmo que o aluno que marcou a alternativa (C) – UM COELHO NO COQUEIRO, pois muito provavelmente não conseguiu avançar na leitura do texto-base com entendimento de seu conteúdo, identificando um trecho logo na segunda palavra após o título, onde aparece a palavra MACACO. Ademais, aqueles que escolhem a alternativa (C) precisam “resistir” à obviedade das alternativas (A) e (B), pois ambas aparecem anteriormente no texto, sendo assim, caso o aluno não soubesse reconhecer o assunto do texto de maneira correta, ele compreende que palavras sozinhas não conseguem dar o sentido para termos o assunto do texto. Todavia, é importante destacar que esse exemplo e essa reflexão estão no bojo da plausibilidade, dessa forma, podemos encontrar outros caminhos pedagógicos para compreender cada uma das escolhas neste item, contudo, temos clareza que essas possibilidades nos mostram caminhos para intervenções pedagógicas, mas que fogem da responsabilidade da avaliação.

Retomando o material disponibilizado pelo Inep, o “Guia de correção” da PB traz uma apresentação de boas-vindas para as professoras e os professores, com uma explicação breve sobre o material que compunha o “kit da Provinha Brasil”, além de explicar, em poucas palavras, o que havia no documento “Reflexões Sobre a Prática”<sup>8</sup> esse por sua vez, foi

---

<sup>8</sup> Esses documentos apresentavam reflexões desde realizar corretamente a interpretação dos resultados de desempenho dos alunos, como também, realizar a divulgação desses resultados, principalmente, para os próprios alunos. Relacionar esses resultados com: a formação continuada de alfabetizadores; o uso e escolha dos materiais didáticos; o próprio livro didático; os dicionários; os livros da literatura infantil; os programas de formação continuada do MEC, a Rede Nacional de Formação e o Pró-Letramento.

descontinuado em 2011. Além de uma abrangente explicação sobre como deveria ser feita a correção (apontamento das alternativas corretas) do instrumento. O guia também trazia algumas informações sobre o processo de alfabetização e letramento, Inep (2009, p. 7), indicando que são dois processos complementares e inseparáveis. Além dessas informações, o guia apresenta uma explicação de como poderia ser feita a interpretação pedagógica do desempenho dos alunos, para tanto, era indicado a existência de uma tabela, em cada teste, como encontra-se na Figura 4, que relacionava quantidade de acerto e Nível.

**Figura 4** – Provinha Brasil. Quantidade de acertos e níveis de desempenho. 2016

#### **INTERPRETAÇÃO DOS NÍVEIS DE DESEMPENHO – LEITURA**

No teste 1 da Provinha Brasil de Leitura de 2016, aplicado no primeiro semestre, os números de acertos adotados para a identificação dos níveis de desempenho dos(as) estudantes foram:

<b>Teste 1 – 2016</b>
Nível 1 – até 3 acertos
Nível 2 – de 4 a 9 acertos
Nível 3 – de 10 a 14 acertos
Nível 4 – de 15 a 16 acertos
Nível 5 – de 17 a 20 acertos

Neste Teste 2 da Provinha Brasil de 2016, são adotados os seguintes números de acertos para identificar os níveis de desempenho dos(as) estudantes:

<b>Teste 2 – 2016</b>
Nível 1 – até 2 acertos
Nível 2 – de 3 a 7 acertos
Nível 3 – de 8 a 13 acertos
Nível 4 – de 14 a 15 acertos
Nível 5 – de 16 a 20 acertos

**Fonte:** Inep (2016, p. 24).

Desde o primeiro teste a PB já continha cinco níveis para contribuir com a sumarização dos resultados e a interpretação pedagógica deles. Esses níveis de desempenho foram definidos, segundo o Inep (2016, p. 23), “após a aplicação de cada Pré-teste, é feita uma análise estatística e pedagógica minuciosa de cada questão e elas são ordenadas da mais fácil até a mais difícil. Em seguida, essas mesmas questões são agregadas em níveis, de acordo com seu grau de dificuldade”. Dessa maneira, os intervalos de acertos que temos na Figura 4, variam, para além

do Teste 1 e Teste 2 de cada ano civil, entre as aplicações da PB ao longo dos anos civis, pois mesmo mantendo uma escala comparável ao longo do tempo, a grande maioria dos itens eram diferentes entre os pré-testes; logo, era comum existirem alterações na quantidade de acertos em cada Nível. A Figura 5 apresenta alguns exemplos de como são as interpretações pedagógicas de cada Nível.

**Figura 5** – Provinha Brasil. Interpretação pedagógica dos Níveis 1, 2 e 3. Teste 2. 2016

<b>Nível 1</b>
<b>até 2 acertos</b>
<p>Neste nível, os(as) estudantes geralmente já podem:</p> <ul style="list-style-type: none"> <li>• diferenciar letras de outros sinais gráficos;</li> <li>• identificar letra ou sequência de letras do alfabeto lida pelo(a) aplicador(a).</li> </ul>
<b>Nível 2</b>
<b>de 3 a 7 acertos</b>
<p>Os(as) estudantes que se encontram neste nível, além de, provavelmente, já terem consolidado as habilidades do nível anterior, geralmente já podem:</p> <ul style="list-style-type: none"> <li>• reconhecer palavras de formação silábica canônica escritas de diferentes formas;</li> <li>• estabelecer relação entre grafemas e fonemas, identificando, por exemplo, a letra ou a sílaba inicial de uma palavra;</li> <li>• ler palavras formadas por sílabas canônicas e não canônicas.</li> </ul>
<b>Nível 3</b>
<b>de 8 a 13 acertos</b>
<p>Os(as) estudantes que se encontram neste nível, além de, provavelmente, já terem consolidado as habilidades dos níveis anteriores, geralmente já podem:</p> <ul style="list-style-type: none"> <li>• identificar o número de sílabas em uma palavra;</li> <li>• ler frases de sintaxe simples com o apoio de imagens ou ditadas pelo(a) aplicador(a);</li> <li>• identificar informação explícita de fácil localização em textos curtos com o apoio da leitura pelo(a) aplicador(a) ou pela leitura individual;</li> <li>• inferir informações em textos curtos de gêneros usuais, pela leitura individual e com o apoio em linguagem não verbal;</li> <li>• reconhecer o assunto do texto com o apoio do título ou de conteúdo informacional trivial, com base nas características gráficas do gênero, pela leitura individual ou com o auxílio da leitura pelo(a) aplicador(a);</li> <li>• reconhecer a finalidade de textos de gêneros usuais (receita, bilhete, curiosidades, cartaz) com base nas características gráficas desses e na leitura individual.</li> </ul>

**Fonte:** Adaptado de Inep (2016, p. 25-27).

O texto que compõe cada uma das partes da interpretação pedagógica pode ser intitulado como “tarefa cognitiva” ou “sentença descritora”, ambas se fundamentam na mesma relação entre:

parte da habilidade do item (verbo da taxonomia e objeto do conhecimento) e o contexto no qual o item foi elaborado. Assim, essa interpretação é realizada com base no conteúdo do item, mas sem descrevê-lo por completo e, como já apresentado, o item é posicionado<sup>9</sup> e a partir desse posicionamento, considerando um conjunto de itens posicionados no intervalo do nível, e produzida a interpretação pedagógica do nível. Porém, sem buscar uma descrição dos acertos e erros dos alunos no teste. De certa maneira, esse trabalho procura problematizar essa forma de interpretar, pois com a utilização do TAI-Leitura 3.0, a individualização do processo passa a ser vislumbrada, como iremos salientar no capítulo 5.

Antes do próximo subcapítulo, é conveniente pontuar que, a partir do Teste 1 de 2011, a matriz de avaliação de ambos os componentes fora incorporada ao “Guia de correção” e, assim, era possível ter em um mesmo documento as informações relacionadas à elaboração dos itens e a interpretação pedagógica do resultado dos alunos. A seguir, vamos apresentar os resultados dos desempenhos dos alunos da Rede Municipal de Ensino de São Paulo (RME-SP) no período entre 2012 e 2015.

## **1.2 – Os resultados da Provinha Brasil – Leitura na Rede Municipal de Educação de São Paulo (2012 a 2015)**

Neste subcapítulo o objetivo é apresentar os aspectos que proporcionaram adquirir as bases de dados da RME-SP, incluindo os motivos que nos fizeram ter essa escolha. Discutir os resultados das aplicações de 2012 a 2015, com base nos gráficos dos níveis e das marcações nas alternativas. A ideia é demonstrar que existe algo interessante na análise das alternativas.

A RME-SP aderiu à PB desde sua primeira aplicação em 2008 e somente no Teste 2 de 2016 é que não houve participação da rede, pois as provas não foram oferecidas de forma impressa, o que inviabilizou, segundo consta na resposta ao pedido de acesso à informação que fizemos à RME-SP, realizar a distribuição do “kit da Provinha Brasil”. Ao mesmo tempo, segundo relatos de colegas que atuam na RME-SP desde antes da implementação da PB, a rede disponibilizava mecanismos para coletar os dados gerados com cada aplicação da PB, o que permitiria termos a possibilidade de acessar milhares de vetores de respostas aos itens, o que facilitaria a implementação do TAI-Leitura 3.0.

---

<sup>9</sup> Esse processo de posicionamento do item é conhecido como “ancoragem” e é realizado com base em uma metodologia estatística. Para mais informações Andrade et al. (2000).

Para que isso pudesse ocorrer, realizamos um pedido via Lei de Acesso à Informação<sup>10</sup> em dezembro de 2020, que ficou sem resposta por praticamente um ano, sendo respondido em setembro de 2021. De qualquer maneira, por conta dos sistemas que a RME-SP utilizou para coletar os dados terem sido descontinuados, dessa maneira, não tivemos acesso à totalidade dos dados, mas a uma parte que englobou as aplicações de 2012 (Teste 2) até 2015 (testes 1 e 2), com exceção ao Teste 2 de 2014 que não estava nos dados disponibilizados. Assim, com base nessas informações que tivemos acesso, elaboramos uma análise descritiva da distribuição dos alunos da RME-SP nos níveis e, de maneira ainda embrionária, começamos a verificar a relação entre escolha das alternativas e classificação no nível. Dessa maneira, é possível observar na Tabela 1 a quantidade de alunos que responderam cada teste.

**Tabela 1** – Provinha Brasil. Quantidade de alunos em cada aplicação na RME-SP. 2012-2015

Ano	Teste	
	1	2
2012	-	44368
2013	42415	44391
2014	-	46476
2015	43523	45780
Total	85938	181015

**Fonte:** Elaboração própria.

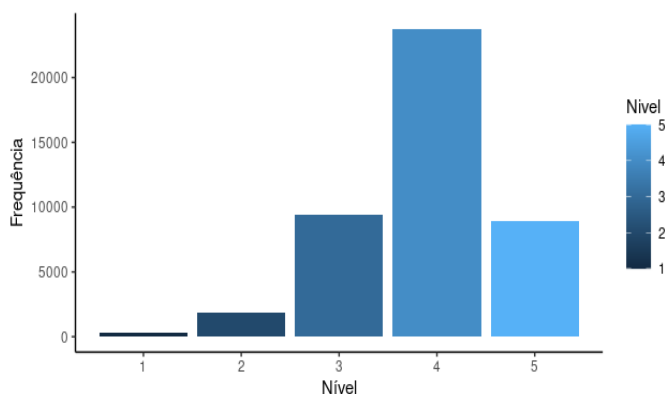
É possível perceber, com base nos dados da Tabela 1, que temos uma grande quantidade de alunos que participaram das aplicações da PB no período analisado, mesmo sem realizar nenhum tipo de análise de frequência relativa ou similar. De qualquer maneira, é importante destacarmos que ter ou não uma maior representatividade do desempenho dos alunos do 2º ano do Ensino Fundamental da RME-SP, não tem grande relevância para o trabalho realizado neste momento, pois com esse quantitativo conseguimos realizar, de maneira significativa, os processamentos estatísticos para obtermos as informações sobre as alternativas.

Sendo assim, não aprofundamos as análises em relação à participação dos alunos na PB aplicada na RME-SP. Outrossim, com esse quantitativo conseguimos analisar a distribuição dos alunos nos níveis e como as respostas ocorreram por item. No Gráfico 1 temos a representação da

<sup>10</sup> Lei nº 12527, de 18 de novembro de 2011. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/12527.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/12527.htm)

distribuição dos alunos nos níveis da PB de Língua Portuguesa (foco em Leitura)<sup>11</sup> no Teste 2 de 2012.

**Gráfico 1** – Provinha Brasil. Distribuição do desempenho dos alunos da RME-SP por níveis. Teste 2. 2012



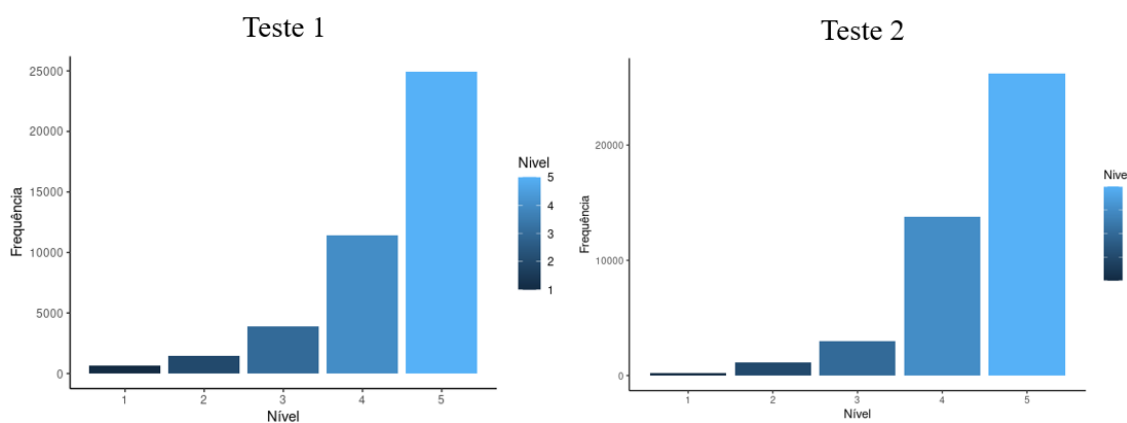
**Fonte:** Elaboração própria.

O Teste 2 é aplicado ao final do 2º ano, logo, é esperado que haja uma maior concentração de desempenho dos alunos nos Níveis 4 e 5. Dessa maneira, o que podemos analisar no Gráfico 1 é uma concentração maior de desempenho dos alunos no Nível 4, além de termos uma quantidade próxima dos trinta mil nos Níveis 4 e 5, algo que corrobora com a expectativa que esse segundo teste gera. Todavia, é preciso perceber que mais de dez mil alunos apresentaram desempenho abaixo desses níveis esperados, o que demonstra uma grande necessidade de se avançar nas abordagens pedagógicas voltadas para a alfabetização no 3º ano do Ensino Fundamental. A Figura 6 apresenta os gráficos de ambos os testes de 2013.

<sup>11</sup> Os dados analisados em todos os capítulos deste trabalho são referentes ao construto de Leitura.



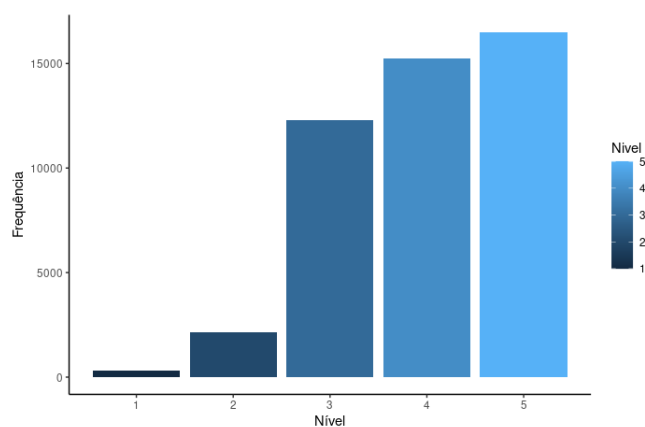
**Figura 6** – Distribuição dos desempenhos dos alunos da RME-SP nos testes 1 e 2. 2013



**Fonte:** Elaboração própria.

Diferente da distribuição que observamos no Gráfico 1, temos uma concentração maior de desempenho dos alunos no Nível 5 em ambos os testes, o que demonstra uma grande diferença nos resultados do Teste 2 de 2012. Além disso, os dois gráficos na Figura 6 apresentam dados de praticamente os mesmos alunos, pois os dois testes foram aplicados no mesmo ano civil, sendo assim, podemos identificar uma diminuição na quantidade de desempenho dos alunos nos Níveis 1, 2 e 3 e um aumento, principalmente, no Nível 4. Isso evidencia o quanto os instrumentos da PB eram capazes de captar os avanços no aprendizado dos alunos. O Gráfico 2 apresenta os dados do Teste 2 de 2014.

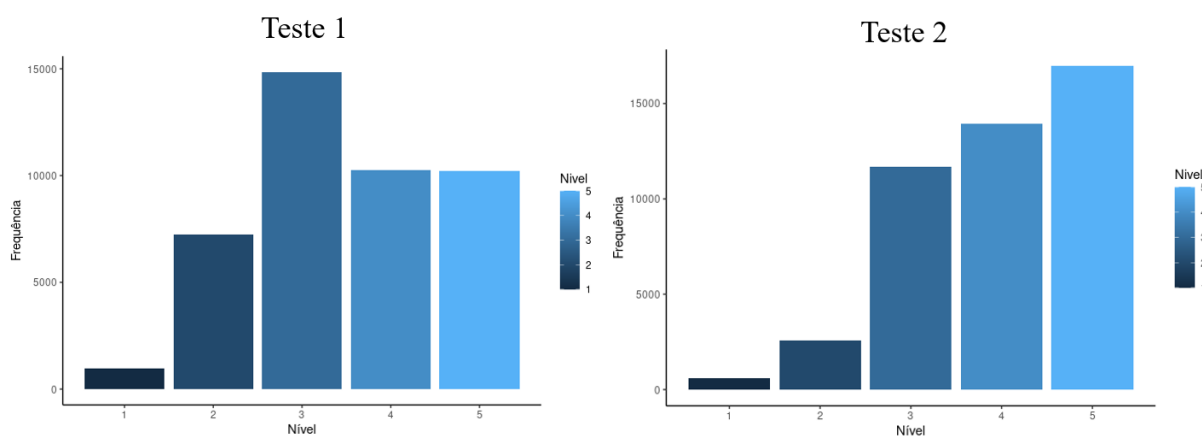
**Gráfico 2** – Distribuição dos desempenhos dos alunos da RME-SP no Teste 2 de 2014



**Fonte:** Elaboração própria.

A aplicação do Teste 2 de 2014 na RME-SP obteve dados de desempenho discrepante dos resultados nos dois testes 2 dos anos anteriores, como é possível verificar no Gráfico 2. Podemos analisar que houve um aumento de desempenho nos Níveis 3 e 4, além de uma diminuição na quantidade de alunos com os seus desempenhos classificados em 5. Com base nessas informações, podemos supor que os alunos que estavam no 1º ano do Ensino Fundamental em 2013 e tiveram um ano de escolarização no 2º ano na RME-SP, apresentaram mais dificuldades no processo de alfabetização que os seus colegas do 3º e 4º ano de 2014. Todavia, sem que essa análise seja determinista, a maior concentração nos três maiores níveis demonstra que esses alunos vão ter uma maior compreensão dos conteúdos abordados no 3º ano. A Figura 7 apresenta a distribuição de desempenho dos alunos nos dois testes de 2015.

**Figura 7** – Distribuição dos desempenhos dos alunos da RME-SP. Testes 1 e 2. 2015

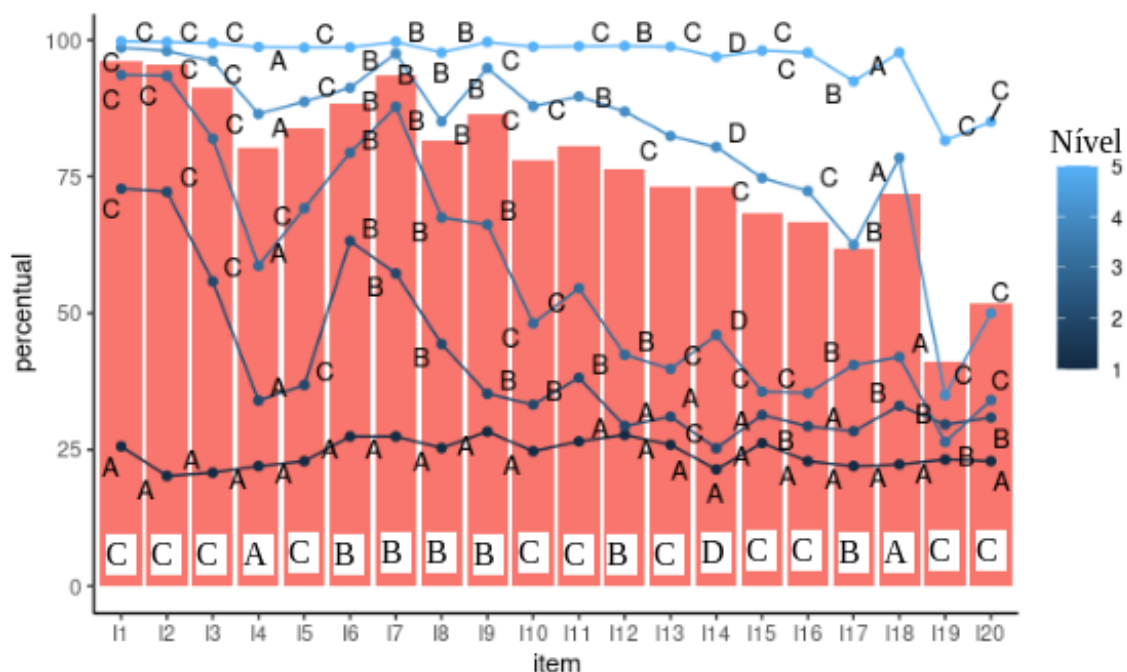


**Fonte:** Elaboração própria.

Na Figura 7 temos uma síntese do que é esperado encontrar nos dados da aplicação dos dois testes da PB, pois temos uma modificação significativa entre as duas aplicações. No Teste 1 temos uma maior concentração dos alunos no Nível 3, além de uma quantidade maior no Nível 2, algo que ainda não havíamos observado. No Teste 2 ocorre um deslocamento na quantidade de desempenho dos alunos para os níveis 4 e 5, com um nítido aumento da proficiência desses alunos. Todavia, a quantidade de alunos no Nível 1 nos chama a atenção, pois mesmo com uma redução em torno de 50%, ainda temos em torno de 1.500 a 2.000 alunos com desempenho neste nível, o que significa avançarem para o 3º ano conseguindo identificar algumas letras e com dificuldades para diferenciar letras de outros símbolos, algo realmente preocupante.

De modo geral, temos um conjunto de dados que possibilita realizarmos processamentos estatísticos das mais variadas formas, principalmente, por dois motivos: o volume de dados é expressivo e temos uma grande variabilidade. Dessa forma, com base no objetivo de nossa pesquisa e com esses dados, foi possível realizar análises das alternativas e acrescentarmos essas informações no TAI-Leitura. Para isso, iniciamos, ainda de maneira embrionária, estudos de como as alternativas dos itens se comportavam em cada um dos níveis, pois gostaríamos de verificar se existia uma distribuição de escolhas das alternativas ao longo de cada teste. Para tanto, utilizamos os dados dos testes 2 de cada ano civil a que tivemos acesso, para não transformar as informações demasiadamente longas e repetitivas. No Gráfico 3 temos o percentual de escolha da alternativa mais assinalada em cada um dos níveis (gráfico de linhas) em conjunto com o percentual de acertos de cada item (gráfico de barras), além da indicação de quais foram as alternativas mais escolhidas em cada nível e o próprio gabarito do item.

**Gráfico 3** – Distribuição do percentual das alternativas mais escolhidas por nível e o percentual de acerto por item. Teste 2. 2012



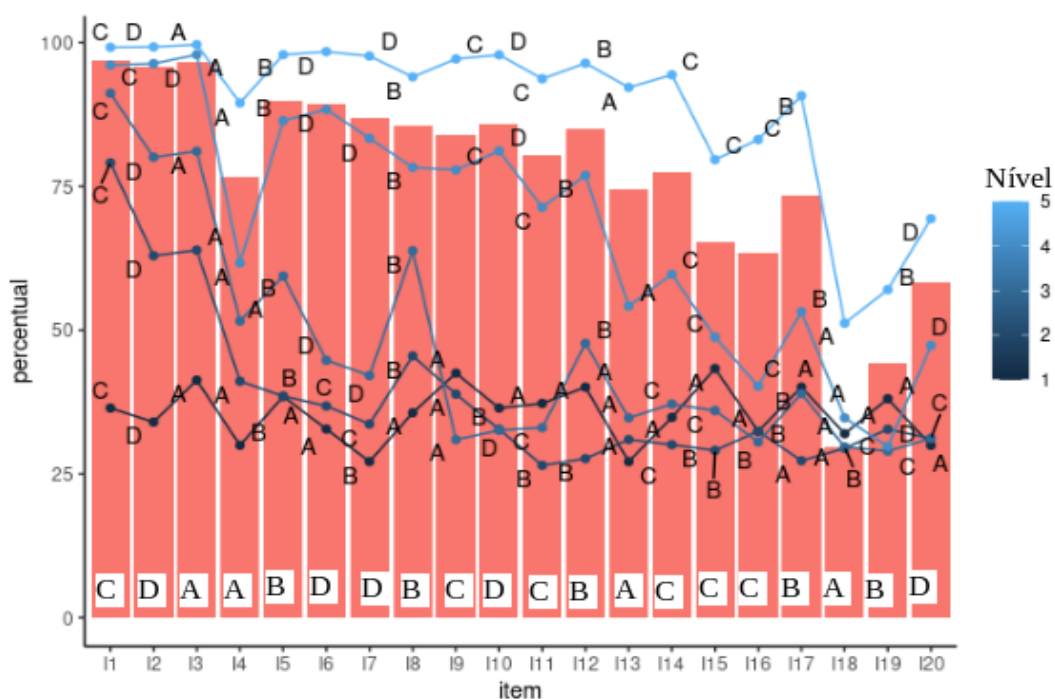
Fonte: Elaboração própria.

As informações que o Gráfico 3 apresentam nos permitem analisar como ocorre a relação entre escolhas das alternativas, levando em consideração aquela que foi mais assinalada em cada

nível e o percentual de acerto em cada item. Dessa maneira, podemos observar que no geral os Níveis 1 e 5 são aqueles que mais se diferenciam em relação à escolha que predomina no grupo, pois temos praticamente a alternativa (A) como a mais marcada em todos os itens no Nível 1. Ao mesmo tempo, no outro extremo, no Nível 5 temos o gabarito sendo escolhido pela maioria dos alunos em todos os itens. As oscilações ficam nos outros níveis, que acabam tendo alternativas corretas ou incorretas sendo escolhidas.

Podemos perceber que existe uma diminuição dos percentuais de escolhas nas alternativas mais marcadas nos Níveis 2, 3 e 4, o que indica uma dispersão maior entre as demais possibilidades de escolhas. Assim, é possível levantarmos a hipótese de que, mesmo com a evidente diferença de escolhas das alternativas internamente aos níveis, os alunos acabam tendo desempenho que os classificam nos respectivos níveis. Isso nos leva a uma segunda hipótese: em um TAI-Leitura que utilize as análises das alternativas, podemos ter diversos “caminhos de escolhas” em cada nível; logo, algo que nos possibilita interpretações pedagógicas diferentes para cada um destes caminhos. No Gráfico 4 temos a distribuição dos percentuais de escolhas no Teste 2 de 2013.

**Gráfico 4** – Distribuição do percentual das alternativas mais escolhidas por nível e o percentual de acerto por item no Teste 2. 2013

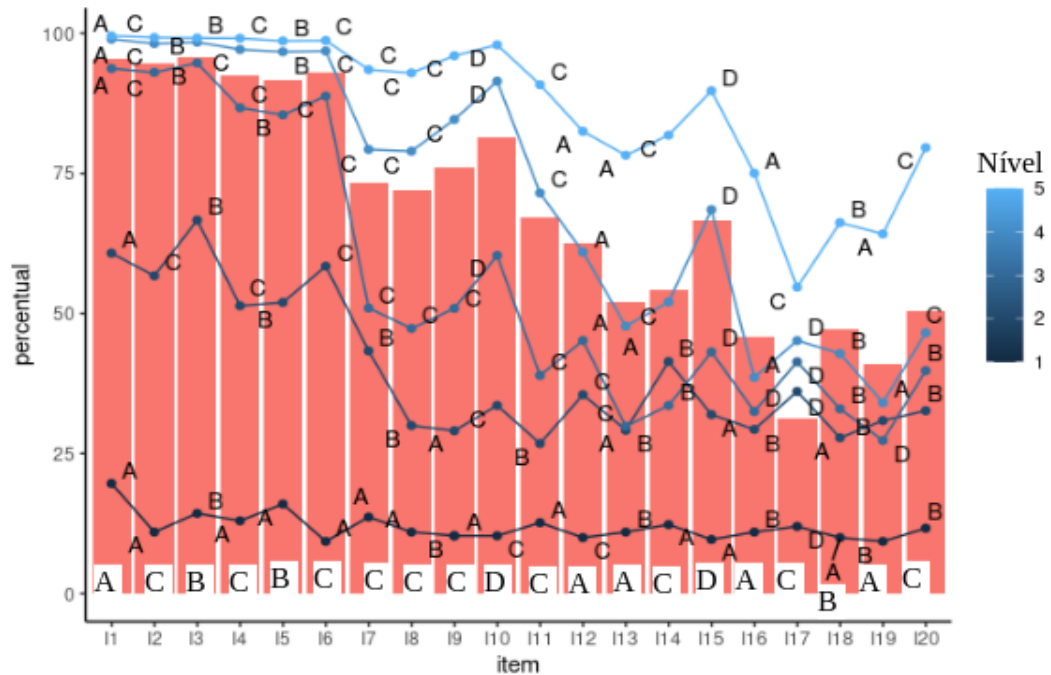


Fonte: Elaboração própria.

O Gráfico 4 nos apresenta informações um pouco diferentes das que analisamos no Gráfico 3, principalmente quanto ao comportamento das escolhas no Nível 1, pois deixa de ter, praticamente, uma única alternativa sendo escolhida, para uma grande quantidade de escolhas consideradas gabarito nos itens e em vários momentos, tendo um percentual de marcação da alternativa correta maior que os Níveis 2, 3 e 4. Para os dados de desempenho do Nível 5 temos percentuais de escolhas que ficam abaixo de 75%, algo que não ocorreu no Teste 2 de 2012, o que pode demonstrar que os itens 18, 19 e 20 foram mais “difíceis”. Contudo, teríamos que verificar tal hipótese com base nas proficiências estimadas. De toda maneira, se utilizarmos o nível como esse alicerce para esta análise, podemos, sim, supor que esses itens exigiram maior proficiência dos alunos que estavam neste nível, principalmente, em comparação com os demais itens.

Os Níveis 2 e 3 apresentam uma proximidade maior nos percentuais de escolhas das alternativas, inclusive com grande parte dos itens tendo assinalada a mesma alternativa. Um ponto importante que ocorre com os três níveis menores é a inversão de posição entre eles em diversos itens, sendo o Nível 1 o que mais aparece com percentual de escolha maior que os demais. Entretanto, não podemos confundir o percentual de escolha com assinalar a alternativa correta, pelo contrário, e isso fica perceptível em diversos itens, por exemplo, nos itens 11 e 19, para os quais o Nível 1 tem um percentual de escolha de uma das alternativas maior que os Níveis 2 e 3. Porém, a alternativa mais escolhida não é a considerada correta para o item, ao mesmo tempo que se diferencia das que os demais níveis mais assinalaram internamente em cada grupo. O Gráfico 5 nos traz os dados das alternativas para o Teste 2 de 2014.

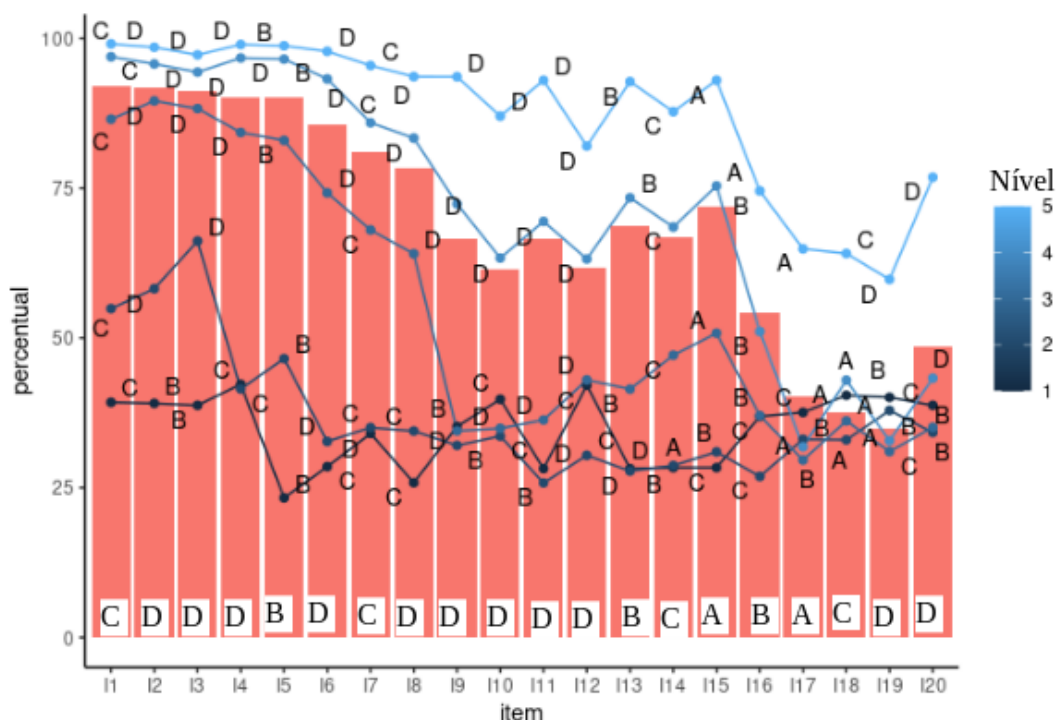
**Gráfico 5** – Distribuição do percentual das alternativas mais escolhidas por nível e o percentual de acerto por item. Teste 2. 2014



**Fonte:** Elaboração própria.

Em relação às análises das escolhas das alternativas em cada nível, temos algo muito semelhante ao que encontramos no Gráfico 3, com uma oscilação maior das alternativas mais escolhidas no Nível 1. Outrossim, o Gráfico 5 nos traz percentuais de acertos menores que os dois gráficos anteriormente analisados, em que todos os itens após o décimo apresentam percentual menor que 75% e com cinco itens com valores menores ou próximos a 50% de acerto. O que demonstra que o instrumento ficou difícil para os alunos ao final do 2º ano do Ensino Fundamental. O Gráfico 6 nos traz as informações do desempenho dos alunos no Teste 2 de 2015.

**Gráfico 6** – Distribuição do percentual das alternativas mais escolhidas por nível e o percentual de acerto por item. Teste 2. 2015



**Fonte:** Elaboração própria.

O Gráfico 6 nos traz dados do Teste 2 de 2015 que muito se assemelham aos demais gráficos analisados com a mesma metodologia, mas com a diferença de unir, no mesmo gráfico, todas as tendências que aparecem, de certa forma, separadas nos anteriores. Entre alguns pontos que chamam atenção no gráfico, temos 12 itens que estão com percentuais abaixo de 75% e três itens com valores próximos a 30%, o que demonstra, como já mencionado no Teste 2 de 2014, que o instrumento ficou mais “difícil” para o conjunto de alunos do 2º ano de 2015.

No geral, é possível identificar que: existem diferenças de escolhas das alternativas entre os níveis; os itens acabam sendo mais “fáceis” para os alunos do Nível 5 e difíceis para os do Nível 1; existem alternativas que atraem mais alguns níveis, mesmo nós não realizando uma análise completa das alternativas de cada item e; os itens mais difíceis são os que mais geram dispersões entre as escolhas das alternativas, algo trivial, mas que é importante ser destacado, pois contribui para o cerne do TAI-Leitura, adaptabilidade conforme o conhecimento do aluno, logo, aproximando a alternativa correta ao aluno que a consegue escolher.

O próximo capítulo vai nos apresentar algumas características das principais teorias das medidas educacionais e psicométricas, com o intuito de contribuir para a definição de um modelo estatístico que nos permita ter um TAI-Leitura que considere as alternativas no processo de aplicação.



## **2 – Medidas educacionais e os desafios para descobrir o que sabem os alunos: alguns delineamentos e resultados**

O uso de medidas em avaliação educacional é uma prática nas escolas, contudo tais medidas se baseiam, geralmente, em instrumentos que não foram construídos com base em padrões psicométricos válidos e aceitáveis, como, por exemplo, serem submetidos a um pré-teste dos itens e, sobretudo, dos próprios instrumentos. Outra constatação é que as professoras e os professores raramente se dão conta dessas limitações e fazem uso de resultados que não indicam o quanto, efetivamente, seus alunos sabem daquilo que constitui o conteúdo desses instrumentos. As consequências podem ser, igualmente, caracterizadas por limitações que chegam a comprometer o fluxo dos alunos, como seria reprovar um aluno por supor que ele “não sabe o suficiente” quando isso não corresponde à realidade, ou assumir que alunos aprenderam quando não foi isso que teria ocorrido e, portanto, ficam com lacunas no aprendizado que, ao contrário, deveriam ter sido enfrentadas com revisões de conteúdo.

De toda forma, uma medida que represente quantitativamente o domínio de cada estudante sobre o conteúdo da avaliação, associado aos componentes curriculares, é frequentemente atribuída em cada bimestre e ao final do ano, servindo de balizadora na decisão, por parte das professoras e dos professores, em promover ou não um estudante. Obviamente, essas afirmações se caracterizam como um exemplo que se encontra na maioria das escolas em nosso país.

Nesse sentido, este capítulo tem por objetivo abordar as duas principais teorias que possibilitam a constituição da medida educacional. Todavia, é importante que destaquemos um pressuposto que necessita ser considerado na análise de qualquer medida, principalmente da medida relacionada à área psicossocial, a saber: toda e qualquer medida relaciona um número a um atributo do fenômeno que estamos a medir. Por exemplo, quando queremos medir o comprimento de algum objeto, usamos um instrumento que também guarda o mesmo atributo do objeto, no caso, o comprimento. Pasquali (2013, p. 37) define esse processo como “medida fundamental”, em que se pode estabelecer uma unidade-base natural e específica e, dessa forma, a relação entre o que está se medindo e o instrumento de medida permite que a definição dos números que compõem o instrumento seja transferida para o objeto, permitindo, portanto, que se obtenha a medida dele.

Entretanto, existem medidas para as quais essa relação não consegue ser feita de maneira direta, como, por exemplo, a distância entre estrelas, ou a massa de uma determinada galáxia ou a

capacidade de leitura de um estudante. Os dois primeiros exemplos necessitam de uma medida indireta, porém que se diferenciam em relação à medida baseada na teoria psicossocial, pois existem fenômenos físicos que respeitam determinadas leis físicas e que, com conhecimento atual da humanidade, não vão desrespeitá-las e, assim, a medida pode ser derivada. Contudo, no caso de processos psicológicos, que ocorrem internamente às pessoas, não temos esses fenômenos externos que podem ser utilizados como base, mas, sim, definições e compreensões sobre o que venham a ser processos concretos derivados, ou melhor, causados por esse fenômeno que estamos querendo medir. Para Pasquali (2013, p. 40), esse tipo de medida é delimitada como “medida por teoria” que,

[...] quando nem leis existem relacionando variáveis, pode-se recorrer a teorias que hipotetizam relações entre os atributos da realidade, permitindo assim a medida indireta de um atributo através de fenômenos a ele relacionados via teoria. O importante neste caso é garantir que haja instrumentos calibrados para medir (fundamentalmente ou de outra forma válida) os fenômenos com os quais o atributo em questão esteja relacionado pela teoria.

A medida por teoria, conforme define Pasquali (2013, p. 45), pode ser dividida em: teoria dos jogos, teoria psicofísica e teoria psicométrica ou a teoria dos testes psicológicos. A última, por sua vez, nos interessa para a análise do objeto de trabalho, pois como Pasquali (2013, p. 45-46) a define,

Esta teoria trabalha igualmente com dois parâmetros, a saber, a resposta (comportamento) do sujeito e o critério. Pelo fato de que o critério é entendido de diferentes maneiras, surgem duas teorias psicométricas bastante distintas, quais sejam, a Teoria Clássica dos Testes (TCT), que entende o critério como comportamento (futuro), e a Teoria da Resposta ao Item (TRI), que entende como critério o traço latente.

É importante enfatizar que a relação entre medida e Psicologia não é unânime, pelo contrário, é conturbada e com diversas correntes de entendimento, principalmente, no que tange às análises dos instrumentos e, por consequência, dos construtos definidos. Para Guttman (1971), existem diversos questionamentos sobre a medida na área da Psicologia, que o autor define como “Teoria da Medida”, sobretudo sua relação com a teoria estatística. Pasquali (2013, p. 53), além de explorar o emprego da medida na Psicologia, traz uma distinção entre a Psicometria, que é um ramo da Psicologia, e a Estatística. Nesse sentido, o autor considera que “a Psicologia não tem, como objeto de estudo, parâmetros e sim processos comportamentais, processos psíquicos [...]”. Dessa maneira, o processo de medir na Psicologia, assim como na

Educação, se fundamenta em algo que não é observado diretamente, pois está interno ao respondente e o que observamos, por meio das respostas dadas aos itens e de maneira indireta, são os efeitos daquilo que estamos querendo medir.

Nestes termos, os efeitos de um processo psíquico, cuja magnitude estamos querendo medir, são apreendidos de forma indireta, dado ser impossível medir diretamente, verificando respostas a itens – tarefas – que respondentes apresentam sobre determinado conteúdo que compõe os itens, supondo que esse comportamento pode ser compreendido como ancorado em uma capacidade que, por ser interna a esse respondente, é considerada um traço latente. De qualquer forma, assim como define Pasquali (2013, p. 63), “[p]ostula-se que, ao se operar sobre o sistema comportamento, está-se operando sobre os traços latentes (isomorficamente). Assim, a medida que se faz ao nível comportamental é a medida dos traços latentes”. Nessa perspectiva, assim como já citado, temos duas teorias que buscam explicar as relações entre respostas dadas aos itens de instrumentos e a medida da capacidade do respondente da qual decorrem essas respostas: a Teoria Clássica dos Testes (TCT) e a Teoria da Resposta ao Item (TRI).

### **2.1 – As duas teorias mais relevantes para as medidas educacionais: a TCT e a TRI**

No processo de medidas educacionais, uma etapa incontornável é a análise dos itens que, entre outros objetivos, busca informações que assegurem que os itens e os instrumentos aos quais estão associados são adequados ao processo de medição para o qual são utilizados. Dessa maneira, buscamos ter evidências do quanto uma pessoa sabe em relação a um determinado objeto de conhecimento que, por sua vez, pode ser o objeto de avaliação.

Para tanto, desde o século XIX, diversos pesquisadores procuram compreender as relações entre os processos que ocorrem internamente às pessoas e o seu comportamento no mundo em que vivem e interagem. Entre as mais variadas análises estatísticas, duas teorias ganharam notoriedade e que serão exploradas na sequência, sem exaurir as possibilidades analíticas e suas problematizações. Tratam-se da Teoria Clássica dos Testes (TCT) e da Teoria da Resposta ao Item (TRI).

### 2.1.1 – Teoria Clássica dos Testes (TCT): algumas características

A TCT tem como principal função contribuir para a compreensão do escore total do teste, o que implica uma análise do instrumento como um todo e o modelo matemático que fundamenta a TCT pode ser escrito da seguinte forma:

$$T = V + E$$

Onde:

T = escore bruto do respondente, sua pontuação no teste ou o resultado empírico.

V = escore verdadeiro – *true score* em inglês –, ou seja, aquilo que de fato queremos medir com o teste desconsiderando o erro de medida do resultado do teste.

E = erro da medida associado ao resultado empírico.

É importante destacar dois pontos que estão presentes no modelo da TCT, a saber: o escore verdadeiro é baseado na ideia de Esperança matemática, que seria um resultado que teríamos se fosse possível aplicar infinitos testes para um determinado respondente, sem que houvesse influência de uma aplicação na outra, além de nenhum efeito interno ou externo ao respondente, como por exemplo, o cansaço ou o avanço do aprendizado. Importante salientar que as análises aqui apresentadas estão vinculadas a itens de múltipla escolha, embora isso possa ser aplicado, com algumas adaptações para itens de resposta construída ou itens de atuação.

Dessa maneira, em um determinado momento teríamos o que poderia ser considerado como o escore que de fato corresponde à verdadeira medida a ser analisada: a quantidade de acertos que o respondente teria com o domínio que possui do conteúdo do teste **sem erro**. Ao mesmo tempo, também, devemos considerar que o erro de medida em cada uma das aplicações é independente, e que decorre de fatores não sistemáticos, como, por exemplo, a dúvida do respondente em algum item sobre qual seria a resposta a ser apresentada; logo, se calcularmos a média dos erros teremos um resultado igual a zero ao final. Segundo Primi (2012, p. 300),

O escore de erro pode ser entendido como uma variável aleatória associada a eventuais erros associados às condições particulares da aplicação. [...] Assuma-se que o erro seja assistemático, aleatório, ou seja, não mostra tendência sistemática de assumir valores positivos ou negativos.

Além da análise do teste como um todo, temos a possibilidade de investigar o comportamento de cada um dos itens do teste e para tanto, no âmbito da TCT, podemos classificar o item em relação à sua dificuldade, com base na proporção de marcações na alternativa considerada correta, assim como as proporções para as demais alternativas. O que nos possibilita verificar se existe alguma alternativa incorreta que esteja sendo mais assinalada, principalmente, pelos respondentes que obtiveram um maior escore no teste. Em relação ao conjunto de itens do teste, esperamos que existam itens com diversos níveis de dificuldade, porém, como a TCT tem um enfoque no teste, essa dificuldade pode variar conforme a população para a qual o teste é submetido, sendo, sempre, desejável que o teste tenha variabilidade de dificuldade, pois, assim, será possível diferenciar melhor os respondentes conforme a capacidade de acertar ou não cada item.

Essa interpretação da diferenciação entre aqueles que conseguem escolher a alternativa considerada correta e os que escolhem outras alternativas, pode ser feita por meio da correlação entre a proporção de acertos do item e o escore total do teste. Na análise de itens, essa diferenciação é intitulada de “discriminação dos itens” e na análise pela TCT temos algumas formas de calcularmos esses índices, como: grupos critérios (índice D e teste t); coeficiente bisserial e correlação ponto-bisserial.

Para calcular a discriminação por meio dos grupos critérios, primeiro é necessário calcular o escore no teste para cada respondente e classificar os 27% com menor (grupo inferior) e maior (grupo superior) escore no teste. Após essa classificação, o índice D é definido a partir da diferença entre os percentuais de acerto para o grupo superior e grupo inferior. Esperamos que esse valor seja positivo, pois assim, teremos um percentual de acerto maior para o grupo superior. Já o teste t consiste em um teste estatístico que busca verificar se as médias de acertos obtidas em ambos os grupos são diferentes, com um índice positivo, pois, caso contrário, teremos um item em que o grupo inferior terá uma média maior que o grupo superior.

O coeficiente bisserial, conforme define Pasquali (2013, p. 136), “é uma estimativa da correlação de Pearson e é utilizada na situação na qual as variáveis correlacionadas são contínuas, mas uma delas (no caso, o item) foi artificialmente reduzida a duas categorias”. Dessa maneira, temos que o coeficiente bisserial é calculado por meio da fórmula,

$$r_b = \frac{\bar{X}_i - \bar{X}_t}{S_t} \cdot \frac{pq}{y}$$

Onde:

$\bar{X}_i$  = média dos respondentes que acertam o item;

$\bar{X}_t$  = média de acertos no teste;

$S_t$  = desvio padrão dos acertos no teste;

p = proporção de respondentes que acertam o teste;

q = 1 – p; e

y = área da curva normal à esquerda do valor da proporção de respondentes que acertam o teste (p).

Já o coeficiente ponto-bisserial – a correlação de Pearson –, segundo Primi (2012, p. 306), “expressa a correlação entre uma variável categórica dicotômica (acerto ou erro) e uma variável intervalar (o escore no teste que, embora não possa ser considerada uma variável intervalar genuína, para fins práticos é considerada como tal)”. A fórmula para calcular este coeficiente, com os mesmos elementos da equação anterior, é

$$r_{pb} = \frac{\bar{X}_i - \bar{X}_t}{S_t} \cdot \sqrt{pq}$$

As análises que buscam verificar a capacidade do item em diferenciar alunos com escores diferentes, ou melhor, de discriminar esses escores, em certa monta, estão baseadas em uma medida de um contínuo. Sendo assim, a pontuação obtida no teste precisa ser considerada como uma medida contínua, que esteja ancorada em um construto, um traço latente. Entretanto, na análise pela TCT não estamos de fato medindo esse traço, mas o comportamento em relação aos itens. Ao mesmo tempo, é necessário que tenhamos a certeza de que o instrumento esteja medindo o que, a priori, queríamos que fosse medido, pois, considerando que com o uso da TCT queremos somar os escores de cada item para obtermos o escore total, não podemos somar coisas diferentes, o que não teria sentido prático.

Essa necessidade de “coisas iguais” para serem somadas, que constituiria um traço latente comum para medirmos, pode ser definida como análise de dimensionalidade do instrumento. Pasquali (2013, p. 114-115) apresenta esse conceito de dimensionalidade da seguinte forma:

Quando se está analisando uma série de itens, tanto pela TCT quanto pela TRI, supõem que todos eles estejam medindo a mesma coisa. No caso da TCT, isto ocorre porque ela trabalha com o escore total e cada item é avaliado em função deste escore total; acontece, porém, que o escore total consiste na soma das

respostas dadas aos itens; assim, ela faz a suposição que eles são somáveis e isto faz sentido somente se eles se referem à mesma coisa, pois não dá para somar alhos e bugalhos. Alhos e bugalhos dão um agregado, não uma soma. No caso da TRI, todos os itens são avaliados em função de um traço latente.

Para realizarmos a análise de dimensionalidade existem diversas técnicas estatísticas que a propiciam a partir das respostas dadas aos itens de modo a reduzirmos a um fator comum, ou melhor, a uma única dimensão. Dentre essas técnicas a mais conhecida é a Análise Fatorial (AF)<sup>12</sup>, que por meio de técnicas estatísticas multivariadas, permite reduzirmos o conjunto de respostas à menor quantidade de dimensões possíveis e que possam fazer sentido para os dados analisados.

Com maior confiança em relação ao que estamos medindo, as análises de itens pela TCT ganham maior veracidade, pois temos maior certeza de que o escore total se refere ao que queríamos medir. Sendo assim, as possibilidades de análises com a TCT ficam mais fidedignas e confiáveis, porém devemos destacar que existem limites para as investigações pela TCT, principalmente no que se refere à necessidade de manter o instrumento inalterado em novas aplicações, sobretudo para comparabilidade de respondentes e populações, pois qualquer mudança de algum item no instrumento, muito provavelmente, irá gerar uma nova e diferente escala de medida, o que impossibilita comparações com a medida anterior. Existem, na literatura, afirmações sobre a possibilidade da criação de testes paralelos, porém sem grandes demonstrações de que realmente seja possível realizar comparações de maneira direta e confiável.

Dessa forma, ao implementarmos um instrumento que venha medir, por exemplo, a capacidade de leitura nos anos iniciais do Ensino Fundamental ao longo de um período, o uso da TCT acaba nos obrigando a manter o mesmo instrumento. Essa fixação do instrumento, impediria que, por exemplo, itens mais difíceis fossem incorporados ao teste em função de uma suposta evolução da capacidade dos alunos em decorrência do ensino e do estudo; itens que num determinado momento seriam tão difíceis que não permitiriam informações pedagógicas, com acertos eventuais sendo fruto do acaso. Ainda, o emprego da TCT, no sentido contrário, impediria que itens muito fáceis fossem eliminados, pois deixariam de ser um “desafio” pedagógico para os alunos, inclusive com a possibilidade de erro devido ao “desinteresse” ou por descuido ao assinalar a alternativa correta, ou seja, igualmente sem informações pedagógicas relevantes.

---

<sup>12</sup> Para maiores informações, consultar Hair (2006).

Dessa forma, se realizarmos mudanças nos itens que compõem o instrumento, muito provavelmente criaremos uma escala nova, o que impossibilitaria a comparação direta.

Diante dessa restrição da TCT, a utilização da Teoria da Resposta ao Item (TRI) ganha relevância, pois essa, como veremos a seguir, nos possibilita manter o foco no item para basear a medida no funcionamento de cada item dentro do teste, sem com isso deixar de nos preocuparmos com o teste como um todo.

### **2.1.2 – Teoria da Resposta ao Item (TRI): algumas características**

A TRI pode ser caracterizada, de modo amplo, como um conjunto de modelos matemáticos que contribuem para calcular a probabilidade de um respondente acertar um item, dada sua magnitude no traço latente que é medido por meio dos itens que responde. Para Muniz (1997, p. 17), a grande contribuição da TRI está na “*posibilidad de obtener mediciones invariantes respecto de los instrumentos utilizados y de los sujetos implicados*”, o que, por sua vez, diferencia a TRI totalmente da TCT, pois centra as análises nos itens ao invés do teste como um todo.

Muniz (1997, p. 15) apresenta os autores que foram os precursores no desenvolvimento da TRI, como, por exemplo, Birnbaum (apud LORD; NOVICK, 1968) e Lord (1977, 1980). Com o avanço da computação, mais estudos foram sendo implementados e diversos programas de computadores foram desenvolvidos para que fosse possível realizar análises com a TRI. Deve-se sublinhar que a TRI foi se materializando em vários modelos, diferenciados em função dos parâmetros utilizados para os cálculos da referida probabilidade.

A TRI, independentemente do modelo, tem alguns pressupostos que precisam ser respeitados para se evitar divergências nas estimativas dos parâmetros e da própria medida do traço latente, denominada proficiência. Além dos objetivos apresentados com base em Muniz (1997), há a possibilidade de se obter medições invariantes independentemente do instrumento utilizado e da população analisada, considerando como pressupostos a unidimensionalidade e a independência local dos itens.

A unidimensionalidade está relacionada com a probabilidade de um indivíduo responder corretamente um item depender somente da capacidade que esse indivíduo tem, sendo que essa capacidade é a mesma que o teste quer medir, isto é, a dimensão do teste que é, praticamente, única para todos os itens e se relaciona com o conteúdo do teste. Em outras palavras, a TRI



pressupõe que o conjunto de itens que compõem o instrumento tem um fator predominante. Sendo que infringir esse pressuposto pode levar a interferências na estimativa dos parâmetros dos itens, principalmente aquela relativa à propriedade do item em discriminar respondentes, quanto à probabilidade em acertá-lo em função da proficiência desses respondentes – o parâmetro  $a$  –, que será melhor detalhada mais adiante.

Para identificar a existência de uma única dimensão no instrumento analisado, para além da Análise Fatorial (AF), temos a *Full Information Factor Analysis* (FIFA) que é baseada na TRI. Essa análise, segundo Pasquali (2013, p. 120), tem grandes vantagens em relação às demais técnicas, a saber:

- 1) leva em conta toda a informação empírica da aplicação do teste, não somente informações sintetizadas, como é o caso na AF que utiliza as correlações entre as variáveis;
- 2) consegue trabalhar o acerto dado ao acaso;
- 3) consegue tratar os dados omissos; e
- 4) consegue contornar os problemas da matriz não positivo – definida, bem como os casos Heywood<sup>13</sup> que ocorrem frequentemente na AF.

A independência local está intrinsecamente relacionada com a unidimensionalidade, pois, segundo Muñiz (1997, p. 26-27), “*para un sujeto con un determinado valor en la variable unidimensionalidad su respuesta a un ítem no viene influida por sus respuestas en los otros*”. Dessa maneira, cada item é considerado como uma unidade independente dos demais, o que possibilita que diversos estudantes respondam instrumentos diferentes ou parcialmente diferentes e tenham a medida do traço latente na mesma escala, pois não existe uma dependência do conjunto de itens, mas de cada item respondido.

### **2.1.2.1 – Os modelos da TRI**

Os itens dicotômicos ou dicotomizados, geralmente, utilizam, no processamento pela TRI, os modelos logísticos pelo fato de a função logística ter uma, em sentido figurado, maior maleabilidade matemática. Além disso, para Muñiz (1997) os modelos logísticos de 1, 2 e 3 parâmetros,

[...] *assumen (aquí) que la respuesta a los ítems es dicotómica, es decir, o se acierta o se falla el ítem, independientemente del número de alternativas que tenga, o que sea de carácter abierto en el que los sujetos deban generar su*

---

<sup>13</sup> Estimativas de parâmetros que podem ser zero ou negativas.

*propia respuesta, en cuyo caso ésta sólo se considerará correcta o incorrecta, sin grados intermedios.*

Na Tabela 2 apresentamos os modelos logísticos de 1, 2, 3 e 4 parâmetros, sendo importante destacar que o modelo logístico de 1 parâmetro, formulado inicialmente por Rasch, é utilizado no processamento da PB e que os modelos de 2 e 4 parâmetros não são tão usuais nas avaliações em larga escala no Brasil. Quanto aos demais, de todo modo, apresentá-los aqui se faz relevante.

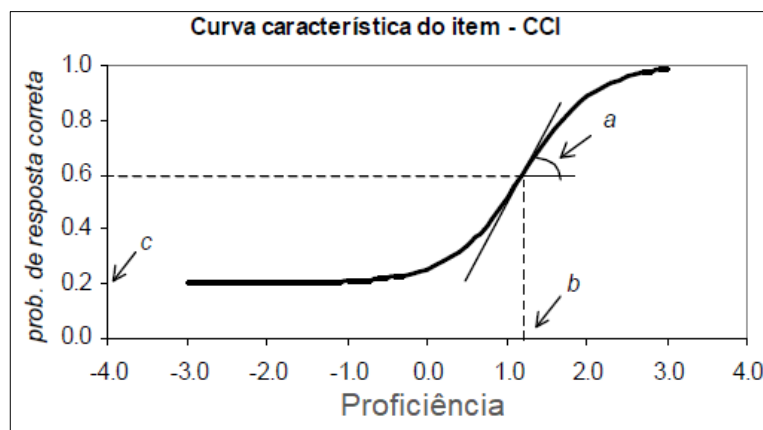
**Tabela 2 – Modelos logísticos da TRI**

Um parâmetro	$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}$
Dois parâmetros	$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$
Três parâmetros	$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$
Quatro parâmetros	$P_i(\theta) = c_i + (Y_i - c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$

**Fonte:** Adaptado de Muñiz (1997).

O aspecto mais relevante é que os modelos da TRI assumem que existe uma função que relaciona a probabilidade de acertar um item e os valores das variáveis que eles medem, e essa função pode ser representada graficamente pela Curva Característica do Item (CCI). De uma maneira mais simplificada, Muñiz (1997) define essa relação de tal modo que a “probabilidade de acertar um item só depende dos valores das variáveis medida pelo item”. Salientamos que a pesquisa na qual se baseia esta Dissertação não estava focada somente na probabilidade de acertar o item, mas em investigar e estimar a probabilidade de marcar cada uma das alternativas de um item, incluindo a considerada correta. O Gráfico 7 apresenta uma CCI de um modelo logístico de três parâmetros.

**Gráfico 7** – Curva Característica de um Item (CCI) para o modelo logístico de 3 parâmetros



Fonte: Andrade, Tavares e Valle (2000).

No Gráfico 7, com uma CCI para itens dicotômicos, a escala de proficiência e os parâmetros do modelo da TRI estão destacados com as letras “a”, “b” e “c”. A escala para dimensionamento do traço latente – a medida da proficiência – é arbitrária, considerando que, segundo Andrade et al. (2000, p. 10), “o importante são as relações de ordem existentes entre seus pontos e não necessariamente sua magnitude”. O parâmetro b está na mesma escala do traço latente e esse parâmetro é análogo à própria medida do traço latente, uma vez que representa o quanto de proficiência o estudante precisa para ter maior probabilidade de acertar um item, de tal forma que quanto maior o parâmetro b menor é a probabilidade de acerto para valores de proficiências muito menores do que o de b, que por sua vez, mais “difícil” é considerado o item. Inversamente, quanto menor o valor do parâmetro b, mais “fácil” é o item com as relativas interpretações no que tange à relação entre o traço latente a probabilidade de acerto do item.

O parâmetro c, cujo valor pode variar entre 0 e 1, não depende diretamente da escala do traço latente e é a probabilidade de um estudante com baixa medida do traço latente acertar o item. Esse parâmetro, por diversas vezes, é relacionado com o “chute” em um item, mas não é desse modo que se deve interpretá-lo.

O parâmetro a representa a discriminação de um item, isto é, o quanto um item consegue separar os estudantes que apresentam maior ou menor probabilidade de acertar ou o errar o item. Andrade et al. (2000, p. 11) definem o parâmetro a como sendo

[...] proporcional à derivada da tangente da curva no ponto de inflexão. Assim, itens com “a” negativo não são esperados sob esse modelo, uma vez que indicariam que a probabilidade de responder corretamente o item diminui com o aumento da habilidade.

O processo de obtenção dos parâmetros dos itens – denominado parametrização dos itens– pode ocorrer em conjunto com a estimação dos parâmetros do traço latente, possibilidade considerada um problema de estimação por Andrade et al. (2000, p. 27) e pode ser dividida da seguinte forma:

[...] do ponto de vista teórico, podemos dividir o problema em três situações, quando já conhecemos os parâmetros dos itens, temos apenas que estimar as habilidades; se já conhecemos as habilidades dos respondentes, estaremos interessados apenas na estimação dos parâmetros dos itens e, por fim, a situação mais comum, em que desejamos estimar os parâmetros dos itens e as habilidades dos indivíduos simultaneamente.

De toda maneira, o processo de parametrização em duas etapas tem sido usual nos processamentos de avaliações educacionais, principalmente pelo Sistema de Avaliação da Educação Básica (Saeb) que utiliza esse procedimento desde 1997, quando houve a introdução da TRI. Os procedimentos implementados no Saeb são mais bem explicados em Klein (2003), de onde destacamos a estimação dos parâmetros dos itens e traços latentes para múltiplos grupos, processo que ocorre quando se quer comparar, em uma mesma escala, estudantes de anos escolares diferentes e, de maneira muito comum, estudantes de anos civis diferentes. Tal tema será tratado de modo mais aprofundado adiante, quando abordaremos o número de populações.

### **2.1.2.2 – Os modelos politômicos**

Existem diversos modelos para o tratamento de respostas a itens politômicos, ou seja, os itens para os quais não temos somente uma resposta tratada exclusivamente como certa ou errada, mas com uma gradação – com categorias – das respostas. Na pesquisa buscamos modelos que tenham como foco os itens de múltipla escolha em que suas alternativas não tenham uma ordenação prévia, por mais que exista, por parte dos elaboradores de itens, uma ideia de plausibilidade sobre os processos cognitivos associados à marcação de cada uma das alternativas que podem sugerir uma diferenciação na proficiência dos respondentes.

Nesse sentido, identificamos, inicialmente, dois modelos, o Nominal e o Aninhado (*Nested Logit*, em inglês), que poderiam nos ajudar na elaboração de um TAI que considere a análise das alternativas em seu processo de aplicação. A Tabela 3 apresenta os dois modelos e como os parâmetros se relacionam.

**Tabela 3** – Modelos Nominal e Aninhado da TRI para itens politômicos

Modelos politômicos da TRI	
Nominal	$P_{iv}(\theta_j) = \frac{\exp^{Z_{iv}(\theta_j)}}{\sum_{k=1}^{m_i} \exp^{Z_{iv}(\theta_j)}}$
Aninhado	$P(U_{ij} = 1 \vee \theta_j) = \frac{1}{1 + \exp^{-Da_i(\theta_j - b_i)}}$ <p>E</p> $P_{iv}(U_{ij} = 0 \vee \theta_j) = \left\{ 1 - \left[ \frac{1}{1 + \exp^{-Da_i(\theta_j - b_i)}} \right] \right\} \left[ \frac{\exp^{Z_{iv}(\theta_j)}}{\sum_{k=1}^{m_i} \exp^{Z_{iv}(\theta_j)}} \right]$

Fonte: Elaboração própria.

Onde:

$P_{iv}(\theta_j)$  = probabilidade de um respondente escolher uma alternativa dado o item, a alternativa e o theta;

$P(U_{ij} = 1 \vee \theta_j)$  = probabilidade de um respondente acertar um item dada a medida do seu theta;

$P_{iv}(U_{ij} = 0 \vee \theta_j)$  = probabilidade de um respondente escolher uma alternativa que seja considerada como errada no item, dada a medida do seu theta;

$a_i$  = parâmetro de discriminação do item;

$b_i$  = parâmetro de dificuldade do item; e

$D$  = constante que pode ser igual a 1, quando queremos que a função seja logística ou, igual a 1,7 para termos uma aproximação da função normal.

O modelo Nominal estabelece, segundo Bock (1972, p. 30), a igualdade  $Z_{iv}(\theta_j) = a_{ik}\theta_j + c_{ik}$ , em que os parâmetros  $a$  e  $c$  representam, respectivamente, a discriminação e o intercepto, sendo

que, para o intercepto, Andrade et al. (2000, p. X) o consideram como sendo o parâmetro de dificuldade da alternativa. De toda maneira, a modelagem das alternativas representada, pelo modelo Nominal, nos dá condições de obter probabilidades para cada categoria escolhida por cada estudante, o que, por sua vez, nos permite definir um parâmetro de posição na escala de proficiência e, assim, utilizar no algoritmo de escolha do item do TAI.

Em linhas gerais, a probabilidade de um estudante escolher uma alternativa está relacionada com os parâmetros da alternativa escolhida e das demais. Dessa maneira, temos como possibilidade esperada que a alternativa considerada como a correta seja aquela que terá os maiores valores para ambos os parâmetros; logo, em uma ordenação, teremos essa alternativa como sendo a primeira. Todavia, essa não é uma afirmação que pode ser considerada como totalmente aceitável, pois como o modelo nominal pressupõe que não existe uma ordenação nas categorias, é possível que uma alternativa considerada como incorreta possa ficar à frente da alternativa considerada como o gabarito. Obviamente, tal situação indicaria algum problema na formulação do item ou na população para a qual o item foi submetido não tem conhecimento suficiente para respondê-lo corretamente e, de qualquer maneira, será necessário analisar esse item.

O modelo Aninhado, por sua vez, faz uma junção entre os modelos dicotômicos e o Nominal, de maneira que a categoria considerada como a correta possa ser modelada por um modelo dicotômico da TRI (2PL ou 3PL) e as categorias consideradas incorretas sejam modeladas pelo modelo Nominal. Essa relação possibilita que se estabeleça uma hierarquia dentre as alternativas, mantendo a categoria considerada correta como sendo a do primeiro nível e as demais como de segundo. Em Bolt et al. (2010, p. 457), o modelo “Nested Logit”, que neste trabalho estamos intitulado por “Aninhado”, é apresentado com a relação entre o 3PL e Nominal, porém deixa claro que não existe nenhuma diferença entre a maneira como os dois modelos se relacionam, somente com relação ao modelo dicotômico utilizado.

Na Tabela 3 apresentamos o modelo Aninhado utilizando o modelo de 2PL para a parte dicotômica, pois compreendemos que seja o modelo que mais se aproxima do Nominal e introduz somente um parâmetro em relação ao modelo Rasch, o que, por sua vez, pode facilitar algumas comparações entre modelos, principalmente nos dois politômicos. Entretanto, antes de avançarmos para as comparações, precisamos apresentar os conceitos que contribuem para a

análise dos itens, são eles: a Função de Informação do Item (FII) e Função de Informação do Teste (FIT).

### 2.1.2.3 – Função de Informação do Item (FII) e Função de Informação do Teste (FIT)

A FII é uma medida amplamente utilizada para verificar o quanto um determinado item contém de informação em relação ao traço latente medido. Para Hambleton (1991, p. 91) as FII e as FIT são “um método poderoso de descrever itens e testes, selecionar itens de testes e comparar testes, fornecido pela Teoria da Resposta ao Item. O método envolve o uso de junções de função de informação do item”. Este mesmo autor destaca que o responsável por elaborar o modelo para ambas as funções foi Birnbaum em 1968. De qualquer maneira, temos modelos matemáticos de FII para cada um dos modelos da TRI, porém, como estamos analisando os modelos de 1PL, 2PL, Nominal e Aninhado, iremos somente apresentar esses modelos neste trabalho.

Para os modelos logísticos de 1 e 2 parâmetros, a Tabela 4 nos apresenta o modelo de probabilidade e a FIT.

**Tabela 4** – Modelos politômicos da TRI

Modelo	Derivada da $P_i(\theta)$	Informação
1 parâmetro	$DP_i(\theta)Q_i(\theta)$	$\sum_{i=1}^n D^2 P_i(\theta)Q_i(\theta)$
2 parâmetros	$Da_i P_i(\theta)Q_i(\theta)$	$\sum_{i=1}^n D^2 a_i^2 P_i(\theta)Q_i(\theta)$

**Fonte:** Adaptado de Muñiz (1997, p. 127).

Onde:

$$Q_i(\theta) = 1 - P_i(\theta)$$

Na Tabela 3 somos capazes de verificar a derivada da  $P(\theta)$  que podemos escrever como  $P_i'(\theta)$ , para cada um dos modelos logísticos, além das próprias Funções de Informação do Teste (FIT), pelas quais é possível observar a importância dos parâmetros para identificarmos o quanto cada

item pode contribuir para com a medida do traço latente. Já a FII, com base em Muñiz (1997, p. 127), cujos símbolos são os mesmos da equação anterior, é dada por:

$$I(\theta) \frac{[P_i'(\theta)]}{P_i(\theta)Q_i(\theta)}$$

A relação entre a derivada da CCI e as probabilidade de escolher a alternativa correta e as alternativas incorretas é a mesma para todos os modelos, apenas com diferença na quantidade de parâmetros considerada em cada modelo. Para a FII, retira-se o somatório, pois, assim, temos a FII para cada item que considerarmos no cálculo.

Para os modelos politômicos existem algumas diferenças em relação ao modelo utilizado para se calcular a FII e FIT, entretanto vamos nos ater às funções dos modelos Nominal e Aninhado neste trabalho. É importante destacar que existe uma relação entre esses dois modelos, pois, como já apresentado, o Aninhado é a junção dos modelos logísticos com o Nominal e, também, existe uma similaridade entre as FII e FIT de ambos. Para o Nominal, com base em Bock (1972, p. 44), a FII é dada

$$I_{iv}(\theta) = a_i' W_i a_i P_{iv}(\theta)$$

É importante observar que o  $W_i(\theta)$  é o Delta de Kronecker que permite expandir ou reduzir os índices dos coeficientes das expressões as quais ele aparece. Neste sentido, os pontos em que  $i$  e  $j$  são iguais,  $\delta$  recebe o valor 1, caso contrário, recebe o valor 0. Dessa maneira, temos uma variação nos valores de  $P_{iv}(\theta)(\delta - P_{iv}(\theta))$ . Para o modelo Aninhado, a FII é dada por

$$I_i(\theta_j) = P_{ij} I_{iu}(\theta_j) + \sum_{i=1}^{m_i} P_{ijv} I_{iv}(\theta_j)$$

Podemos observar que a FII para o modelo Aninhado é a composição entre a informação da alternativa considerada correta e a soma das informações de cada alternativa incorreta e, dessa maneira, temos uma ampliação na quantidade de informação que cada item pode fornecer em relação ao traço latente. O Erro Padrão de cada item em cada medida de traço latente é a função inversa da FII.

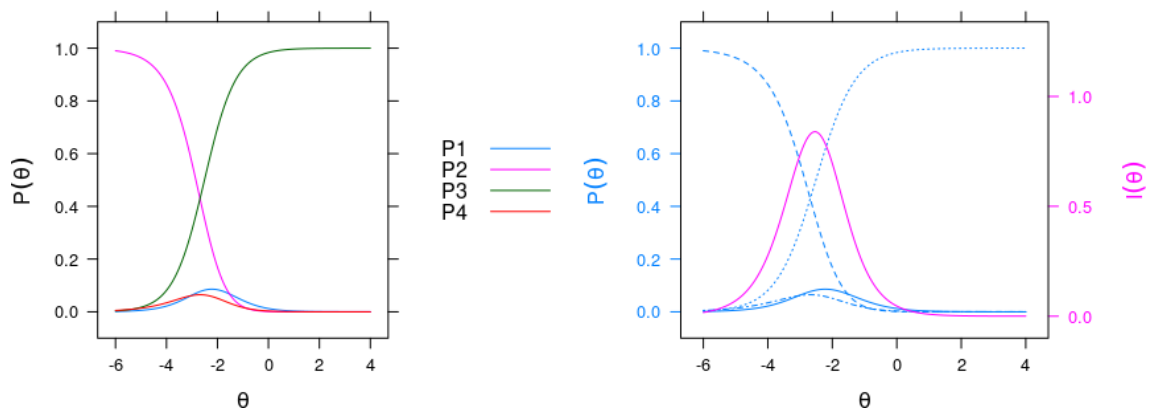
Em todos os modelos da TRI, o EP é o inverso da FII e, independentemente de estarmos considerando as alternativas incorretas ou somente as alternativas corretas, o erro em cada



medida do traço latente terá relação direta com a quantidade de informação do modelo no ponto analisado. É importante destacar que as distribuições da FII e do EP contribuem para que possamos compreender as potencialidades e limites que o instrumento utilizado pode nos oferecer.

Além das FII e FIT que geram dados para cada ponto da escala do traço latente, temos a representação gráfica de cada uma delas, que são conhecidas como Curva de Informação do Item (CII) e Curva de Informação do Teste (CIT). Ambas as representações podem ser acompanhadas pela também representação gráfica do EP e é possível observar a relação entre a informação do item ou do teste e o erro. Na Figura 8 temos as duas curvas demonstradas, sendo o gráfico da direita a CII em conjunto com uma CCI e o da esquerda a CII.

**Figura 8** – Curva Informação do Item em conjunto com a Curva Característica do Item



**Fonte:** Elaboração própria.

A CII combinada com a CCI possibilita verificarmos em qual região da escala do traço latente o item consegue obter maior informação. Isso é importante por dois motivos: itens com mais informações demonstram maior relação com o traço latente; logo, conseguem contribuir mais com a medida desse construto. Em um TAI a informação será essencial para que o item possa ser escolhido e, dessa maneira, itens menos informativos, a depender dos algoritmos utilizados, podem ficar sem serem utilizados em aplicações do TAI.

No caso do modelo Aninhado, com sua CCI exemplificada na Figura 8, temos a junção dos modelos dicotômico e politômico e, logo, esperamos uma curva logística que represente a categoria considerada como o gabarito e outras curvas que modelem as demais alternativas.

Então, devemos observar a CCI do modelo logístico aumentando o valor da probabilidade conforme avançamos nos valores dos  $\theta$ s. Ao mesmo tempo, quanto às curvas que modelam as demais alternativas, é esperado que exista uma diminuição da probabilidade conforme se avança nas proficiências ou um aumento até um ponto máximo em um determinado valor do traço latente para, depois, diminuir gradativamente. Essa relação entre as curvas, pressupõe que elas não vão ficar totalmente sobrepostas, porém, no caso do modelo Aninhado, pode ser que exista uma certa sobreposição, pois, como já apresentado, ele se baseia no Nominal, que não prevê uma ordenação das categorias.

Já a CIT permite que possamos analisar o espectro da escala para a qual os itens que compõem o instrumento conseguem cobrir. É de extrema importância essa informação, pois conseguimos ter uma melhor ideia de como o instrumento está cobrindo a escala, o que impacta diretamente a medida do traço latente. Além disso, plotar a CIT com o EP nos possibilita identificar os pontos em que o nosso instrumento apresenta fragilidades na medida do construto.

De maneira geral, todas as informações até aqui apresentadas dependem diretamente da quantidade de populações envolvidas na aplicação do instrumento, pois isso demanda organizações de estruturas do instrumento que vão interferir diretamente na maneira como podemos obter a medida. Sendo assim, o próximo subcapítulo aborda exatamente essa problemática das populações envolvidas em uma aplicação de uma prova.

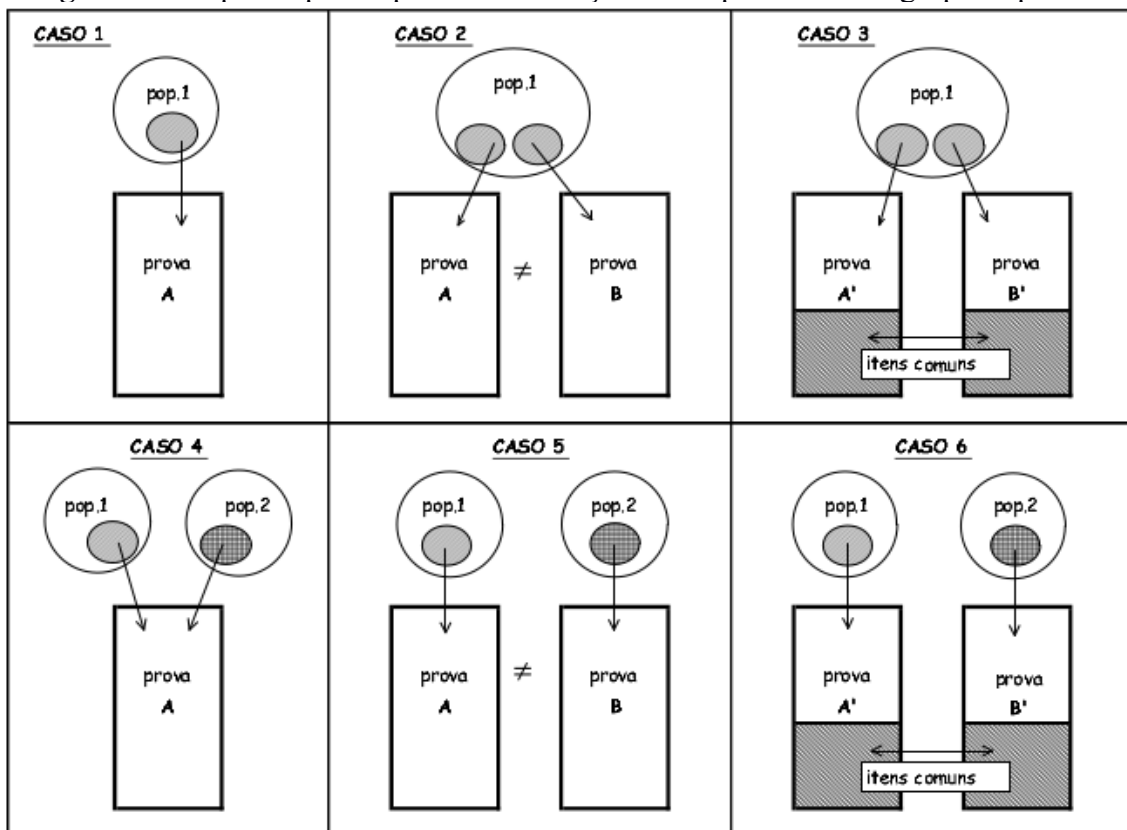
### **2.1.3 – O número de populações envolvidas**

A quantidade de populações envolvidas no universo da avaliação educacional implica estabelecer procedimentos para garantir comparabilidade entre tais populações. Existe uma lista elaborada por Andrade et al. (2000, p. 79) que sistematiza tais procedimentos, a saber:

1. Um único grupo fazendo uma única prova;
2. Um único grupo, dividido em dois subgrupos, fazendo duas provas, totalmente distintas (nenhum item comum);
3. Um único grupo, dividido em dois subgrupos, fazendo duas provas, apenas parcialmente distintas, ou seja, com alguns itens comuns;
4. Dois grupos fazendo uma única prova;
5. Dois grupos fazendo duas provas, totalmente distintas (nenhum item comum); e
6. Dois grupos fazendo duas provas, apenas parcialmente distintas, ou seja, com alguns itens comuns.

Essa lista de procedimentos é chamada de “equalização” e pode ser mais bem entendida em um esquema representado na Figura 9. Nele é possível verificar, mesmo que somente entre duas provas, como essas relações entre grupos e provas podem ocorrer. Nesta pesquisa, precisávamos estabelecer uma relação entre as diversas populações que fizeram a PB, sendo que existiam duas aplicações ao longo de um ano. Sendo assim, o procedimento mais adequado aos dados que dispúnhamos é o “Caso 6”, em que temos populações diferentes e instrumentos parcialmente diferentes. Entretanto, essa ligação será feita e explicada melhor no Capítulo 2.

**Figura 9** – Esquema para representar a relação entre quantidade de grupos e provas.



Fonte: Andrade et al. (2000, p. 80).

Os seis casos representados na Figura 9 sistematizam as relações que podem ocorrer entre grupos e provas, possibilitando entender como a estrutura de uma prova deve ser organizada para garantir que o processamento gere resultados que, efetivamente, sejam equalizados numa mesma escala. Todavia, é importante destacar que nas avaliações educacionais e, mais especificamente nas avaliações externas, o Caso 6 acaba sendo o mais utilizado, pois permite ter provas praticamente diferentes, mas com comparabilidade. Dessa maneira, para estarem na

mesma escala, as provas devem ser comparáveis; então, no Caso 6 existem provas aplicadas a populações diferentes e com parte dos itens iguais e, assim, preserva-se a comparabilidade entre as provas e as especificidades de cada uma delas.

Para que pudéssemos avançar em nosso objetivo na pesquisa, na busca de um TAI-Leitura que possibilitasse analisar as alternativas dos itens, precisávamos de um modelo da TRI que nos permitisse essa modelagem. Com esse intuito, realizamos um processamento estatístico com os modelos politômicos Nominal e Aninhado, com a finalidade de escolher entre um destes modelos. Ao mesmo tempo, como forma de compararmos com alguns dos modelos logísticos, também processamos os mesmos dados para os modelos de 1 e 2 parâmetros logísticos.

## **2.2 – Processamento estatístico para definição de um modelo da TRI: análise dos dados da Provinha Brasil de Leitura aplicada na Rede Municipal de Ensino de São Paulo**

Para que se obter um Banco de Itens que tenha parâmetros para todas as alternativas que compõe cada item, precisamos definir um modelo da TRI que nos possibilite essa modelagem. Para tanto, e com base nos dados da PB no período de 2012 a 2015, que obtivemos junto à RME, realizamos uma primeira análise com uma das bases de dados, na qual definimos a primeira aplicação de 2015, simplesmente por ser uma das bases mais recentes e ter sido aplicada no início do ano, que intuitivamente teria uma quantidade maior de itens em níveis menores da PB.

Destarte, além da definição do modelo que iríamos utilizar na composição do TAI-Leitura 3.0, precisamos calibrar todos os itens que seriam incorporados ao Banco de Itens (BI) do TAI. Sendo assim, a segunda parte desse subcapítulo será composta com os pontos principais da análise para calibração dos itens, incluindo as justificativas para as decisões tomadas e um certo detalhamento dos procedimentos realizados.

### **2.2.1 – Definição do modelo da TRI: processamento da Provinha Brasil primeira aplicação de 2015**

Como já indicado anteriormente, os dados da primeira aplicação de 2015 foram escolhidos para realizar o processamento com a TRI e por ele definir qual modelo utilizaríamos na composição do BI do TAI-Leitura. Para realizar os processamentos pela TRI, os dados foram organizados para atender aos padrões de cada um dos modelos utilizados para processá-los. Para os modelos

Rasch e 2PL, por serem dicotomizados, realizamos a binarização das alternativas, transformando a alternativa correta em 1 e as alternativas incorretas em 0. Já para os modelos politômicos, fizemos uma transformação direta, onde a alternativa A = 0, B = 1, C = 2 e D = 3.

O processamento pela TRI foi realizado por meio da linguagem R<sup>14</sup>, amplamente utilizada na área da estatística, ciências de dados, sociais, biológicas etc. No Ambiente de Desenvolvimento Integrado (IDE em inglês) trabalhamos com o *Rstudio*<sup>15</sup> e para realizar o processamento utilizamos o pacote *mirt*<sup>16</sup> (CHALMERS, 2016). Além deste, outros pacotes foram utilizados para organização e preparação das bases de dados e serão oportunamente apresentados.

Em linhas gerais, o processamento pelo *mirt* é dividido em três momentos diferentes, a saber:

- No caso de uma única população, a função *mirt* é utilizada para realizar todo o processo de iterações e estimativas iniciais dos parâmetros dos itens e das proficiências dos indivíduos. Se for mais de uma população no processamento, ao invés de utilizar a função *mirt*, devemos utilizar a *multiplegroup*, indicando qual grupo deve ser considerado como população de referência;
- Após essa primeira fase, já temos as estimativas dos parâmetros e proficiências, porém, sem que elas estejam organizadas em algum objeto do R. Dessa maneira, utilizamos a função *coef* para obtermos os parâmetros dos itens. Essa função possibilita que os parâmetros de dificuldade (b) sejam apresentados no formato conhecido na literatura ou multiplicado por menos o parâmetro de discriminação (a); e
- Por fim, para que seja possível manipularmos as proficiências estimadas para o traço latente analisado, devemos utilizar a função *fscore*. É importante destacar que na primeira fase os vetores que apresentaram somente valores faltantes, foram retirados automaticamente da análise e, logo, o resultado da função *fscore* não será compatível com a base de dados inicial. Assim, é importante retirar esses vetores (linhas) antes de iniciar o processamento da primeira fase, pois assim, a saída das proficiências será referente aos vetores de cada linha inicialmente na base de dados.

---

14 <https://www.r-project.org/>

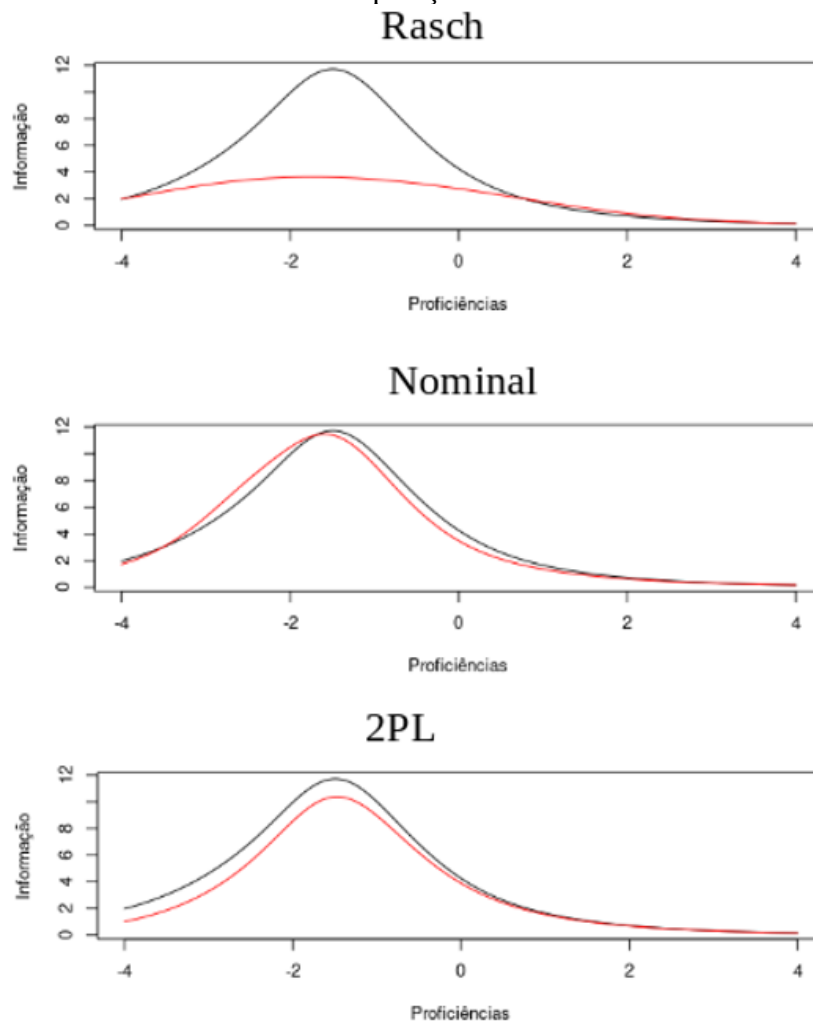
15 <https://www.rstudio.com/products/rstudio/>

16 <https://cran.r-project.org/web/packages/mirt/index.html>

Podemos solicitar que seja feita a estimativa do EP na primeira fase, assim, nas fases seguintes, será possível indicar que o erro deve ser impresso junto com os demais dados e obtermos uma estimativa completa. Além disso, também existe a possibilidade de gerar gráficos para a CCI, CCI em conjunto com a CII e a CIT, incluindo a distribuição do EP.

Para que possamos analisar os resultados, apresentamos na Figura 10 as CIT para os modelos 1PL, 2PL e para o Nominal em comparação com o modelo Aninhado, pois temos como hipótese inicial que pode ser o modelo que melhor se adéqua na modelagem das alternativas. Assim, podemos avaliar qual deles pode ser mais informativo em relação ao traço latente proficiência em leitura.

**Figura 10** – Comparação das Curvas de Informação do Teste (CIT). Modelos Rasch, 2PL, Nominal em comparação com o Aninhado



Fonte: elaboração própria.

Para se compreender o que está exposto na Figura 10, a linha preta em cada um dos gráficos representa a CIT do modelo Aninhado e a linha vermelha se refere ao modelo que dá título a cada gráfico. Quando comparamos as CIT de todos os modelos, podemos observar o quanto o modelo Aninhado é mais informativo que os modelos de Rasch e 2PL, porém com maior destaque para a diferença de informação em relação ao de Rasch. Entretanto, em comparação ao modelo Nominal, o Aninhado apresenta a mesma quantidade de informação, com um leve deslocamento positivo na escala de proficiência. Essas comparações nos mostram que, com os modelos que levam em consideração a análise das alternativas, o instrumento apresenta um ganho de informação, que pode contribuir também, como hipótese, para um ganho de informação pedagógica.

Com base na Figura 10, onde é evidente que os dois modelos politômicos que analisam as alternativas dos itens são mais informativos, continuamos com a mesma necessidade de definir qual dos dois modelos utilizar na implementação do TAI-Leitura. Analisando o que ambos os modelos podem nos fornecer de informações, não só no teste geral, mas de cada item, compreendemos que o modelo Nominal pode ter uma limitação ao não prever a indicação de uma das categorias como sendo a correta.

Já o modelo Aninhado tem demonstrado superar isso que estamos definindo como sendo uma limitação do Nominal, ao amoldar a alternativa considerada como correta com um modelo logístico. Ao mesmo tempo, sem impor uma ordenação a priori, que é marca do modelo Nominal. Desta maneira, entendemos que a melhor escolha para a pesquisa foi seguir na construção de um TAI-Leitura a partir da análise dos itens da PB por meio do modelo Aninhado da TRI. A seguir iremos apresentar como essa análise foi feita e quantos itens foram calibrados, além das informações gerais sobre os parâmetros dos itens.

### **2.2.2 – Análise com o modelo Aninhado: calibração dos itens que compõem o Banco de Itens do TAI-Leitura 3.0**

O processamento dos itens da PB de Leitura foi realizado em algumas etapas: primeiro, estimamos os parâmetros dos itens que compuseram ambas as aplicações de 2012 a 2015, com base nos resultados dos pré-testes obtidos junto ao Inep por meio da LAI; segundo, elaboramos uma base clone dicotômica (base fictícia) gerada a partir dos parâmetros dos 160 itens; terceiro, estimamos os parâmetros das alternativas dos itens das aplicações da PB, utilizando as bases de

dados da RME-SP, fixando a base clone como a referência. A seguir detalhamos cada uma dessas etapas.

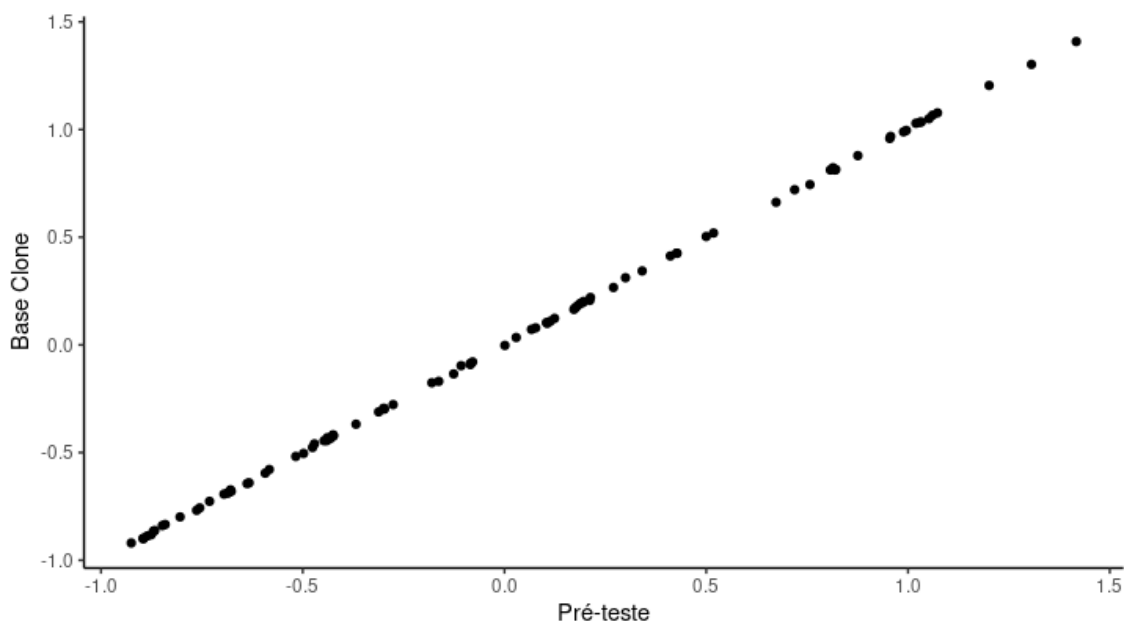
Para que fosse possível construir uma escala única, tivemos que inicialmente organizar as bases de dados dos pré-testes de 2012, 2013 e 2014, pois foram nesses pré-testes, realizados pelo Inep, que os parâmetros desses itens foram estimados. Segundo Catalani (2019, p. 55), com base em documentos elaborados por especialistas e assessores do Inep, a PB utilizava o modelo de um parâmetro logístico, com o mesmo valor de parâmetro  $a$  para todos os itens da PB.

Com base nesta informação, organizamos as bases de dados dos pré-testes e processamos com base no modelo de múltiplos grupos por meio do software *Bilog – MG*. Dessa maneira, conseguimos obter os parâmetros dos itens que haviam sido escolhidos para compor os instrumentos da PB na segunda aplicação de 2012, e nas duas aplicações de 2013, 2014 e 2015. Em posse dos parâmetros, elaboramos uma base clone dicotômica com 100 mil respondentes e selecionamos, dentre os 160 itens, aqueles que agregados tinham uma média para o parâmetro  $b$  igual ou muito próxima a zero. Assim, do total de itens anterior, ficamos com 120 itens para gerar a base clone.

As respostas fictícias foram elaboradas a partir de uma distribuição normal  $(0,1)$  com 100 mil pontos. Esses valores foram utilizados como referência de proficiência no cálculo da probabilidade de acerto ou erro, com base no modelo de um parâmetro da TRI. Dessa maneira, se a probabilidade fosse menor que cinquenta por cento, de maneira aleatória, a resposta para aquele respondente fictício seria apontada como tendo acertado ou errado o item, o que gerou vetores de respostas coerentes com o valor da proficiência utilizada como referência.

Com a base clone elaborada, realizamos um processamento utilizando o pacote *mirt*, para validar se a base conseguia reproduzir os parâmetros iniciais, pois isso garante a equalização dos instrumentos. O Gráfico 8 apresenta a distribuição dos parâmetros  $b$  estimados com o pré-teste em relação os parâmetros  $b$  a partir da base clone.



**Gráfico 8** – Distribuição dos parâmetros b dos 120 itens da PB

**Fonte:** Elaboração própria.

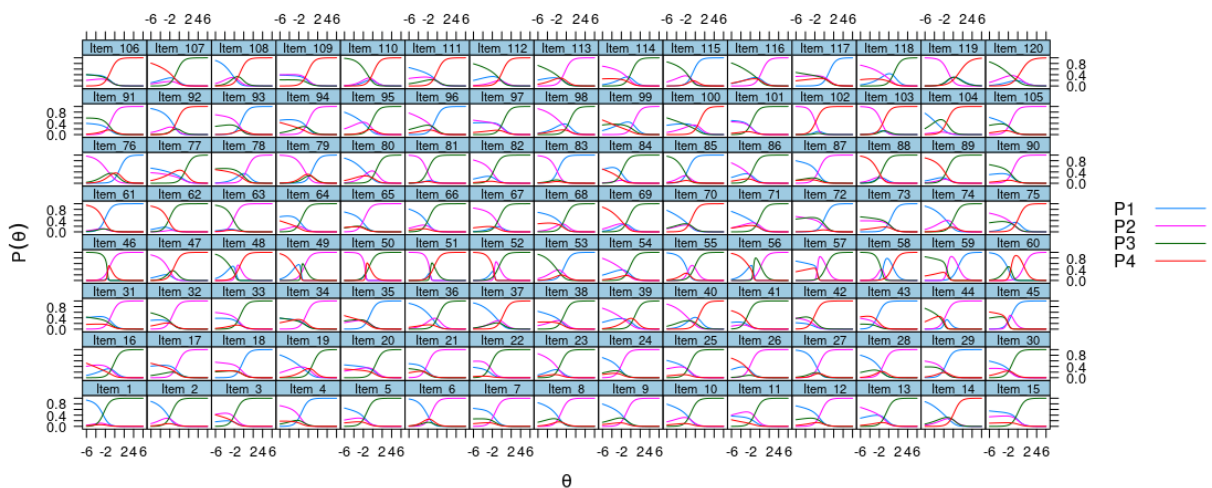
É perceptível como os parâmetros estimados a partir da base clone eram muito próximos dos parâmetros originais. Se faz importante destacar que no processamento com a base clone não houve fixação do parâmetro b, mas, dado que os parâmetros a não podiam ser diferentes, pois o modelo é de 1 PL, realizamos a fixação deles no valor igual à média dos parâmetros a originais e, ademais, eles não apresentavam diferenças significativas para o valor que foi fixado, pois esses itens já haviam sido selecionados para comporem a PB, por parte do Inep, exatamente por conta desse comportamento. Além da distribuição que o Gráfico 8 nos apresenta, também temos a Correlação de Pearson que foi igual a 0,999967, ou seja, praticamente 1 entre as duas estimativas de parâmetros b.

A partir da certeza de que a base clone conseguia manter a escala inicial, organizamos uma nova base de dados com os dados da RME-SP. Essa base foi estruturada para ser processada pela metodologia de múltiplos grupos do pacote mirt e, assim, tivemos que acrescentar a variável “grupo”, sendo a base clone representada pelo número 0 e as demais em uma sequência até o número 6. Além disso, a estrutura dos dados estava no formato longo; havia os 120 itens da base clone para usar como referência para a parte dicotômica do modelo Aninhado e os vinte itens das aplicações de 2012 e 2014 (segunda aplicação) e de 2013 e 2015 (ambas as aplicações).

Dessa maneira, tínhamos uma parte da base de dados que estava dicotomizada, inclusive esses itens de cada aplicação da PB, e outra parte que estava com as alternativas indicadas.

Esse procedimento foi necessário porque precisávamos neste momento estimar os parâmetros da parte nominal do modelo Aninhado e, sendo assim, fixamos os parâmetros da parte dicotômica para estimar os demais. A ligação da base clone com as outras bases se deu por meio dos itens que estavam dicotomizados, que tiveram que ser replicados em cada uma das bases, mas no formato categórico, incluindo o gabarito, pois a ideia foi ampliar o vetor de respostas dos alunos, separando os itens que compuseram a parte processada pelo modelo de 2PL, fixando o parâmetro  $a$ , e aqueles que serviram de base para os parâmetros do modelo Aninhado. A Figura 11 apresenta as CCI dos 120 itens para o modelo Aninhado.

**Figura 11** – Curva Característica do Item para o modelo Aninhado



**Fonte:** Elaboração própria.

O conjunto de CCI da Figura 11 nos mostra que temos algumas categorias que conseguem apresentar um comportamento semelhante ao que se espera de um item politômico, em que as categorias tenham uma certa distribuição ao longo dos valores de proficiência, ao mesmo tempo, sem que haja uma grande sobreposição. Todavia, para ampla maioria dos itens, esse comportamento esperado não ocorreu, as categorias sobrepuseram na mesma região de  $\theta$  e, em muitos casos, não apresentaram ponto de máximo dentro do intervalo analisado. De qualquer maneira, os itens tiveram ajuste interessante, pois nenhum EP ficou acima dos valores

apresentado nos processamentos anteriores. Todas as CCI dos itens podem ser visualizadas no Anexo I.

Neste sentido, definimos os itens para compor o BI do TAI-Leitura 3.0 com base no modelo Aninhado. No próximo capítulo, iremos apresentar os elementos centrais de um TAI e as características principais do TAI-Leitura 3.0.

### **3 – TAI-Leitura na perspectiva da análise das alternativas**

Este capítulo apresenta um breve histórico da elaboração e implementação do TAI-Leitura, as principais características que um TAI tem e como o TAI-Leitura 3.0 e TAI-Leitura 3.0 pseudoB se diferenciam. Neste último tópico, elaboramos dois algoritmos para as aplicações do TAI-Leitura, com o intuito de verificar as diferenças nas análises das alternativas em um TAI.

#### **3.1 – Breve histórico sobre o TAI-Leitura**

O TAI-Leitura foi desenvolvido no bojo de uma pesquisa que o Grupo de Estudo e Pesquisa em Avaliação Educacional (Gepave) realizou em parceria com a Secretaria Municipal de Educação de São Paulo no ano de 2016. Segundo Catalani (2019, p. 132), o TAI da Provinha Brasil – Leitura (TAI-Leitura) teve sua construção “planejada para ocorrer mediante um processo de aplicação experimental que viabiliza a obtenção de dados de alunos reais, pertencentes a escolas, com equipes docentes e gestoras igualmente reais”.

Nessa perspectiva, o projeto de criação do TAI-Leitura estava envolto a um processo de formação continuada das professoras e dos professores, potencializando o TAI como uma ferramenta voltada, no limite, para a avaliação do aprendizado, como relatam Alavarse, Catalani, Santos e Costa (2019). Sendo um instrumento com potencial para ser utilizado pelas professoras e pelos professores em seus processos pelos quais avaliam o aprendizado de seus estudantes. No universo dessa pesquisa, 15 escolas da Rede Municipal de Ensino de São Paulo se dispuseram a participar, propiciando, assim, ambientes com grandes possibilidades de se fazer discussões sobre avaliação educacional, avaliações externas, processos de medidas, avanço das tecnologias no ambiente escolar etc.

Ainda no âmbito do projeto, houve três aplicações de instrumentos que abarcavam itens da Provinha Brasil. Um desses instrumentos era composto por vinte itens dispostos no papel e outro que mantinha a disposição dos itens, porém utilizava como suporte para a apresentá-los aos estudantes tablet ou laptop, transpondo o teste, originalmente com impressão em papel, para o modo digital. Contudo, o formato continuava fixo, isto é, com o mesmo número de itens para cada aluno e sem que a ordem de apresentação dos itens fosse alterada. Por fim, um novo instrumento foi composto pelo próprio TAI que, por sua vez, repassa para o final de cada aplicação e para cada estudante, a constituição da estrutura do instrumento, na qual devemos considerar a quantidade de itens e a própria ordem de apresentação destes.

Nesse cenário, o projeto do TAI-Leitura em 2016 conseguiu implementar um TAI que avaliasse a proficiência em leitura dos estudantes do 1º e 2º anos, com grandes potenciais para ampliar a base de aplicação, passando a incluir todos os anos do Ensino Fundamental. Além disso, como condição para que o TAI-Leitura tivesse potencial de uso em sala de aula, foram realizadas sequências de formações em cada escola, com o intuito de apresentar todo o universo ao qual o TAI estava vinculado, como, por exemplo, a Matriz de Referência da PB, incluindo aspectos importantes sobre a compreensão da competência leitora. Ainda no âmbito das formações, outros aspectos sobre as avaliações também foram abordados, em uma tentativa de aproximar todo o ferramental do TAI à sala de aula.

A continuidade do TAI-Leitura se concretizou, a partir de 2019, com a reconstrução de todos os algoritmos estatísticos e o desenvolvimento de uma nova plataforma web para a aplicação do TAI. Sendo que a sua hospedagem, hoje, se encontra em um servidor da Universidade de São Paulo. Nesse novo formato de arquitetura de software e de hospedagem, é possível evoluir o sistema conforme as pesquisas, presentes e futuras, do Gepave em relação ao TAI, se concretizam, como também, a sua utilização em parcerias com escolas ou Redes de Ensino.

### **3.2 – As principais características do TAI-Leitura**

O TAI-Leitura, assim como outros TAI, tem por objetivo central a apresentação de itens numa sequência mais “adequada” ao estudante que o responde. Nesse mesmo sentido, Olea e Ponsoda (2002, p. 5) consideram que

[u]m Teste Adaptativo Informatizado (TAI) é uma prova construída para fins de avaliação psicológica ou educativa, cujos itens se apresentam e são respondidos mediante um computador, sendo sua característica fundamental a adaptação ao nível de competência progressivo que manifeste a pessoa.

Dessa maneira, o TAI-Leitura se diferencia da PB, sua referência pedagógica e, parcialmente, psicométrica, pois nas condições explicitadas no capítulo 1, ao invés do estudante responder vinte itens, sendo necessário um ritmo coletivo para tal, pois a professora ou o professor precisam realizar a leitura de trechos do item para que seja possível respondê-lo, ele passa a responder a quantidade que for necessária para o teste parar, algo que pode variar para cada estudante. Além da melhoria considerável da precisão da estimativa da proficiência, vinculado a um Banco de Itens (BI) que consiga cobrir de maneira satisfatória a escala de proficiência.

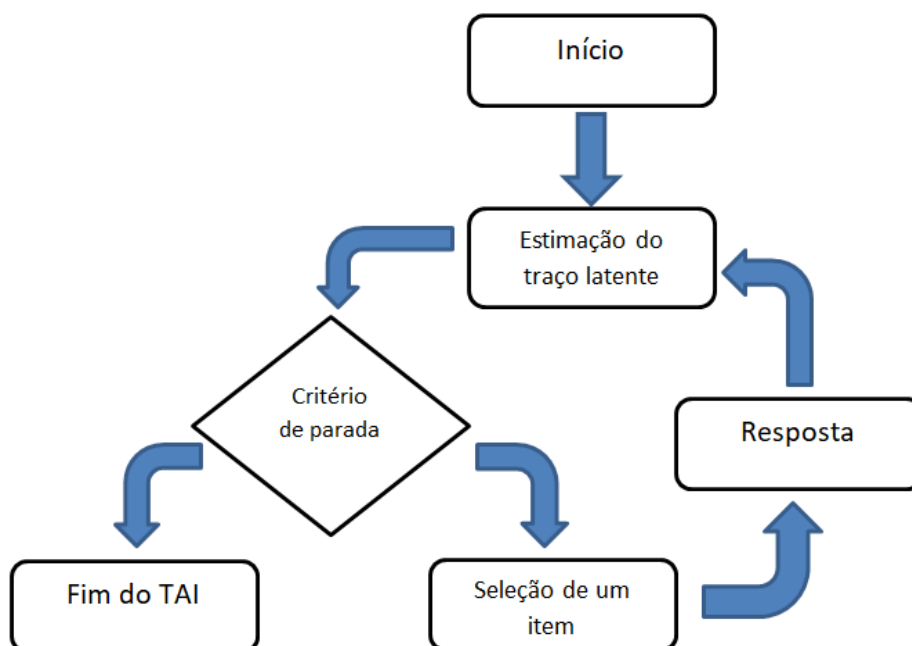
No entanto, para a constituição do TAI-Leitura é preciso mais do que um BI, sendo necessário algoritmos de partida, de seleção do próximo item e de parada do teste, além de todo o desenvolvimento do *back-end* e *front-end* do sistema. Em linha gerais, a parte que cabe ao desenvolvimento de software pode variar conforme a tecnologia disponível à época, da mesma maneira, que as preferências daqueles que estão desenvolvendo podem interferir em qual linguagem de programação será utilizada para desenvolvê-lo.

Esses elementos da informatização do teste não diminuem a importância do BI que continua a ocupar um lugar central em qualquer aplicação de TAI, pois é por meio de sua composição que a medida pode ser constituída. No TAI-Leitura os itens fazem parte do acervo histórico de aplicações da PB, em torno de 300 itens, quantitativo que se tornaria um obstáculo excessivo se fosse necessária a elaboração de cada um dos itens. Adicionalmente, sublinhamos que os parâmetros dos itens, estimados pela TRI, são condição necessária para a existência de um BI que possa compor um TAI e para que os processamentos estatísticos envolvendo respostas dadas pelos estudantes da RME-SP e aos pré-testes a fim de se estimar suas proficiências.

O BI do TAI-Leitura estava, à época, composto com 60 itens provenientes dos instrumentos de 2015, primeira e segunda aplicações, e de 2016, segunda aplicação. Os parâmetros até então utilizados foram obtidos por meio do modelo de 1 parâmetro, porém, no bojo dessa pesquisa, outros modelos paralelos foram incorporados e o BI agora também apresenta parâmetros do modelo Aninhado. Todavia, somente o Banco de Itens do TAI-Leitura 3.0 contém os itens com a nova metodologia de estimação dos parâmetros. Nesse sentido, a continuidade da pesquisa se baseou no BI para o novo TAI, o TAI-Leitura 3.0.

Os algoritmos são um conjunto de procedimentos estatísticos que garantem que o teste seja transformado num teste adaptativo, potencializando a importância do BI; em seu desenvolvimento ulterior pudemos chegar ao TAI-Leitura 3.0. Para que possamos compreender como esse conjunto atua, devemos pensar no fluxo que um TAI tem e para tanto, com apoio em Barrada (2012, p. 290), na Figura 12 apresentamos um diagrama que resume o fluxo que minimamente precisa ser seguido.

**Figura 12** – Diagrama do fluxo de um Teste Adaptativo



Fonte: BARRADA (2012).

Com base no fluxo da Figura 12 podemos observar alguns momentos em que existe uma tomada de decisão e o processo ao qual o respondente está submetido é alterado. Nessa linha, podemos destacar o “Início”, o “Critério de parada” e a “Seleção de um item”, como sendo os momentos em que tais tomadas de decisões ocorrem e de fato alteram o fluxo do respondente dentro do TAI. De maneira um pouco mais detalhada, se nos determos em cada um destes momento poderemos compreender melhor os motivos dessa ilação.

O “Início” do TAI se assemelha em muito ao algoritmo de “Seleção de um item” ou à “Estimação do traço latente”, porém sem guardar uma estrita relação com o significado que a palavra “estimativa” apresenta quando nos referimos a ela no bojo do momento de estimação do traço latente, pois no início do teste a maior preocupação é com o primeiro item a ser apresentado e nem tanto com a “verdadeira” proficiência do respondente.

Dessa maneira, o “Início” do TAI pode se valer de várias informações para apresentar o primeiro item, desde resgatar a última estimativa de proficiência do respondente ou utilizar uma outra medida como, no caso do TAI-Leitura, com estudantes associados a uma turma de uma escola, a média da sua turma ou da escola, ou o Nível Socioeconômico do estudante ou da escola. No TAI-Leitura 3.0, mantivemos o critério que define como priori a média de proficiência da escala

da PB que, teoricamente, varia de 0 a 1000, ou seja, o item inicial selecionado como padrão tinha dificuldade próxima do valor 500 e com maior informação neste ponto.

Essa decisão de utilizar o item mais ao centro da escala pode gerar problemas com um efeito conhecido como “viés de exposição do item”, que tem como significado a quantidade de vezes que um item é exposto aos respondentes, logo, influenciando na independência local, pois passa a não ser somente a medida do traço latente o fator predominante no momento que o estudante dá a resposta, mas um certo conhecimento prévio de qual poderia ser a resposta correta do item. Como perspectiva, temos outras possibilidades de filtros para apresentar o primeiro item, como, por exemplo, o conteúdo dos itens relacionado ao ano escolar do estudante que vai responder o TAI.

O algoritmo de “Seleção de um item” pode variar quanto ao método estatístico que utiliza para obter qual é o melhor item para ser apresentado. Neste caso, o termo “melhor” está relacionado com a maior informação que podemos obter, quando cotejamos parâmetros do item e a proficiência estimada. Sendo assim, a seleção do próximo item vai depender de quanta informação um determinado item tem quando relacionado com a proficiência estimada enquanto o estudante realiza o teste. Existem, na literatura, diversos métodos estatísticos que poderiam ser utilizados, no caso do TAI- Leitura 3.0 a Máxima Informação de Fisher, ou simplesmente Informação de Fisher, foi a escolhida para a seleção do próximo item. Para mais informações sobre esse e outros métodos estatísticos relacionados ao TAI, é possível consultar alguns escritos sobre o tema, como se encontra em Barrada (2012), Chang (2014), Costa (2009), Magis, Yan e von Davier (2017), Olea e Ponsoda (2002) e Olea, Ponsoda e Prieto (1999).

Ainda sobre o algoritmo de seleção do próximo item, mas que não consta no TAI-Leitura, é possível que os filtros mencionados anteriormente sejam incorporados ao algoritmo, restringindo assim os itens que podem ser utilizados para a maximização da informação. Essa interferência na constituição do BI, durante a aplicação do TAI, pode suscitar questionamentos, pois é sabido que dispormos de uma maior quantidade de itens é fundamental para obtermos um significativo sucesso na aplicação de um TAI, no entanto, neste caso, nos parece que tal restrição pode ampliar a eficiência pedagógica do teste, incluindo a possibilidade de cobrir uma quantidade maior de conhecimentos relacionados ao traço latente medido, dirimindo, em certa monta, as possíveis críticas em relação a diminuição da quantidade de itens que o respondente vai ser submetido.



O terceiro momento, “Critério de parada”, configura-se como um algoritmo que pretende definir os critérios pelos quais o teste vai parar de ser aplicado ao estudante. Em um teste no formato de papel, esse algoritmo de parada está relacionado para com a quantidade de itens e o tempo de aplicação do instrumento, e, assim, não deve causar nenhuma estranheza existir critérios para que a aplicação de um teste seja finalizada. No caso do TAI-Leitura 3.0, esse critério está relacionado, principalmente, segundo Alavarse, Catalani, Meneghetti e Travitzki (2017, p. 75), com um critério misto, a saber:

a) número de itens do teste (mínimo de 8 e máximo de 20 itens); b) limite permitido de incerteza (Erro Padrão menor que 35) na estimação da proficiência; e c) confiabilidade do nível de proficiência, tendo sido considerado confiável quando o intervalo de confiança da estimação da proficiência corresponde a um único nível de proficiência da Provinha Brasil (PB).

Dessa maneira, o TAI-Leitura 3.0 tem sua parada, em uma aplicação, assim que a estimativa de proficiência mais o erro de medida ponderado por um valor crítico (intervalo de confiança), estiverem incluídos em um nível de proficiência, sendo que o TAI-Leitura manteve os cinco níveis da PB. Outros critérios de parada podem ser a estabilização do erro de medida entorno de um valor definido a priori (erro padrão) ou a definição de um valor mínimo ou máximo de itens aplicados.

Em linhas gerais, essas são as principais características do TAI-Leitura, que não se diferencia, grosso modo, daquilo que tem se encontrado na literatura sobre o tema, porém, sem desconsiderar novos desdobramento a serem pesquisados e incorporados, com sua principal inovação no critério de parada, justamente para se converter numa ferramenta que possa ser utilizada na sala de aula, onde a interpretação pedagógica é um elemento fulcral na perspectiva de uso formativo dos resultados do TAI.

### **3.3 – TAI-Leitura 3.0 e pseudoB: as principais diferenças desses dois testes**

Pelo que foi apresentado nos subcapítulos anteriores, o desenvolvimento do TAI-Leitura, processo ao qual a pesquisa se integrou, decorre do esforço de pesquisadoras e pesquisadores do Gepave. Nessa perspectiva, a ideia de estimar a proficiência dos respondentes com base nas respostas a quaisquer alternativas, se aprofundou com a definição do modelo Aninhado da TRI

como sendo a base para o BI do TAI, e avançamos na compreensão sobre quais poderiam ser as vantagens dessa escolha em contribuir com o uso do TAI em sala de aula.

Os algoritmos do TAI-Leitura 3.0 foram implementados e testados em simulações, como detalhado no capítulo 4, e percebemos que existia a possibilidade de obtermos informações da relação que a CCI do item nos proporcionava, mais especificamente em análises politômicas, como é o caso do modelo Aninhado, em que temos curvas para cada uma das alternativas que, sendo assim, exprimindo uma relação com as probabilidades de cada uma das alternativas ser assinalada em função da proficiência do respondente. Conseqüentemente, levantamos a hipótese de que poderíamos definir a dificuldade de cada uma das alternativas com base nessas relações das categorias.

Ao analisarmos as CCI de todos os itens do BI, identificamos que as projeções dos pontos de intersecção das curvas das alternativas consideradas como incorretas com a curva do gabarito no eixo das proficiências (abscissa), geravam um valor menor que o próprio parâmetro  $b$  estimado. Destarte, identificamos que em todos os itens esse mesmo fenômeno ocorria; sendo assim, realizamos essas projeções para todos os itens e estimamos os valores dos parâmetros de “dificuldade” das alternativas incorretas que, por sua vez, intitulamos de “pseudoB”.

Com a possibilidade de utilizarmos esse pseudoB para a dificuldade de cada uma das alternativas, realizamos adaptações nos algoritmos já existentes, porém sem perder o que já havíamos desenvolvido. Dentre as diversas modificações que buscamos implementar, a que mais se apresentou viável foi a utilização do pseudoB no algoritmo de escolha do próximo item. Assim, após o TAI-Leitura 3.0 ter estimado a proficiência e o erro padrão, verificado que essas medidas não se enquadram dentro do critério de parada, o algoritmo que determina o próximo item utiliza os pseudoB para calcular a máxima informação e verificar qual item maximiza a informação no ponto que representa a proficiência estimada.

Em testes iniciais esse mecanismo se demonstrou interessante e com poucas diferenças para o algoritmo inicial do TAI-Leitura 3.0. Então, resolvemos implementar no sistema do TAI esses dois testes e, de maneira moderada, sem considerar como iniciativas com pesos iguais, decidimos realizar as simulações e utilizar ambas no pré-teste e na aplicação geral do TAI, porém com uma quantidade menor de respondentes para o TAI com o pseudoB, pois, em nosso entendimento, esse algoritmo ainda necessitava de mais estudos e testes.

## **4 – Análise das simulações e dos pré-testes do TAI-Leitura 3.0**

A validação dos algoritmos desenvolvidos pode ser feita, em um primeiro momento, por meio de simulações que, por sua vez, nos permitem ampliar a quantidade de respondentes e diferenciar facilmente os tipos de amostras submetidas ao algoritmo. Entretanto, somente as simulações não são suficientes para desvelar as potencialidades e os limites que os algoritmos para o TAI-Leitura 3.0 podem apresentar. Dessa maneira, além das simulações, realizamos dois pré-testes em escolas parceiras no projeto do Gepave.

Mediante os testes dos algoritmos com simulações e aplicações com alunos dessas escolas, cujos resultados foram analisados e serão apresentados neste capítulo. Todavia, é importante salientar que o mais importante, como decorrência desses pré-testes, não são os resultados dos alunos, ainda que sejam relevantes, mas como o TAI-Leitura 3.0 se comportou nesses diferentes cenários, especialmente em face do objetivo principal da pesquisa, como utilizar quaisquer alternativas assinaladas para a estimativa da proficiência, sem desconsiderar a compreensão dos erros cometidos ao longo do teste, isto a marcação das alternativas consideradas incorretas, que nesta pesquisa assumem a característica de serem aquelas com menor parâmetro de dificuldade em relação àquela considerada como gabarito. Evidentemente, continua-se com a preocupação com a classificação em determinado nível de proficiência, sobretudo, pelo uso pedagógico que isso possui.

Este capítulo está dividido nos seguintes tópicos: A estrutura da base simulada e resultados descritivos das aplicações; Informações gerais e os resultados descritivos sobre o pré-teste nas escolas; Análise comparativa das aplicações: o que podemos concluir? Compreendemos que esses tópicos contribuem com a compreensão do objeto pesquisado, destarte, podemos justificar as escolhas realizadas para a aplicação geral do TAI-Leitura 3.0.

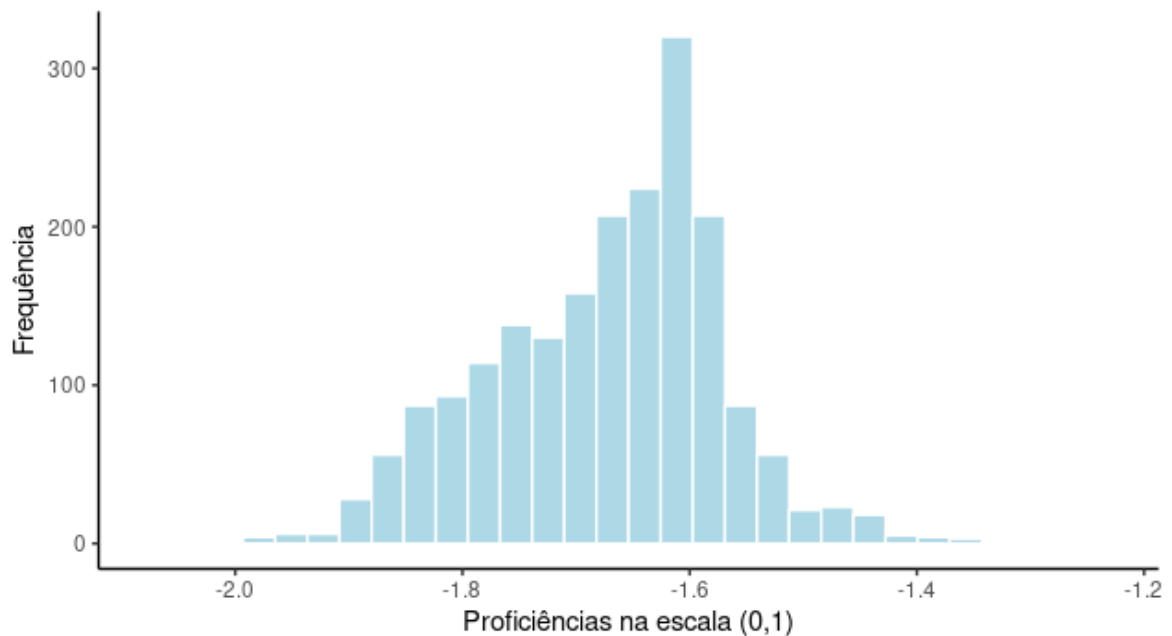
### **4.1 – A estrutura da base simulada e resultados descritivos das aplicações**

Para realizarmos as simulações, precisamos de uma base de dados que contenha respostas aos itens e possam ser diferenciadas entre cada respondente simulado. Além dessa diferenciação, o fato de estarmos utilizando um algoritmo que modela dados politômicos nos obriga a utilizar respostas que tenham categorias, ou melhor, que não sejam somente consideradas como corretas ou incorretas. Sendo assim, a simulação precisa modelar as categorias com base nos parâmetros dos 120 itens e uma distribuição de proficiências que consiga trazer variabilidade para a

estimativa dos parâmetros. Essa distribuição, em geral, tem como média o valor 0 e desvio-padrão igual a 1, sendo conhecida por distribuição normal (0,1) e, teoricamente, varia de menos infinito ( $-\infty$ ) a mais infinito ( $+\infty$ ), mas é comum variar entre -4 e 4 desvios-padrão, com distâncias definidas pela quantidade de pontos de quadratura que, por sua vez, representam os pontos em que as estimativas serão ancoradas.

A simulação da base de dados foi feita com a função *simdata* do pacote *mirt*, que permite, conforme apresenta Chalmers (2012, p. 9 e 28), com base nos parâmetros dos itens e a indicação da média e desvio-padrão como referência da população na modelagem, simular uma grande quantidade de dados. No Gráfico 9 temos a distribuição das proficiências estimadas como a função *fscore* do pacote *mirt*, para os dados da base de respostas simuladas.

**Gráfico 9** – Distribuição das proficiências estimadas por meio do pacote *mirt*



**Fonte:** Elaboração própria.

Com base na distribuição que podemos observar no Gráfico 9, identificamos que existe um deslocamento nas proficiências para uma região do eixo das abscissas menor que zero devido aos parâmetros de “dificuldade” dos itens estarem localizados, em sua grande maioria, abaixo do valor -1. Dessa maneira, temos uma simulação de dados que se concentrou entorno de uma média igual a -1,6735 e com caudas indo de -2,5 a -1, o que nos permitiu utilizar os vetores de

respostas para testar os algoritmos do TAI-Leitura 3.0, pois estavam muito coerente com a distribuição dos parâmetros  $b$  dos itens do BI.

Essa distribuição de proficiências com a base simulada, nos permitiu classificar cada uma delas nos níveis que eram utilizados na PB. A Tabela 5 apresenta a distribuição das proficiências nos níveis.

**Tabela 5** – Informações gerais sobre a análise estatística da base simulada

Níveis	Quantidade de respondente	Percentual de respondente	Média Proficiêncica	Média do Erro Padrão
1	482	24,1%	318,4	10,6
2	1518	75,9%	337,1	8,9
3				
4				
Total	1999			

**Fonte:** Elaboração própria.

É possível verificar na Tabela 5 que, como já identificamos no histograma do Gráfico 9, as proficiências estão concentradas nos Níveis 1 e 2, que correspondem aos intervalos de proficiências com valores, respectivamente, menores que -1,75 e entre -1,75 e -0,75. Se realizarmos a transformação da escala (0,1) para uma escala com média 500 e desvio-padrão 100, teremos um intervalo para os Níveis 1 e 2, com valores menores que 325 e entre 325 e 425. Além dessa informação, também podemos perceber que as médias de proficiências dos níveis ficaram próximas entre si, com uma diferença na ordem dos 18 pontos. Entretanto, com base em um teste-t de *Student*, com 1098 graus de liberdade e valor de  $p$  menor que 0,05, podemos afirmar que os dois níveis contêm valores de proficiências estatisticamente diferentes.

Com a base de dados simulada e verificada todas as informações descritivas que seus dados contêm, realizamos as simulações utilizando os dois algoritmos elaborados para modelar as alternativas em um TAI. Como já apresentado no capítulo 3, os algoritmos TAI-Leitura 3.0 e TAI-Leitura 3.0 – PseudoB, estão baseados no mesmo modelo da TRI, o modelo aninhado, porém se diferenciam pela forma como selecionam o próximo item a ser apresentado ao respondente, antes do final do teste. O primeiro algoritmo utiliza a estimativa de proficiência, obtida a cada item respondido, como ponto de ancoragem da escolha do item. Já o segundo

algoritmo executa o mesmo procedimento, porém apoiado no pseudoB da alternativa escolhida pelo respondente e esse parâmetro é utilizado como “estimativa da proficiência” do respondente naquele momento do teste.

Para ambas as simulações, os critérios de parada não foram modificados; logo, a única diferença foi a metodologia utilizada para escolher o próximo item.

A Tabela 6 apresenta a distribuição dos dados simulados em relação aos níveis e as médias de proficiência em cada nível e relaciona as estimativas de proficiências do algoritmo com as da função *fscore* do pacote *mirt*.

**Tabela 6** – Informações gerais sobre a simulação com o TAI-Leitura 3.0

Níveis	Quantidade de respondente	Percentual de respondente	Média Proficiência	Média do Erro Padrão	Quantidade de respondente - Mirt	Percentual de respondente - Mirt	Média Proficiência - Mirt	Média do Erro Padrão - mirt
1	1422	71,1%	263,9	40,7	1150	71,1%	297,5	34,5
2	507	25,4%	350,7	21,9	827	25,4%	351,9	34
3	64	3,2%	462,7	31,2	22	3,2%	444,4	35,6
4	6	0,3%	584,6	34,7		0,3%		
Total	1999				1999			

**Fonte:** Elaboração própria.

A partir das informações que encontramos na Tabela 6, fica evidente que a utilização do TAI-Leitura 3.0 modifica, ainda que parcialmente, os resultados que já havíamos encontrado ao processar a base simulada. Confrontando com as informações da Tabela 5, percebemos que apareceram, mesmo que em menor magnitude, dados simulados nos Níveis 3 e 4, com valores para as médias de proficiências iguais a, respectivamente, 462,7 e 584,6. Para além do aumento dos níveis contidos nos resultados da base simulada, podemos afirmar que o TAI-Leitura 3.0 conseguiu diferenciar os respectivos vetores (respostas simuladas) apresentados em estimativas de proficiências diferentes e, por consequência, níveis distintos.

Entretanto, quando comparamos os dados das estimativas de proficiências do TAI-Leitura 3.0 e do pacote *mirt* para os mesmos vetores de respostas, podemos destacar pelo menos dois pontos. O primeiro é que as médias de proficiências para cada nível têm diferenças, na maioria dos casos, que não são superiores a 20 pontos de proficiência; somente no Nível 1 é que temos uma diferença que supera esse valor. O segundo é o fato de não termos, nas estimativas do pacote *mirt*, nenhum dado simulado que venha a ser classificado no Nível 5, o que nos sugere

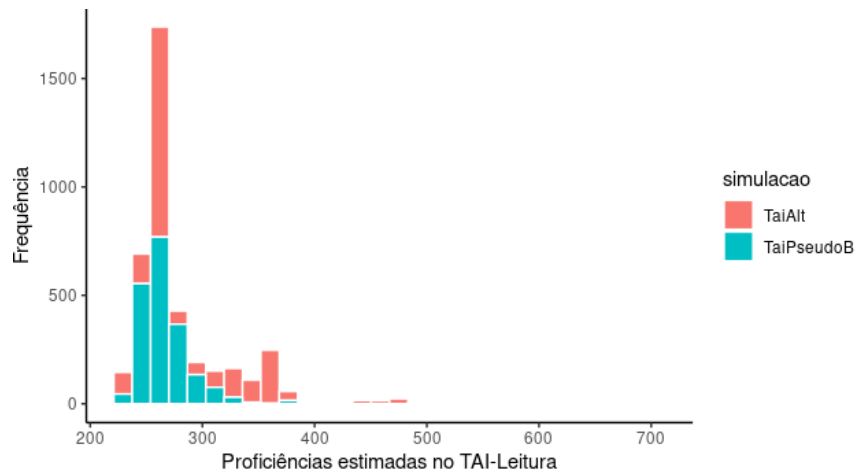
que os resultados de distribuição dos dados simulados nos níveis tenderam a ficar mais próximos do processamento geral da base simulada. Já os resultados para o TAI-Leitura 3.0 – pseudoB, são apresentados na Tabela 7.

**Tabela 7** – Informações sobre a simulação do TAI-Leitura 3.0 pseudoB

Níveis	Quantidade de respondente	Percentual de respondente	Média Proficiência	Média do Erro Padrão	Quantidade de respondente - Mirt	Percentual de respondente - Mirt	Média Proficiência Mirt	Média Erro Padrão Mirt
1	1958	97,9%	264,5	40,3	108	5,4%	304,6	43,4
2	41	2,1%	349,7	30,9	1269	63,5%	381,8	38,5
3					612	30,6%	451,5	42,1
4					10	0,5%	546,4	48,3
Total	1999				1999			

**Fonte:** Elaboração própria.

É possível identificar que existe uma grande diferença entre os resultados dos dois algoritmos, principalmente por aumentar a concentração de resultados simulados nos Níveis 1 e 2, além de não haver nenhum resultado nos demais níveis. Em relação ao comparativo com as estimativas do pacote mirt, temos uma diferença em relação ao que encontramos na Tabela 6, pois passamos a ter resultados simulados no Nível 4, porém continuamos sem nenhum resultado simulado no Nível 5. Entretanto, temos, quando comparamos os resultados das simulações dos dois algoritmos, médias para os Níveis 1 e 2 muito próximas, com diferenças de no máximo 1 ponto de proficiência. O Gráfico 10 distingue a diferença entre os resultados das duas simulações.

**Gráfico 10** – TAI-Leitura. Distribuição das proficiências estimadas

**Fonte:** Elaboração própria.

Além da possibilidade de distinção de resultados, conseguimos perceber que aqueles do TAI-PseudoB apresentam uma distribuição mais próxima do que encontramos no processamento geral, observado no Gráfico 9. Entretanto, mesmo com distribuições diferentes, podemos notar que a maior concentração em ambas as análises está na região de 250 a 260, logo, não diferenciando muito do que temos no processamento geral.

Ao avançarmos as análises dos resultados das simulações no que tange à quantidade de itens apresentados para os respondentes situados em cada nível, na Tabela 8 encontramos a distribuição da quantidade de itens por nível.

**Tabela 8** – Distribuição da quantidade de itens por nível em ambas as simulações

Nível	Média de itens por simulação			
	TAI-Alt	TAI-Alt/Mirt	TAI-pseudoB	TAI-pseudoB/Mirt
1	12,4	13,6	13,6	12,5
2	15,5	13	16,5	13,8
3	19,2	16,7		13,4
4	20			9,7
5				
Total	13,4	13,4	13,6	13,6

**Fonte:** Elaboração própria.

A análise da Tabela 8, entre outras informações, indica uma diferença pequena entre os Níveis 1 e 2 que pode, muito provavelmente, ser ocasionada por serem os níveis com maior concentração em todas as análises. Contudo, para os demais níveis temos diferenças maiores,



sendo a maior delas em relação ao Nível 4 que, para o TAI-Leitura 3.0, chegou a 20 itens aplicados em média e, para o TAI-Leitura pseudoB/*mirt*, com o mesmo algoritmo do TAI-pseudoB, chegou a 9,7 itens.

Todas as informações apresentadas neste tópico nos levam a acreditar que as simulações dos dois algoritmos foram bem-sucedidas, pois conseguimos verificar diferenças nas estimativas de proficiências para cada simulação, cada uma delas com um algoritmo próprio. Temos, também, por conta dos percursos diferentes, diferenças entre as simulações, o que demonstraria a capacidade de analisar cada alternativa assinalada, o que indicaria uma medição mais precisa do objeto do teste, a proficiência em leitura. Nesse sentido, avançamos para analisar os resultados dos pré-testes que foram aplicados utilizando ambos os algoritmos.

#### **4.2 – O pré-teste nas escolas e os resultados descritivos das aplicações**

A continuidade da pesquisa, diante da necessidade de validarmos o TAI-Leitura 3.0 e o TAI-Leitura 3.0 pseudoB, deu-se com a realização de um pré-teste com alunos, sem nenhum tipo de simulação, pois somente dessa maneira poderíamos submeter ambos os algoritmos ou, pelo menos, um deles ao processo de aplicação com um número maior de alunos. Dessa maneira, conseguimos mobilizar, junto a escolas parceiras do Gepave, alunos do 2º ao 5º anos para participarem da aplicação dos dois algoritmos do TAI.

Todavia, devemos destacar que não houve nenhum tipo de amostragem ou sorteio, pois os algoritmos ficaram finalizados para entrarem em produção<sup>17</sup> em momentos diferentes. Sendo assim, em uma das escolas, local com a maior concentração de alunos, somente o TAI-Leitura 3.0 foi aplicado. Após 10 dias da primeira aplicação do pré-teste, conseguimos aplicar o TAI-Leitura 3.0 pseudoB em duas turmas de 2º e 3º ano, em outras duas escolas.

A Tabela 9 nos apresenta os dados da aplicação na escola denominada Zimbábue, em que temos o maior quantitativo de alunos que realizaram o pré-teste.

---

<sup>17</sup> Versão de um sistema/plataforma/site que está acessível para qualquer usuário da internet. Porém, caso seja necessário ter usuário e senha para acessá-lo, somente a página inicial estará disponível.

**Tabela 9** – TAI-Leitura 3.0. Aplicação na Escola Zimbábue. Março 2023

Níveis	Quantidade de respondente	Percentual de respondente	Média Proficiência	Média do Erro Padrão
1	9	4,1%	298,2	27,7
2	40	18,4%	364,2	23,1
3	50	23,0%	467,4	30,4
4	23	10,6%	555,3	29,8
5	95	43,8%	647,3	42,1
Total	217			

**Fonte:** Elaboração própria.

Conseguimos observar na Tabela 9 que o TAI-Leitura 3.0 é capaz de medir a proficiência em leitura dos alunos que estejam localizados no Nível 5; logo, nas simulações de fato não existiam dados simulados que poderiam estar localizados no maior nível. Destarte, temos as médias aumentando conforme avançamos o nível, além do erro padrão ser maior do que 40 somente no Nível 5, muito provavelmente, porque o Banco de Itens (BI) está limitado na sua parte superior, em relação ao parâmetro b. Com essa informação sobre uma possível limitação do BI, a Tabela 10 traz as informações sobre a média de itens respondidos por alunos classificados em cada nível.

**Tabela 10** – TAI-Leitura 3.0. Média de itens respondidos por nível. Escola Zimbábue. Março 2023

Nível	Quantidade de respondente	Média de itens
1	9	19,4
2	40	16,2
3	50	19,7
4	23	20
5	95	13,1
Total	217	16,2

**Fonte:** Elaboração própria.

Inicialmente, seria possível supor que, ao considerarmos as alternativas escolhidas por cada aluno, haveria uma diminuição da quantidade de itens, pois o fato de termos mais informação do que o modelo dicotômico, faria com que a estimativa de proficiência demandasse menos itens para sua finalização. Entretanto, o que encontramos na aplicação do pré-teste foi o contrário, um aumento na quantidade de itens, principalmente nos Níveis 3 e 4. Provavelmente, esse aumento de itens esteja relacionado com o aumento de informação coletada, pois o fato de

considerar a trajetória de respostas amplia a “certeza” de o teste parar em um nível o que, em muitos casos, amplia a quantidade de itens aplicados. De qualquer maneira, o aumento na quantidade de itens não implicou medidas piores; revela, tudo indica, a necessidade de ampliar a variedade de itens que compõem o BI.

A aplicação do TAI-Leitura 3.0 pseudoB teve uma quantidade menor de respondentes. Como já explicado, houve uma diferença nas datas de finalização de cada algoritmo, o que fez com que a entrada em produção ocorresse somente após a aplicação do pré-teste na escola Zimbábue. De qualquer maneira, 32 alunos realizaram o pré-teste, sendo alunos do 2º e 3º anos. A Tabela 11 apresenta a distribuição destes alunos nos níveis e as médias de proficiência.

**Tabela 11** – TAI-Leitura pseudoB. Aplicação em Escolas parceiras. Março 2023

Níveis	Quantidade de respondente	Percentual de respondente	Média Proficiência	Média do Erro Padrão
1	5	15,6%	290,2	34,2
2	9	28,1%	383,8	32,2
3	7	21,9%	460,5	37,7
4	6	18,8%	547,1	41,8
5	5	15,6%	630,2	43,2
Total	32			

**Fonte:** Elaboração própria.

Podemos identificar que mesmo com menos alunos participando do pré-teste, o TAI-Leitura 3.0 pseudoB consegue realizar uma distribuição dos alunos ao longo dos níveis, o que nos permite afirmar que o algoritmo não limitou a disponibilidade de itens conforme o teste avançava. Por serem escolas e algoritmos diferentes, a comparação não pode ser direta, pois não temos nenhum tipo de variável que pudesse servir como referência, porém, analisando as Tabelas 6 e 7, podemos perceber que existe uma coerência nas médias de cada nível, além da concentração de alunos nos níveis. A Tabela 12 traz a média de itens apresentados por nível, sendo possível analisarmos se houve aumento ou diminuição na quantidade de itens com a mudança de algoritmo.

**Tabela 12** – TAI-Leitura pseudoB. Média dos itens respondidos. Escolas parceiras. Março 2023

Nível	Quantidade de respondente	Média de itens
1	5	17,4
2	9	18,8
3	7	20
4	6	20
5	5	17,4
Total	32	18,8

**Fonte:** Elaboração própria.

Na aplicação do TAI-Leitura 3.0 pseudoB tivemos, no geral, um aumento na quantidade de itens administrados em relação ao TAI-Leitura 3.0, sendo, com base na informação contida na Tabela 12, os Níveis 3 e 4 aqueles para os quais os alunos aí alocados mais responderam itens ao longo do teste. Além disso, é importante destacar que o Nível 4 tem um intervalo menor que os demais níveis e, logo, é compreensível este aumento de itens, pois, ao se considerar as alternativas, a variação da quantidade de itens fica maior, sendo que a cada resposta um novo caminho pode ser iniciado, o que pode garantir maior precisão para o TAI. De qualquer maneira, ambas as aplicações demonstram que os algoritmos de análise das alternativas não funcionaram de maneira inesperada, o que já traz credibilidade para uma aplicação maior. No próximo tópico, são apresentados e analisados os resultados numa perspectiva das trajetórias de cada nível.

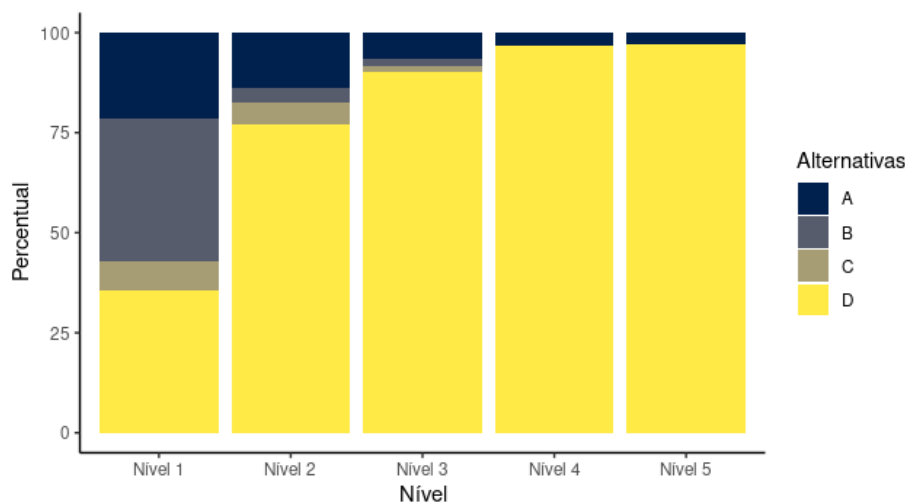
#### **4.3 – Análise comparativa das aplicações: o que podemos concluir?**

As aplicações dos pré-testes nos deram a oportunidade de avaliar se a ideia de modelar as alternativas em um teste, na perspectiva de conseguirmos implementar essa modelagem em um TAI, foi bem-sucedida. À primeira vista, com base nas informações detalhadas até o momento, temos fortes indícios de que os algoritmos desenvolvidos conseguem diferenciar cada nível pela trajetória que cada estudante, com conhecimento descrito pelo nível, realiza ao longo da aplicação do TAI.

Nesse sentido, vamos analisar as duas aplicações conjuntamente, pois, assim, diminuimos os impactos que as diferenças entre escolas podem ter, principalmente, as diferenças quantitativas. A ideia é realizarmos uma análise das marcações às alternativas dos dois primeiros itens, incluindo o item de exemplo, pois são iguais para todos os alunos que responderam ao TAI. A

seguir vamos cotejar as trajetórias que os alunos classificados tiveram, em relação ao pseudoB, dentro do teste. O Gráfico 11 apresenta as marcações por alternativa agregadas por nível.

**Gráfico 11** – TAI-Leitura 3.0. Distribuição das escolhas por alternativa no item exemplo



**Fonte:** Elaboração própria.

O primeiro item do TAI-Leitura<sup>18</sup> é um exemplo para que os alunos compreendam como devem proceder para responder a cada item e passar para o próximo. Esse item, ao ser respondido, como parte do processo de compreensão do procedimento, não tem a resposta considerada para fins de estimativa de proficiência, mas essa resposta fica registrada no banco de dados e pode ser analisada em conjunto com as respostas de todos os respondentes de uma sessão de aplicação do TAI-Leitura. O Gráfico 11 apresenta o percentual de marcação por alternativa para cada nível no TAI-Leitura 3.0 e, de maneira geral, podemos perceber que a alternativa “D” (onde está escrita a palavra “BALÃO”), gabarito do item, foi a mais escolhida em todos os níveis.

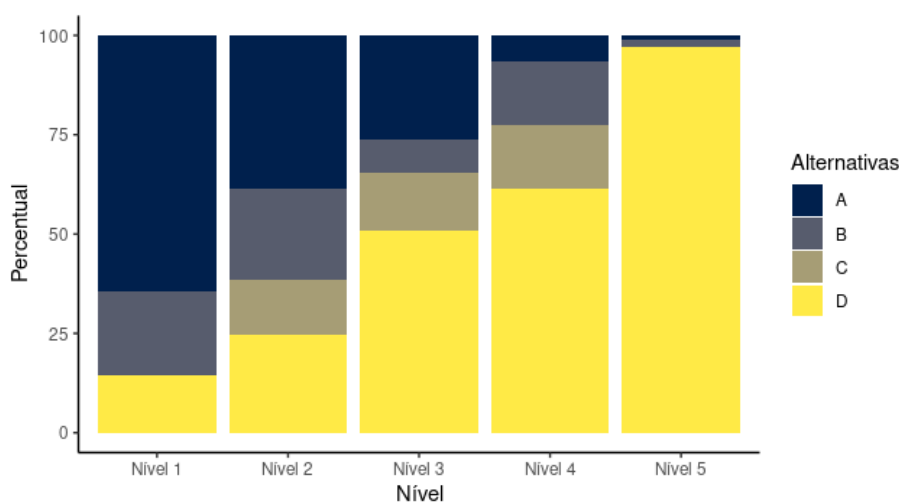
Entretanto, essa escolha não foi unânime, mesmo no Nível 5. Sendo assim, podemos compreender que houve, pelo menos, duas possibilidades de escolha, principalmente, no maior nível: ou a alternativa “A” (onde está escrita a palavra “PILÃO”) exerce uma certa atração sobre os alunos, mesmo aqueles que estão em níveis mais avançados, ou os alunos compreendem que se trata de uma ‘questão’ exemplo e respondem a primeira alternativa para avançar rapidamente

<sup>18</sup> Para esse capítulo considere TAI-Leitura a expressão para os TAI-Leitura 3.0 e TAI-Leitura 3.0 PseudoB.

para o próximo item. De qualquer maneira, conseguimos observar que os alunos que estão no Nível 1 acabam buscando escolher outras alternativas, principalmente, a alternativa “B” (onde está escrita a palavra “VILÃO”). Nos demais níveis, temos uma diminuição rápida da quantidade de marcações nas alternativas consideradas incorretas e um aumento das marcações na correta.

Após os alunos responderem ao item exemplo, é apresentado um item comum a todos os respondentes, como já foi explicado no capítulo 3, item escolhido com base na informação de uma “estimativa” de proficiência prévia. Neste momento, as aspas na palavra estimativa indicam que estamos usando um valor fixo para todos os respondentes do TAI-Leitura, independente da proficiência que de fato o estudante tenha no momento da aplicação, o primeiro item virá em uma determinada região da escala. O Gráfico 12 apresenta as informações sobre as marcações às alternativas do item 1, o item comum.

**Gráfico 12** – TAI-Leitura 3.0. Distribuição das escolhas por alternativa no item 1



**Fonte:** Elaboração própria.

Os dados representados no Gráfico 12 nos mostram que aquilo que observamos no item exemplo ocorre no item que de fato estava “valendo” para o teste. De modo geral, a alternativa “D” (onde está escrito “UMA BRINCADEIRA DE ESCONDER”), o gabarito do item, foi a alternativa mais escolhida a partir do Nível 3. Todavia, por ser um item que está no meio da escala, com um parâmetro b aproximadamente com um valor igual a 500 e, posicionado no

Nível 3, se torna um item mais “difícil” para alunos classificados nos Níveis 1 e 2, algo confirmado com os dados do gráfico.

Por conseguinte, as alternativas incorretas, as que não são consideradas como o gabarito do item, são mais escolhidas, o que nos faz perceber que existem entendimentos diferentes, por parte dos alunos, em relação ao que venha a ser a resposta correta para o item. No Nível 1 temos uma concentração maior de marcações nas alternativas “A” e “B”, sendo alternativas que indicam uma localização no texto base do conteúdo da alternativa. Além de serem as duas alternativas com os menores pseudoB que compõem o item.

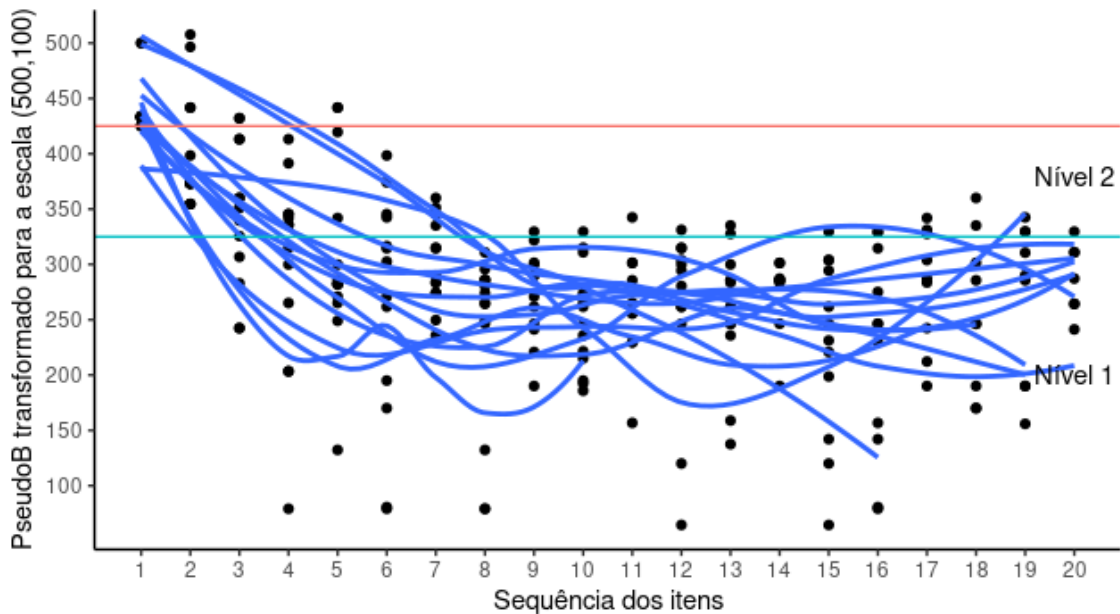
Um outro aspecto interessante na análise dos dados desse item, é o fato dos Níveis 3 e 4 terem, para além do gabarito mais marcado, a alternativa “C” (onde está escrito “UM COELHO NO COQUEIRO”), duas palavras “COELHO” e “COQUEIRO” que estão presentes no texto, como uma das alternativas mais escolhidas. Essa alternativa é a que tem o segundo maior valor de pseudoB, o que a torna como a que mais se aproxima do gabarito. Nesse sentido, o que temos é de fato uma diversificação de escolhas ao longo dos níveis, o que contribui para que o próximo item não seja igual para todos os alunos, mas que se diferencie entre as escolhas.

À vista disso, podemos esperar que exista um movimento mais uniforme nas escolhas das alternativas quando analisamos por nível. Essa hipótese pode ser alimentada quando investigamos as marcações pelos pseudoB também por níveis. O Gráfico 13 apresenta as trajetórias elaboradas a partir de uma regressão local (loess)<sup>19</sup>, que permitem avaliar as marcações a partir do pseudoB.

---

<sup>19</sup> Para mais informações consultar o link: [projetos:saudavel:loess.pdf\(ufpr.br\)](http://projetos:saudavel:loess.pdf(ufpr.br)).

**Gráfico 13** – TAI-Leitura 3.0. Distribuição no Nível 1 dos pseudoB das alternativas



**Fonte:** Elaboração própria.

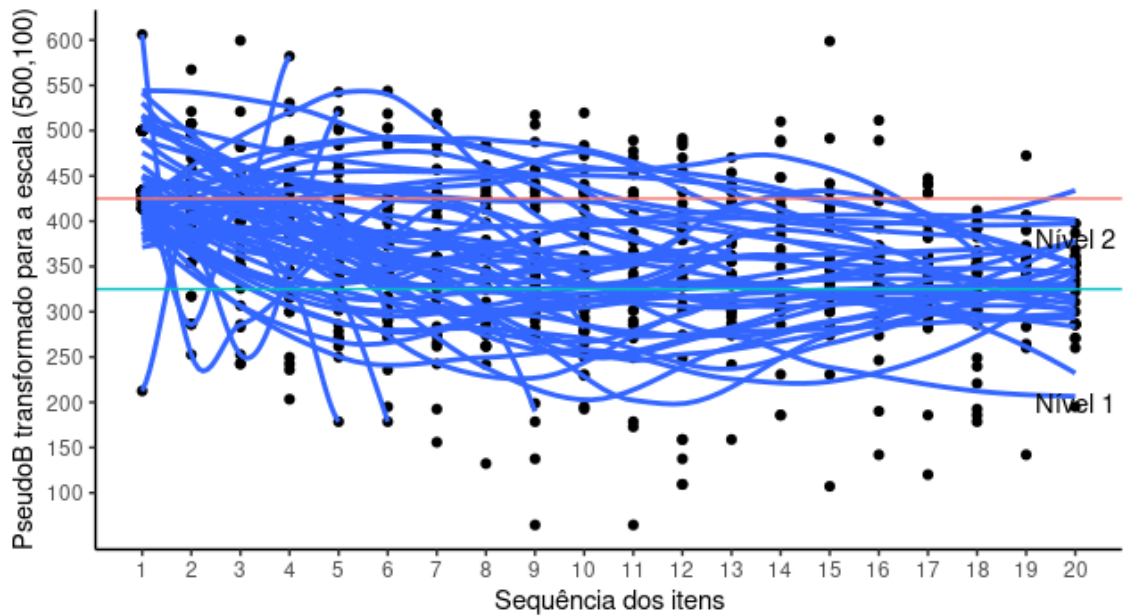
No Gráfico 13 o eixo das abscissas representa os 20 itens, com suas respectivas alternativas, possíveis de serem administrados aos alunos. O eixo das ordenadas é a escala de proficiência utilizada na PB. Os pontos são os pseudoB de cada alternativa escolhida, por cada aluno classificado no Nível 1, para cada item oferecido, sendo que os itens podem ser diferentes, pois dependem da resposta ao item anterior. As linhas azuis são regressões locais, que buscam suavizar a relação entre conjuntos de pontos em cada trecho do gráfico. Entretanto, por meio dessas linhas, podemos compreender como as trajetórias das marcações dos alunos estão representadas.

Ao longo dos 20 itens que podem compor o teste, temos uma diminuição gradativa da medida do pseudoB, chegando a estar, praticamente, localizados no Nível 1. Isso significa que os itens que foram oferecidos a partir do oitavo item, têm em suas alternativas, incluindo o próprio gabarito, abaixo de 325 ou muito próximo desse valor. Deste modo, temos uma trajetória das escolhas das alternativas muito alinhada com as exigências para quem está no Nível 1.

Nesta mesma linha de raciocínio, o Gráfico 14 apresenta, como dados para o Nível 2, a mesma metodologia de análise.



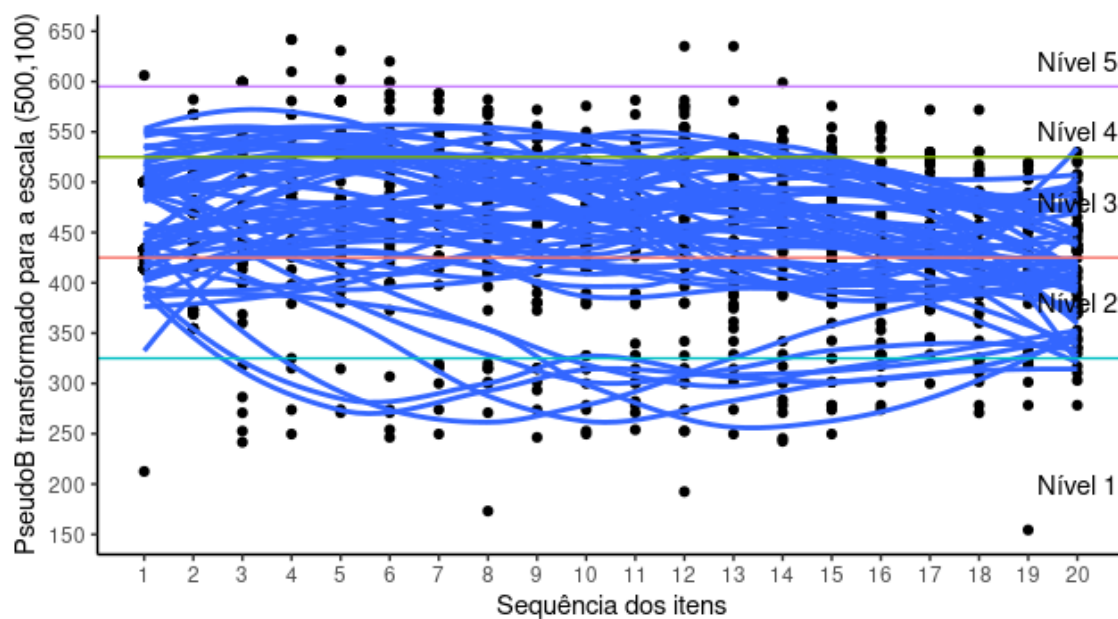
**Gráfico 14** – TAI-Leitura 3.0. Distribuição no Nível 2 dos pseudoB das alternativas



**Fonte:** Elaboração própria.

As escolhas das alternativas dos alunos que foram caracterizados como Nível 2, se diferencia em relação ao nível anterior, pois acaba ocorrendo uma maior variabilidade de pseudoB das alternativas, incluindo um item com valor de 600 pontos de proficiência para um pseudoB. Entretanto, as trajetórias demonstram que a tendência do teste foi apresentar itens que estivessem localizados nos Níveis 1 e 2, o que é excelente, pois consegue de fato oferecer itens que estejam próximos da proficiência dos alunos deste nível. O Gráfico 15 traz os dados para o Nível 3, o que possibilita analisar a hipótese de diversificação de marcações levantada anteriormente.

**Gráfico 15** – TAI-Leitura 3.0. Distribuição no Nível 3 dos pseudoB das alternativas



**Fonte:** Elaboração própria.

Temos uma grande diversificação nas opções de escolhas das alternativas, que pode ser comprovada com a dispersão dos pontos pseudoB de cada alternativa no gráfico, e as duas trajetórias separadas que aparecem na análise. Essa diferenciação nas trajetórias, pode nos mostrar o quanto é fundamental a análise pelas alternativas, pois, mesmo com caminhos tão diversos, os alunos foram classificados como Nível 3. Obviamente, que essa classificação depende da estimativa da proficiência, e somente só, porém esse número representa a capacidade de um aluno enfrentar itens de Leitura e, logo, ter caminhos diversos, o que não deveria influenciar a sua estimativa, somente, a interpretação pedagógica.

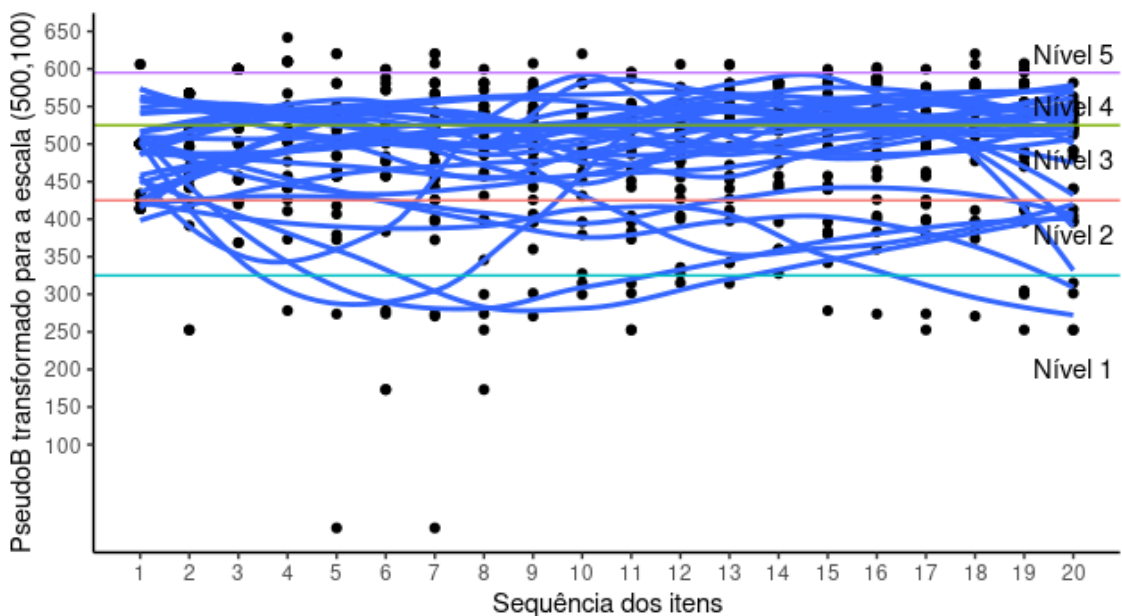
Outro ponto que pode ser destacado do Gráfico 15, é a concentração de pseudoB próximo aos Níveis 1, 2 e 3 ao final da aplicação do TAI-Leitura. Isso corrobora com duas ideias que estão no cerne de qualquer TAI: buscar a adaptabilidade do instrumento ao nível do respondente e dispor de uma grande variabilidade de itens no banco, aliada a uma maior quantidade de itens. Com isso posto e analisando o Gráfico 16, podemos identificar que o TAI-Leitura conseguiu se adaptar ao nível do aluno, independente se ele tem maior ou menor proficiência, pois as marcações representadas no gráfico são de alternativas que cada aluno de fato marcou; logo, existia a possibilidade de ele encontrar uma alternativa, no item que lhe foi oferecido, que

melhor se adequava ao seu nível de proficiência no momento de resolução da tarefa proposta no item.

Destarte, mesmo que tenha uma ampliação na quantidade de itens oferecido, não estamos em uma corrida de itens para saber quem chegará mais rápido ao final do teste, pois os itens estavam alinhados com a capacidade dos alunos, por conta disso, ao final do teste, temos alternativas que não ultrapassam o Nível 3, porém diversas ocupam níveis inferiores, o que pode indicar a escolha de uma alternativa errada, mas que contribui para a estimativa da proficiência e para conhecer os acertos e erros de cada aluno.

O Gráfico 16 apresenta os dados dos pseudoB por alternativa para cada aluno classificado no Nível 4.

**Gráfico 16** – TAI-Leitura 3.0. Distribuição no Nível 4 dos pseudoB das alternativas



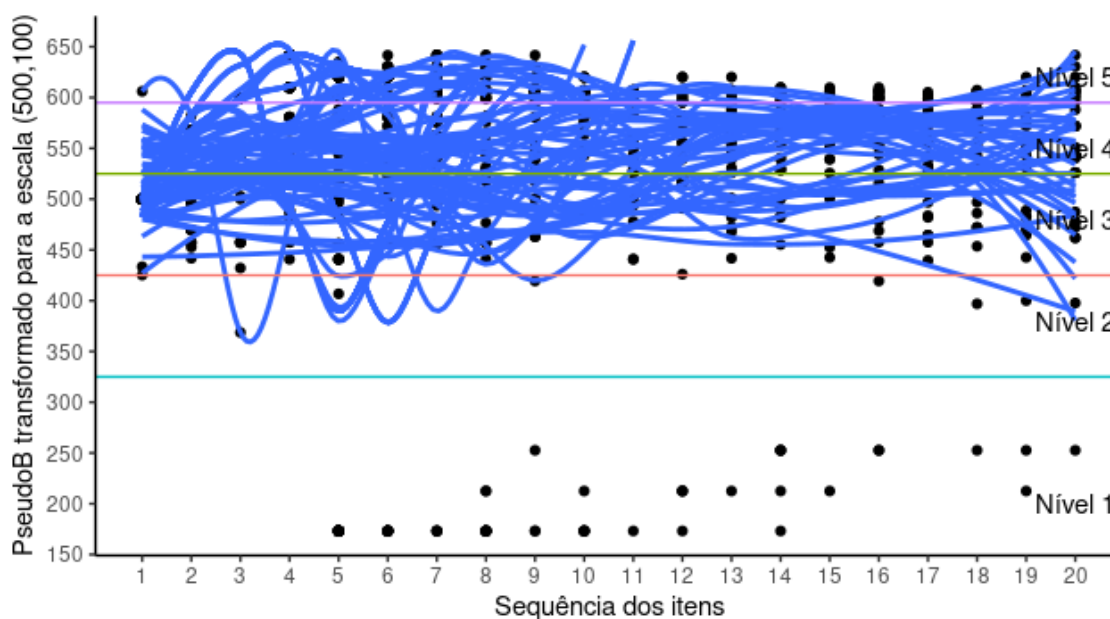
**Fonte:** Elaboração própria.

A análise do Gráfico 16 nos leva a considerar que os Níveis 3 e 4 têm uma concentração de escolhas um pouco maior na alternativa com o segundo maior pseudoB dentre as demais, se diferenciando consideravelmente dos Níveis 1 e 5. Para o Nível 3, pudemos identificar que de fato houve uma dispersão maior entre as marcações, o que pode ser ocasionado, como já mencionado, por uma variabilidade nas escolhas das alternativas. No Gráfico 16 temos algo

muito parecido, porém com uma grande concentração de dados (pseudoB) na região do próprio Nível 4, com uma diversificação para modelagens que tentam dar conta das marcações que se concentraram nos Níveis 1, 2 e 3. De qualquer maneira, temos ao final, o pseudoB das alternativas que estão localizados abaixo do Nível 5, algo que ocorreu para todos os níveis até o momento.

Por fim, o Gráfico 17 nos traz os dados dos alunos que estão classificados no Nível 5.

**Gráfico 17** – TAI-Leitura 3.0. Distribuição no Nível 5 dos pseudoB das alternativas



**Fonte:** Elaboração própria.

Diferentemente das interpretações dos dados dos demais gráficos apresentados para os níveis inferiores, para os alunos que estão no Nível 5 temos uma maior dificuldade para modelar as escolhas, pois, como é possível observar no Gráfico 17, temos algumas marcações que estão localizadas na região do Nível 1 e que podem representar escolhas de alternativas incorretas, mas que contribuem para a compreensão e diferenciação dos alunos classificados neste nível. Outro ponto que podemos destacar, é o deslocamento da ampla maioria dos pseudoB para os níveis acima do 2, mesmo com um resquício no Nível 1.

De maneira geral, podemos compreender que ambos os algoritmos do TAI-Leitura, com análise das alternativas, conseguiram diferenciar os alunos de cada nível, oferecendo itens que de fato se adéquam à estimativa ao nível de proficiência que cada aluno está no momento da aplicação.

Com efeito, é possível sustentar que a confirmação da hipótese de se estabelecer um pseudoB para atender à necessidade de modelagem no tratamento de respostas, politomizando os itens que originalmente eram tratados como dicotômicos – acertando ou errando –, considerando uma “dificuldade” para cada alternativa. Contudo, por conta do aumento de itens no TAI-Leitura 3.0 pseudoB, nos parece mais indicado, em uma aplicação com uma quantidade maior de alunos, concentrarmos a grande maioria no TAI-Leitura 3.0.

Todavia, a ideia é continuar utilizando o pseudoB para analisar as trajetórias e na interpretação pedagógica geral e de cada aluno. Dessa maneira, iremos no próximo capítulo apresentar os resultados da aplicação em cinco escolas parceiras, contando com um quantitativo de aproximadamente 2.200 alunos respondendo ao TAI-Leitura. Desse montante, iremos direcionar, com base em um sorteio, 220 alunos (10% do total) para realizarem o TAI-Leitura 3.0 pseudoB, assim, continuamos analisando a validade desse algoritmo e ampliamos a taxa de respostas para ambos os algoritmos.

## 5 – Análise dos resultados da aplicação do TAI-Leitura 3.0

Como forma de validar o que foi desenvolvido ao longo da pesquisa e do próprio pré-teste do TAI-Leitura 3.0, realizamos uma aplicação para alunos do 1º ao 9º ano do Ensino Fundamental de três escolas da RME-SP: Escola Maurîtânia, Escola Senegal e Escola Zimbábue. As aplicações do TAI-Leitura 3.0 ocorreram ao longo das duas primeiras semanas do mês de abril de 2023, com uma maior concentração nos dias 3, 4 e 5 desse mês. A Tabela 13 apresenta as quantidades de participantes em cada escola e o percentual que essas quantidades representam em relação ao total.

**Tabela 13** – TAI-Leitura 3.0. Quantidade e percentual de alunos participantes. Escolas Maurîtânia, Senegal e Zimbábue. Abril 2023

Escolar	Escolas						Total	Percentual de respondente
	Maurîtânia		Senegal		Zimbábue			
	Quantidade de respondentes	Percentual de respondente	Quantidade de respondentes	Percentual de respondente	Quantidade de respondentes	Percentual de respondente		
1	49	79,0%	40	69,0%	134	88,2%	223	82,0%
2	56	90,3%	53	88,3%	52	89,7%	161	89,4%
3	80	89,9%	66	86,8%	50	84,7%	196	87,5%
4	52	81,3%	51	79,7%	101	91,0%	204	85,4%
5	55	85,9%	25	39,1%	85	95,5%	165	76,0%
6	64	100,0%	50	79,4%	161	95,3%	275	92,9%
7	67	98,5%	25	39,1%	152	89,9%	244	81,1%
8	85	92,4%	35	58,3%	100	93,5%	220	84,9%
9	76	96,2%	47	72,3%	121	89,6%	244	87,5%
Total	588	91,3%	480	83,6%	984	93,8%	2052	90,5%

**Fonte:** Elaboração própria.

Podemos observar que houve uma participação acima de 80% em todas as escolas, o que podemos considerar como algo interessante para o teste do TAI-Leitura 3.0. Além do mais, ao analisarmos os dados de maneira detalhada podemos constatar que nas Escolas Maurîtânia e Zimbábue, praticamente, todos os anos escolares ficaram com uma participação próxima ou acima dos 80%. Todavia, não temos esses mesmos resultados na Escola Senegal, em que somente dois anos escolares obtiveram valores percentuais acima dos 80%, os demais, ficaram abaixo, com destaque para o 5º ano, que somente 39,1% dos alunos participaram da aplicação devido a ausências nos dias de aplicação. De toda forma, para este trabalho a Tabela 13 apresenta um panorama geral de como foram as aplicações em relação à participação e nos

permite afirmar que a quantidade de respondentes foi suficiente para testar os algoritmos do TAI-Leitura 3.0.

Com base nesses quantitativos podemos considerar que existe a possibilidade de termos uma distribuição dos alunos em todos os 5 níveis de proficiência, como originalmente estabelecidos na PB, pois com uma maior participação, aumenta a probabilidade de termos alunos com diferentes medidas do traço latente – proficiência em leitura –, sendo assim, uma maior variabilidade de alunos entre os níveis e internamente a eles. A Tabela 14 nos mostra a distribuição dos alunos nos níveis da PB, no geral, sem considerar os anos escolares.

**Tabela 14** – TAI-Leitura 3.0. Distribuição dos alunos por escola em cada nível. Abril 2023

Nível	Escolas						Total	Percentual de respondente
	Mauritânia		Senegal		Zimbábue			
	Quantidade de respondentes	Percentual de respondente	Quantidade de respondentes	Percentual de respondente	Quantidade de respondentes	Percentual de respondente		
1	21	3,6%	31	6,5%	53	5,4%	105	5,1%
2	119	20,2%	151	31,5%	195	19,8%	465	22,7%
3	83	14,1%	86	17,9%	140	14,2%	309	15,1%
4	44	7,5%	46	9,6%	85	8,6%	175	8,5%
5	321	54,6%	166	34,6%	511	51,9%	998	48,6%
Total	588	100,0%	480	100,0%	984	100,0%	2052	100,0%

**Fonte:** Elaboração própria.

Na Tabela 14 temos a quantidade de alunos que foram classificados em cada nível por escola, independente do ano escolar. Ademais, como os itens da PB foram elaborados com base em uma aplicação para o 2º ano do Ensino Fundamental, temos a expectativa de uma maior concentração de alunos no Nível 5, pois temos uma quantidade maior de participantes do 3º ao 9º ano. Essa hipótese é confirmada a partir dos dados da Tabela 14, que nos mostram uma concentração acima de 30% no Nível 5 nas três escolas, além de termos quase 49% de alunos no Nível 5 no geral. Todavia, os Níveis 2 e 3 aparecem como os segundo e terceiro níveis, respectivamente, com percentual maior de alunos. Sendo que na Escola Senegal temos o maior percentual de alunos no Nível 2 em comparação com as demais escolas, porém, sem representar o maior valor absoluto, pois a Escola Zimbábue, por ter uma quantidade maior de alunos participantes, apresenta 195 alunos neste mesmo Nível.

Um fato importante a ser destacado é o período do ano que o TAI-Leitura 3.0 foi aplicado e a quantidade de alunos no Nível 1, que em todas as escolas foi a menor concentração relativa,

pois era esperado que houvesse uma quantidade maior de alunos no Nível 1, principalmente, pelo fato de termos alunos dos 2<sup>os</sup> e 3<sup>os</sup> anos que foram diretamente impactados pelos efeitos da pandemia de Covid-19, como também alunos do 1<sup>o</sup> ano que haviam completado apenas dois meses de escolarização no Ensino Fundamental. De todo modo, mesmo com a participação de anos escolares em que era esperado uma maior quantidade nos Níveis 4 e 5, tivemos os valores percentuais do Nível 4 muito próximos aos valores do Nível 1 e bem abaixo dos apresentados nos demais níveis. Dessa maneira, podemos afirmar que alunos do 3<sup>o</sup> ao 9<sup>o</sup> ano apresentam proficiências estimadas que foram classificadas nos Níveis 2 e 3, além dos Níveis 4 e 5 que já eram esperados.

Como já apresentado no capítulo 4, na aplicação do TAI-Leitura 3.0 utilizamos dois algoritmos para construir o TAI, sendo que a diferença está na utilização ou não do pseudoB na metodologia de escolha do próximo item. Todavia, como ainda precisamos avançar em mais testes para compreender melhor quais são os impactos da utilização do pseudoB nesse processo, optamos, nesta etapa da pesquisa, por utilizar uma amostra relativamente pequena, pois ampliamos a quantidade total de alunos, porém não modificamos o total de alunos que fizeram o TAI-Leitura 3.0 pseudoB. A Tabela 15 nos mostra a quantidade de alunos que fizeram cada um dos dois testes.

**Tabela 15** – TAI-Leitura 3.0. Distribuição dos alunos nos dois algoritmos. Abril 2023

Modelo TAI	Quantidade de respondentes
PseudoB	50
TaiAlt	2002
Total	2052

**Fonte:** Elaboração própria.

No pré-teste dos algoritmos duas turmas fizeram o TAI-Leitura 3.0 pseudoB, o que totalizou 32 alunos. Dessa maneira, para a aplicação geral ampliamos a distribuição desse modelo do TAI para 60 alunos e, para tanto, realizamos um sorteio de 20 alunos contando com todos os anos escolares por escola. Pelos dados que temos na Tabela 15, não conseguimos ter a totalidade de alunos fazendo esse modelo de TAI, porém tivemos uma quantidade relativamente próxima ao que foi a participação no outro modelo. De qualquer maneira, iremos analisar de modo geral o



TAI, sem considerar a diferença entre os modelos. De toda forma, por termos poucos respondentes no TAI-Leitura 3.0 pseudoB, decidimos analisar dados macros, como as distribuições por níveis e anos escolares, e dados específicos, como as análises de trajetória de um aluno no teste, de modo geral. O pseudoB será utilizado para interpretarmos as “dificuldades” de cada alternativa, porém não vamos separar as análises. Nesse sentido, temos a Tabela 16 que apresenta a média de itens utilizados no TAI-Leitura 3.0 para cada nível.

**Tabela 16** – TAI-Leitura 3.0. Distribuição da média de itens aplicados por Nível. Abril 2023

Nível	Média de itens			
	Mauritânia	Senegal	Zimbábue	Geral
1	18,0	16,2	17,8	17,4
2	15,2	15,5	15,6	15,4
3	19,5	19,4	19,4	19,4
4	20,0	20,0	20,0	20,0
5	11,0	13,6	12,0	11,9
Total	14,0	16,0	14,7	14,8

**Fonte:** Elaboração própria.

É possível constatar que temos uma quantidade de itens próxima ao que a PB apresentava em seu instrumento, porém é importante salientar que os itens em um TAI estão mais aproximados ao que de fato cada aluno consegue responder. Assim, ter uma quantidade próxima ao que era o teste da PB não significa algo ruim, pois, pelo contrário, pode indicar uma maior especificidade de cada teste, onde os acertos e erros possibilitam uma melhor compreensão pedagógica do que cada aluno consegue fazer em relação à leitura. No geral, o Nível 4 foi o que teve a maior quantidade de itens utilizados, o que não significa que não houve alunos que responderam poucos itens, mas, em média, o teste parou por conta de alcançar a quantidade máxima de itens estipulada, 20 itens. Dessa forma, podemos ter como suposição para essa quantidade de itens maior no Nível 4 que haja alguma limitação de quantidade e variabilidade de itens no Banco de Itens (BI), o que iremos explorar um pouco mais essa situação adiante.

O Nível 5 foi o que apresentou a menor quantidade de itens em média, o que demonstra que muitos alunos acertavam os primeiros itens e o teste parava rapidamente. Isso foi constatado ao longo das aplicações, principalmente para os anos escolares mais elevados, como 7º, 8º e 9º ano. Essa informação, por mais que o cerne da PB não seja ter itens que deem conta de coletar informações sobre o domínio em leitura no modo geral, mas, mais especificamente, dos anos

iniciais do Ensino Fundamental – uma fase do processo de alfabetização –, pode também nos mostrar uma limitação da abrangência do Banco de Itens, em que os alunos que começam em uma trajetória de muitos acertos, acabam com o teste limitando a medida da sua proficiência por conta dos itens no banco, pois o instrumento, mesmo sendo um TAI, chegou a um limite na medida do traço latente. De todo modo, o TAI-Leitura 3.0 demonstrou ter capacidade de estimar proficiências em diversos pontos do espectro do traço medido, possibilitando classificar as proficiências dos alunos nos diversos níveis estipulados. Na sequência são apresentados e analisados os resultados agrupando os alunos por Nível. A Tabela 17 nos mostra a quantidade de alunos classificados no Nível 1 por ano escolar em cada escola, além da média de proficiência e do erro padrão para os alunos nesse Nível e ano escolar.

**Tabela 17** – TAI-Leitura 3.0. Quantidade de alunos no Nível 1, médias de proficiência e erro padrão por ano escolar. Abril 2023

Escola	Ano Escolar	Nível 1		
		Quantidade de respondentes	Média Proficiência	Média Erro
Mauritânia	1	15	296,6	29,3
	2	1	298,2	26,4
	3	3	303,0	27,5
	4			
	5			
	6	1	260,8	39,3
	7	1	320,1	21,9
	8			
	9			
Senegal	1	16	278,2	36,4
	2	8	296,4	30,5
	3	3	317,1	23,2
	4			
	5			
	6	1	275,8	30,1
	7			
	8			
	9			
Zimbábue	1	41	292,6	31,0
	2	4	302,3	26,2
	3	2	310,0	24,3
	4			
	5	3	286,6	30,8
	6			
	7	1	295,9	28,8
	8			
	9			
Total		105	292,4	31,0

**Fonte:** Elaboração própria.

Como esperado, a maior quantidade de alunos no Nível 1 é do 1º ano do Ensino Fundamental, principalmente, por estarem nos primeiros meses de escolarização. Outro ponto importante, é

o fato de apresentarem as menores médias de proficiências em relação aos demais anos escolares. Todavia, existem alguns alunos em alguns anos escolares que apresentam proficiência menor que a média do 1º ano, por exemplo um aluno do 6º ano na escola Senegal, que obteve 275,8 como medida da sua proficiência em leitura, enquanto a média para o 1º ano desta escola foi 278,2. De maneira similar, na escola Maurîtânia temos um aluno no 6º ano com proficiência menor que a média do 1º ano. Na Tabela 18 temos as informações quantitativas para o Nível 2 em cada ano escolar.

**Tabela 18** – TAI-Leitura 3.0. Quantidade de alunos Nível 2, médias de proficiência e erro padrão por ano escolar. Abril 2023

Escola	Ano Escolar	Nível 2		
		Quantidade de respondentes	Média Proficiência	Média Erro
Maurîtânia	1	31	355,5	22,0
	2	33	368,7	23,9
	3	21	366,7	23,0
	4	11	371,4	23,4
	5	8	366,5	23,2
	6	4	388,4	24,5
	7	5	361,4	23,5
	8	4	367,6	21,9
	9	1	361,7	22,5
Senegal	1	22	357,3	22,6
	2	37	361,9	23,2
	3	35	364,9	22,6
	4	12	368,3	22,7
	5	10	364,0	23,8
	6	5	379,3	25,1
	7	2	347,4	25,2
	8	3	362,9	25,6
	9	4	369,1	25,0
Zimbábue	1	83	356,8	22,1
	2	31	360,8	22,5
	3	19	361,4	22,9
	4	12	364,7	24,3
	5	12	359,7	22,6
	6	15	365,5	23,2
	7	7	373,4	24,1
	8	6	377,8	26,1
	9	5	367,6	24,3
Total		465	362,8	23,0

**Fonte:** Elaboração própria.

De forma diferente do que ocorreu no quantitativo do Nível 1, a Tabela 18 nos apresenta dados que mostram alunos com proficiência no Nível 2 em todos os anos escolares, o que indica a existência de alunos que estão no 7º, 8º e 9º ano do Ensino Fundamental com grandes dificuldades ainda no processo de alfabetização em torno do domínio do Sistema de Escrita Alfabética e ainda mais em leitura propriamente dita. Entretanto, temos uma concentração maior de estudantes nos primeiros anos do Ensino Fundamental, sendo que na Escola Zimbábue a maior quantidade está no 1º ano, diferente do que ocorre nas outras duas escolas, em que a maioria dos alunos no Nível 2 está localizada no 2º ano. Outro ponto que consideramos importante destacar é o fato de não termos nenhum tipo de progressão das médias de proficiências conforme avançamos nos anos escolares, algo que é esperado, porém destacamos que isso demonstra que de fato os alunos estão no Nível 2, mesmo estando em anos escolares mais avançados; logo, as suas proficiências não se diferem muito de alunos que estão no 1º ou 2º ano. A Tabela 19 nos traz os dados do Nível 3.

**Tabela 19 – TAI-Leitura 3.0. Quantidade de alunos no Nível 3, médias de proficiência e erro padrão por ano escolar. Abril 2023**

Escola	Ano Escolar	Nível 3		
		Quantidade de respondentes	Média Proficiência	Média Erro
Mauritânia	1	3	451,3	33,1
	2	15	453,9	31,5
	3	26	470,6	31,2
	4	8	475,5	30,4
	5	14	471,3	30,7
	6	6	485,3	31,1
	7	6	475,4	31,0
	8	4	453,3	30,6
	9	1	459,3	32,5
Senegal	1	2	455,7	30,0
	2	8	455,8	29,9
	3	18	471,9	30,4
	4	8	465,7	31,4
	5	8	478,3	30,9
	6	13	471,3	32,0
	7	3	504,3	29,7
	8	4	468,2	31,8
	9	4	494,4	30,1
Zimbábue	1	10	461,5	30,9
	2	14	458,8	30,5
	3	14	465,4	30,8
	4	21	467,8	30,8
	5	19	473,1	30,6
	6	29	471,9	30,7
	7	17	464,0	30,6
	8	7	470,7	31,3
	9	9	465,8	30,8
Total		309	468,3	30,9

**Fonte:** Elaboração própria.

Para o Nível 3, o esperado é uma menor quantidade de alunos dos dois primeiros anos do Ensino Fundamental, além de uma maior concentração de alunos no 3º e 4º ano, contudo a distribuição dos alunos no Nível 3 nos mostra que, além do 3º e 4º ano, outros anos escolares podem ser considerados como comuns para concentrar alunos classificados nesse nível, como, por exemplo, o 5º ano com 41 alunos com estimativas de proficiências no Nível 3. Ao mesmo tempo, nos surpreende a quantidade de alunos que aparecem do 7º ao 9º ano nesse nível, principalmente, na Escola Zimbábue.

**Tabela 20** – TAI-Leitura 3.0. Quantidade de alunos Nível 4, médias de proficiência e erro padrão por ano escolar. Abril 2023

Escola	Ano Escolar	Nível 4		
		Quantidade de respondentes	Média Proficiência	Média Erro
Mauritânia	1			
	2	2	565,7	29,6
	3	10	554,6	29,9
	4	6	556,7	29,7
	5	7	555,6	29,9
	6	7	558,6	29,6
	7	1	558,2	29,9
	8	3	548,4	30,3
	9	7	552,7	30,3
Senegal	1			
	2			
	3	7	550,8	29,9
	4	10	552,9	30,3
	5	1	572,9	29,5
	6	6	552,2	29,6
	7	2	567,7	29,6
	8	2	545,7	30,0
	9	6	549,8	30,2
Zimbábue	1			
	2	3	555,6	29,6
	3	5	553,4	29,8
	4	18	558,1	30,3
	5	5	563,3	33,3
	6	19	556,8	30,2
	7	12	544,8	30,9
	8	3	550,4	29,9
	9	16	556,2	30,1
Total		175	555,2	30,2

**Fonte:** Elaboração própria.

A Tabela 20 apresenta os dados quantitativos para o Nível 4, na qual é possível observar que em todas as escolas, nenhum aluno do 1º ano não foi classificado no Nível 4, algo que demonstra

existir uma coerência no teste, pois esse nível pode ser considerado bastante elevado para alunos que praticamente estavam começando o Ensino Fundamental quando da aplicação do teste. Outro ponto importante é a diminuição da quantidade de alunos nesse nível, em comparação com os Níveis 2 e 3. De certa maneira, essa percepção de um quantitativo menor, pode indicar que o teste teve dificuldades para considerar alunos nesse nível, provavelmente, como decorrência das limitações do Banco de Itens. De modo geral, a Escola Zimbábue continua apresentando as maiores quantidades de alunos nos anos escolares mais elevados, em contraste com as outras duas escolas.

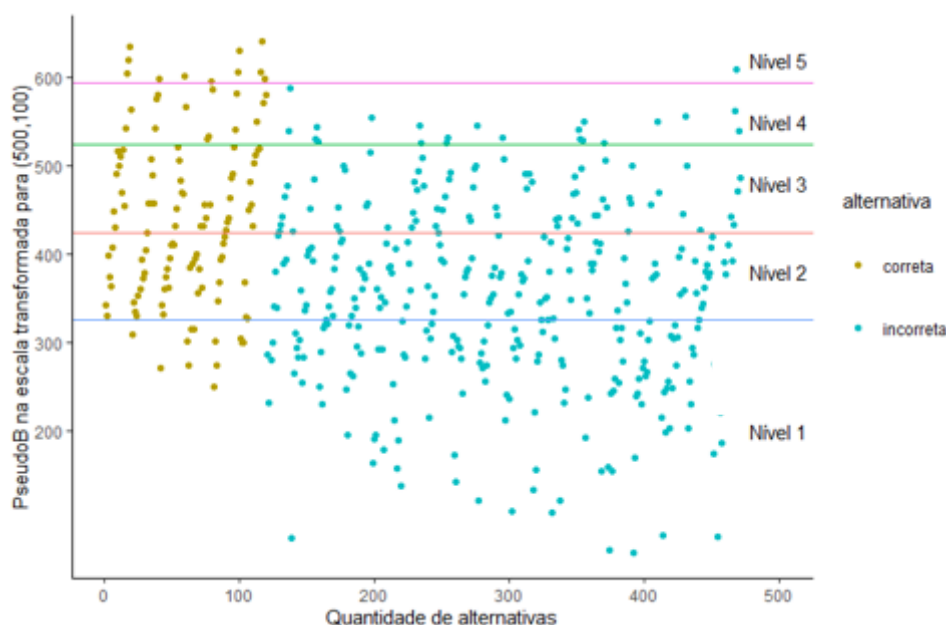
**Tabela 21** – TAI-Leitura 3.0. Quantidade de alunos no Nível 5, médias de proficiência e erro padrão por ano escolar. Abril 2023

Escola	Ano Escolar	Nível 5		
		Quantidade de respondentes	Média Proficiência	Média Erro
Mauritânia	1			
	2	5	616,4	34,4
	3	20	632,3	37,7
	4	27	655,0	44,1
	5	26	655,3	44,8
	6	46	673,5	50,0
	7	54	671,2	49,8
	8	74	668,8	48,6
	9	67	669,3	49,0
Senegal	1			
	2			
	3	3	589,4	30,1
	4	21	623,8	36,3
	5	6	627,1	37,3
	6	25	640,6	41,2
	7	18	657,7	45,1
	8	26	667,0	47,1
	9	33	665,9	48,5
Zimbábue	1			
	2			
	3	10	631,0	36,5
	4	50	655,2	43,8
	5	46	657,0	43,7
	6	98	650,7	43,4
	7	115	657,4	45,3
	8	84	660,3	45,1
	9	91	671,4	49,4
Total		998	658,2	45,4

Fonte: Elaboração própria.

A Tabela 21 apresenta a quantidade de alunos por ano escolar do Nível 5, no qual é esperado que exista uma maior concentração de alunos nos anos escolares mais elevados e isso se confirma com os dados da Tabela, embora a média de proficiência do 9º ano não é a maior em todas as escolas, o que demonstra que nem todos os alunos acertaram os 5 primeiros itens oferecidos e, assim, não conseguiram alcançar o limite que o BI oferece. De qualquer maneira, não temos nenhum aluno do 1º ano e somente na Escola Maurítânia temos cinco alunos do 2º ano que apresentam proficiência para serem classificados nesse nível, algo passível de ocorrer. Os dados apresentados, para o conjunto das tabelas, nos mostram que o TAI-Leitura 3.0 conseguiu, efetivamente, estimar as proficiências de modo coerente com os níveis nos quais foram alocadas, sinalizando que funcionou como um instrumento para medir a capacidade dos alunos em leitura, incluindo aqueles alunos que estão iniciando o processo de alfabetização. Todavia, em diversos momentos apresentamos que houve limitações do Banco de Itens que podem ter impactado a trajetória dos alunos na aplicação do TAI e para investigar essa situação, na sequência, vamos analisar a distribuição dos parâmetros  $b$  e dos pseudoB das alternativas consideradas incorretas. O Gráfico 18 nos apresenta essas informações sobre o Banco de Itens do TAI-Leitura 3.0.

**Gráfico 18** – TAI-Leitura 3.0. Distribuição dos parâmetros  $b$  e pseudoB dos itens do Banco de Itens

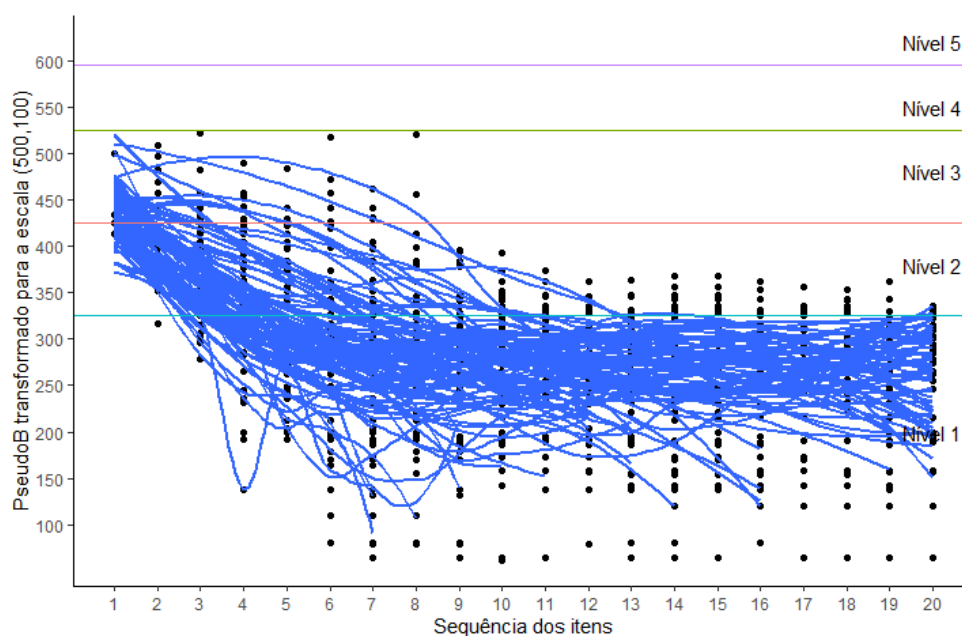


**Fonte:** Elaboração própria.

O Gráfico 18 apresenta a quantidade de parâmetros  $b$  e pseudoB que temos no Banco de Itens e é possível observar uma grande concentração de alternativas corretas nos Níveis 2 e 3 e uma pequena quantidade nos Níveis 1 e 5. Ao mesmo tempo, o Nível 1 apresenta uma grande quantidade de alternativas incorretas, o pseudoB de diversos itens. Entretanto, os Níveis 4 e 5, principalmente o Nível 5, praticamente não apresentam nenhum pseudoB, o que indica a limitação do banco na parte superior, o que pode ser explicado é que nesses níveis estão os itens “mais difíceis”, isto é, com gabaritos – a alternativa que exige maior proficiência para ser assinalada entre as alternativas de um item de múltipla escolha – nessa região de proficiência. Assim, para pseudoB nessa faixa precisaríamos de itens de níveis superiores ao Nível 5. De modo geral, se considerarmos todos os parâmetros como sendo pseudoB temos uma grande variabilidade entre os Níveis 1, 2 e 3, e uma menor nos demais níveis.

Para avançarmos nas análises da aplicação realizada em abril, realizamos a mesma projeção das trajetórias que fizemos no capítulo 4, ou seja, utilizamos a distribuição dos pseudoB para realizarmos uma regressão que suaviza as curvas conforme os pontos se distribuem. O Gráfico 19 apresenta a distribuição dos pseudoB das alternativas escolhidas pelos alunos cuja proficiência foram classificados no Nível 1.

**Gráfico 19** – TAI-Leitura 3.0. Distribuição no Nível 1 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023

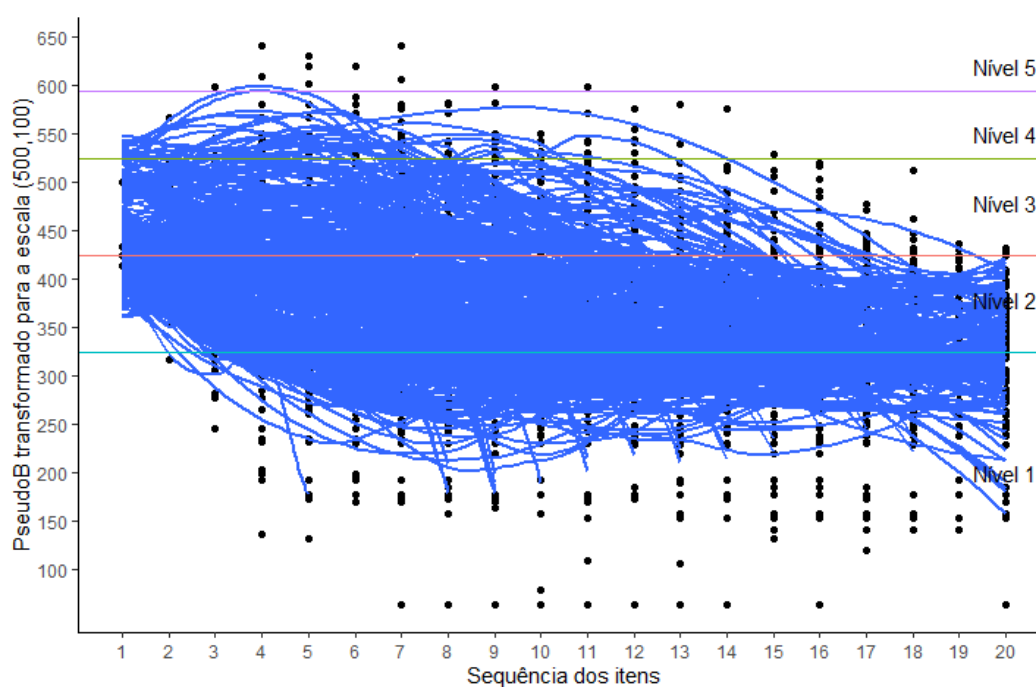


**Fonte:** Elaboração própria.



Podemos observar que a partir do item 9 temos uma queda nos valores do pseudoB de cada alternativa escolhida, indicando que os itens que foram administrados para os alunos do Nível 1 estavam localizados entre os Níveis 1 e 2. Já as trajetórias estimadas com base nos dados, demonstram que realmente essa concentração ocorreu, principalmente, no Nível 1, confirmando que as proficiências desses alunos, de fato, estão nesse nível ou muito próximas do limite com o Nível 2. O Gráfico 20 apresenta a distribuição dos pseudoB para o Nível 2.

**Gráfico 20** – TAI-Leitura 3.0. Distribuição no Nível 2 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023

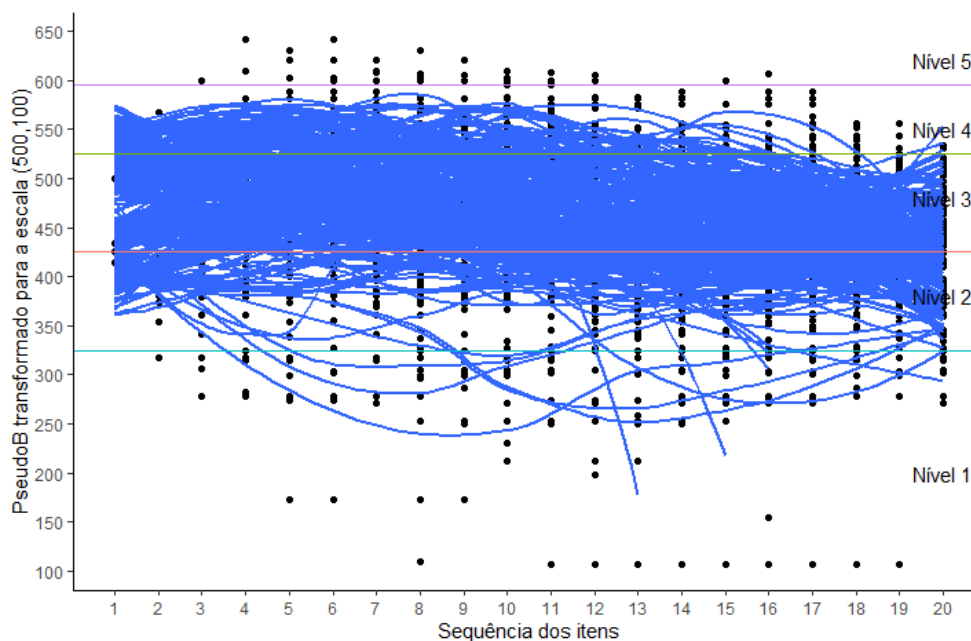


**Fonte:** Elaboração própria.

A concentração de pseudoB nas alternativas escolhidas nesse nível tende a ficar entre os Níveis 1 e 2 ao final dos 20 itens, o que demonstra que mesmo aqueles alunos que precisaram de todos os itens aplicados no teste, tiveram uma trajetória que classificou suas proficiências nesses níveis com alternativas incorretas com pseudoB, de maneira igual nessa faixa de proficiência. Ao mesmo tempo, o fato de termos valores que estão no Nível 5, ratifica ainda mais essa percepção, pois mesmo com a escolha de uma alternativa correta, o TAI-Leitura 3.0 conseguiu estimar de maneira mais precisa a proficiência dos alunos, principalmente por apresentar tarefas que possam ser realmente realizadas pelos alunos. O Gráfico 21 apresenta os dados de

distribuição dos pseudoB dos itens para as alternativas escolhidas entre os alunos que foram classificados como Nível 3.

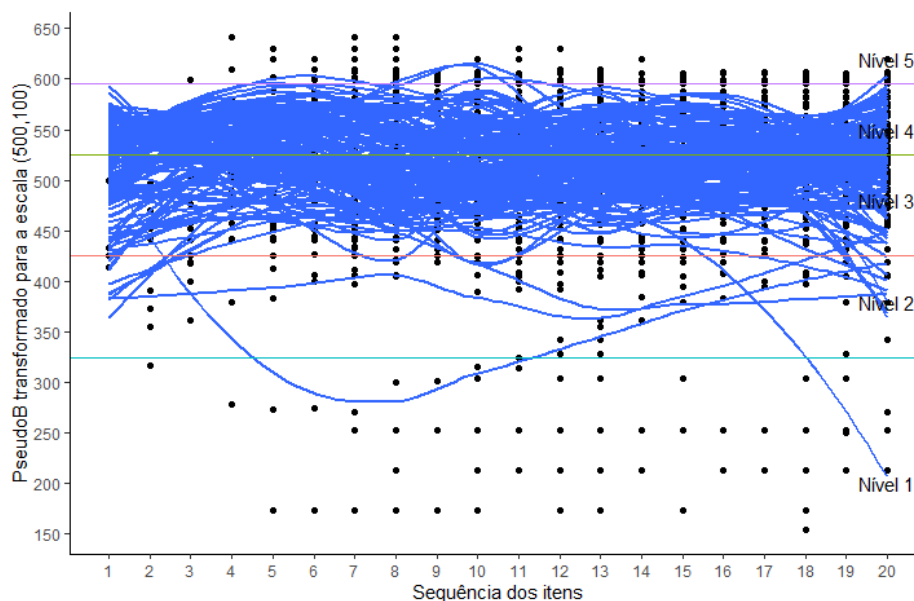
**Gráfico 21** – TAI-Leitura 3.0. Distribuição no Nível 3 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023



**Fonte:** Elaboração própria.

A distribuição dos pseudoB no Nível 3 demonstra que houve uma concentração maior nos Níveis 2 e 3, o que justifica as trajetórias azuis estarem localizadas em sua grande maioria no Nível 3 e, ao final dos 20 itens, se dividirem com o Nível 2. Outrossim, existem pseudoB que apresentam valores próximos do valor 100 na escala de proficiência, o que indica alternativas de itens que podemos considerar elementares, pois os valores para o parâmetro  $b$ , logo, para a alternativa considerada como a correta, são maiores que os do pseudoB. Para o Nível 4, o Gráfico 22 apresenta os pseudoB que foram utilizados na aplicação para os alunos desse nível.

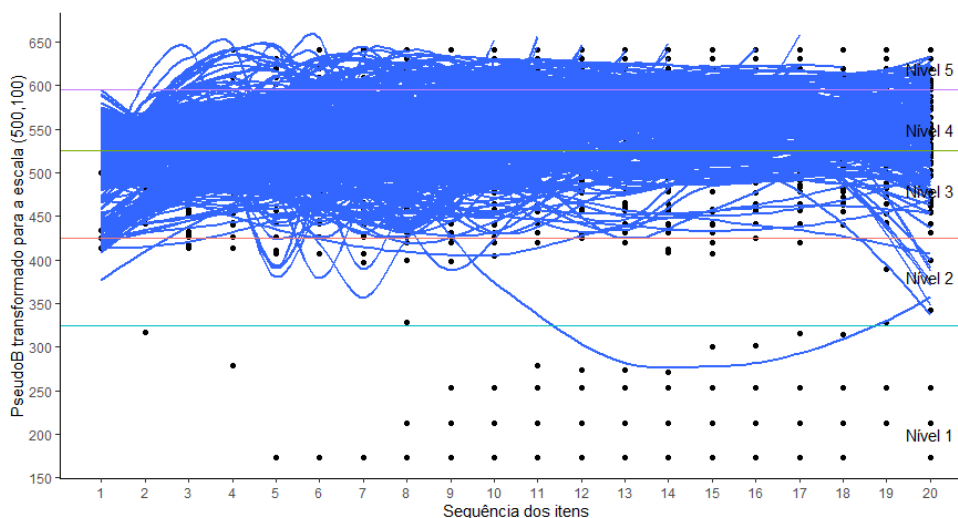
**Gráfico 22** – TAI-Leitura 3.0. Distribuição no Nível 4 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023



**Fonte:** Elaboração própria.

De maneira diferente do que ocorreu nos níveis anteriores, temos uma descontinuidade da concentração de pseudoB entre os Níveis 1 e 2, o que indica que muitos itens apresentam alternativas incorretas concentradas, principalmente, nos Níveis 1 e 3. Ao mesmo tempo, temos a maioria das trajetórias localizadas nos Níveis 3 e 4, fato que demonstra limitações do BI para estimar as proficiências e erro padrão que possam atender ao critério de parada do TAI-Leitura 3.0, quando se espera que a estimativa e o intervalo de confiança fiquem dentro dos limites de um dos níveis. Por fim, para o Nível 5, o Gráfico 23 nos traz os dados da distribuição dos pseudoB dos itens apresentados.

**Gráfico 23** – TAI-Leitura 3.0. Distribuição no Nível 5 dos pseudoB das alternativas em conjunto com regressões de possíveis trajetórias a partir dos dados. Abril 2023

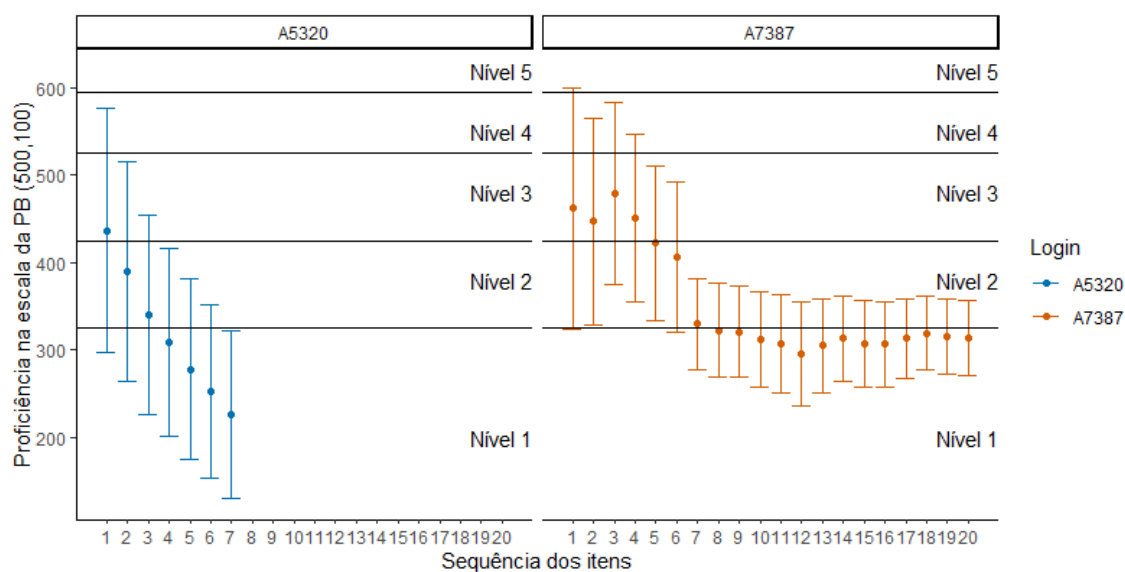


**Fonte:** Elaboração própria.

De maneira semelhante ao que ocorreu no Nível 4, o Gráfico 23 apresenta uma distribuição dos pseudoB dos itens que foram administrados aos alunos classificados no Nível 5, que se concentra mais nos Níveis 3 e 4, além de mostrar uma separação ainda maior entre os pseudoB do Nível 1 e os dos Níveis 3 ao 5, pois, praticamente, não existem pseudoB localizados no Nível 2. De modo geral, as análises por meio da regressão local (*loess*) indicam, em nossa opinião, que os pseudoB podem ser utilizados como indicadores do funcionamento do TAI-Leitura 3.0, pois conseguem demonstrar as trajetórias principais em cada Nível. Entretanto, essas análises também sinalizaram, além do Gráfico 18, existir uma limitação no BI do TAI-Leitura 3.0, principalmente, em conseguir diferenciar alunos que apresentam estimativas de proficiência nos limites dos níveis.

Com base nessa identificação da limitação do BI na variabilidade dos parâmetros das alternativas dos itens, que possibilite estimar as proficiências dos alunos que apresentam medida do traço latente em leitura próxima da divisa de dois níveis. Para tanto, vamos analisar dois casos de cada nível, em que os alunos responderam o maior número de itens possíveis, 20 itens, e a menor quantidade de itens que foram necessários para o TAI-Leitura 3.0 parar em cada nível. Dessa maneira, o Gráfico 24 nos traz as estimativas parciais das proficiências estimadas por resposta a cada item de dois alunos que foram classificados no Nível 1.

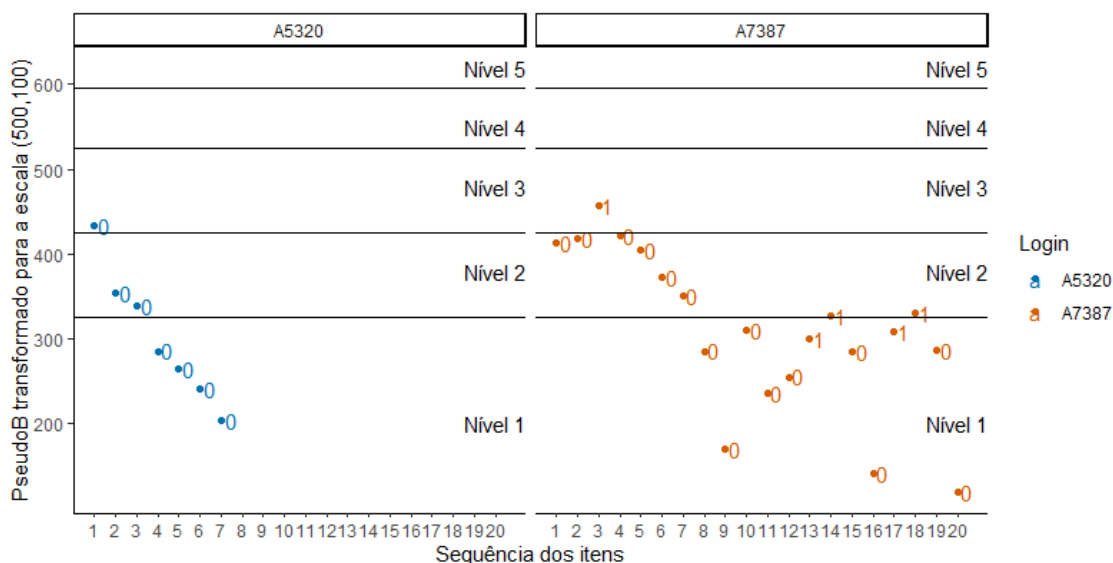
**Gráfico 24** – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 1. Abril 2023



**Fonte:** Elaboração própria.

Podemos observar que os alunos A5320 e A7387 apresentam diferenças significativas nas estimativas parciais, em que o aluno do lado esquerdo do Gráfico 24 tem o teste avançando rapidamente para a finalização, com a estimativa mais o intervalo de confiança dentro dos limites do Nível 1. Todavia, o outro aluno está com as estimativas parciais no limite superior do nível, praticamente entrando no Nível 2, o que dificulta o oferecimento de itens que possibilitem atender ao critério de parada. É importante destacar que a limitação da quantidade de itens pode interferir, pois, como, aparentemente, o algoritmo consegue alcançar uma maior precisão, pois aproveita mais a informação disponível, a limitação em 20 itens pode ser um obstáculo para casos limites. Contudo, a vinculação do projeto com o escopo do que era o instrumento da PB, nos faz estabelecer esse limiar. O Gráfico 25 nos traz a informação do pseudoB de cada item oferecido para cada um dos dois alunos.

**Gráfico 25** – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 1. Abril 2023



**Fonte:** Elaboração própria.

A informação sobre o pseudoB dos itens que foram oferecidos aos dois alunos é completada com a indicação se o item foi “acertado” ou “errado” pelo aluno e, assim, podemos relacionar, além da estimativa das proficiências parciais, o quanto cada aluno escolheu de alternativas corretas ou incorretas. Ademais, é possível relacionar o dado de “acerto” e “erro” com a magnitude de cada intervalo de confiança, o que nos permite analisar se a medida fica mais ou menos precisa conforme o teste avança. No Gráfico 25 nos mostra que o aluno A5320 acaba “errando” todos os itens que são apresentados para ele, o que leva o TAI-Leitura 3.0 a parar no sétimo item, pois o intervalo de confiança e a estimativa da proficiência estavam completamente dentro do Nível 1.

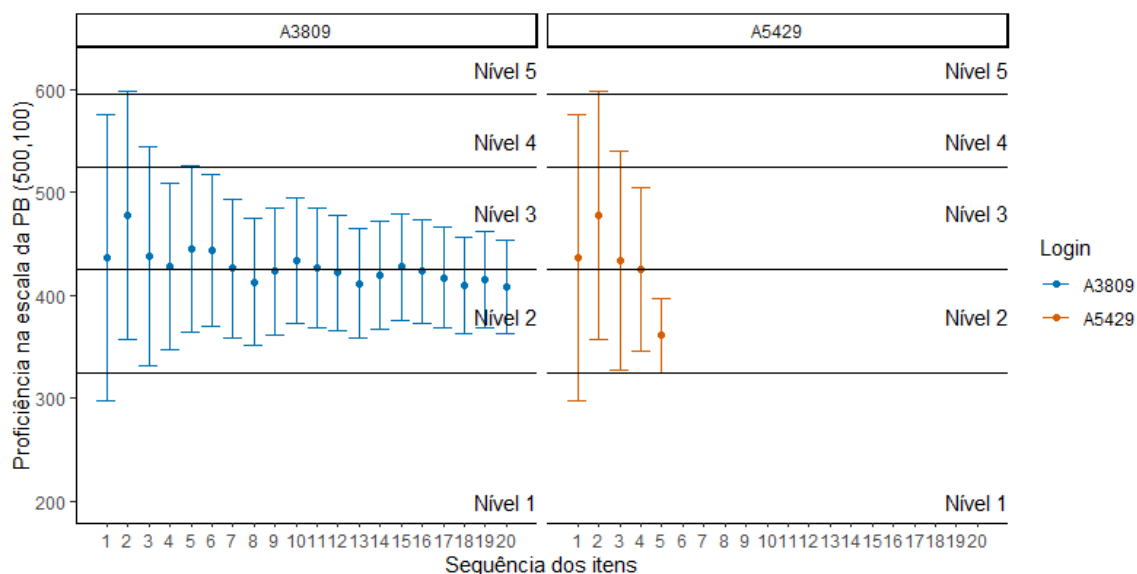
Para o aluno A7387, os 9 itens iniciais do teste apresentam 8 escolhas, por parte do aluno, de alternativas incorretas e isso faz com que as estimativas das proficiências parciais fiquem por 4 itens no Nível 3, pois o terceiro item foi “acertado” pelo aluno. Porém, após esse “acerto”, o aluno “errou” todos até o décimo segundo item, o que fez com que a estimativa da proficiência reduzisse o seu valor até o intervalo do Nível 1. Todavia, os quatro acertos que seguiram ao longo dos sete itens, após o décimo segundo, fizeram com que a estimativa de proficiência permanecesse próxima ao limiar entre o Nível 1 e o Nível 2, alcançando assim a quantidade limite de itens que está configurada no TAI-Leitura 3.0.

Uma diferença que nos parece importante entre as estimativas de proficiências parciais entre os dois alunos é o tamanho do intervalo de confiança. Para o aluno A5320, mesmo com o teste parando em 7 itens, temos uma magnitude de praticamente 200 pontos de proficiência no intervalo de confiança, algo expressivo, pois o aluno pode ter uma medida de proficiência entre 100 e 300 pontos na escala de PB. Contudo, é fundamental frisar que esses valores ficam dentro do Nível 1 e, logo, pedagogicamente o aluno está classificado nesse nível e consegue realizar aquilo que está descrito na interpretação pedagógica desse nível da escala de proficiência.

O aluno A7387, por sua vez, apresenta, nos 9 primeiros itens, uma diminuição no intervalo de confiança na ordem de 180 pontos da escala de proficiência, mesmo “errando” 8 destes itens. Contudo, o intervalo acaba ficando praticamente igual até o final do teste. Além disso, podemos perceber no primeiro item de ambos os alunos, uma diferença nos valores dos pseudoB, o que indica uma escolha de alternativas diferentes, sendo assim, mesmo os dois errando o item, o segundo não foi o mesmo para eles, o que faz com que as estimativas das proficiências parciais sejam diferentes. De modo geral, podemos afirmar que temos muita informação pedagógica desses dois alunos, pois mesmo com uma diferença grande entre as trajetórias nos testes, conseguimos compreender, com base nas respostas deles, o quanto cada um consegue ou não realizar em um teste de leitura.

Nesse mesmo sentido de análise, o Gráfico 26, nos apresenta os dados de dois alunos que foram classificados no Nível 2 e que realizaram, respectivamente, 20 e 5 itens nos testes que responderam. Além disso, há uma oscilação relativamente menor nas estimativas parciais quando comparadas com os dois exemplos do Nível 1.

**Gráfico 26** – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 2. Abril 2023

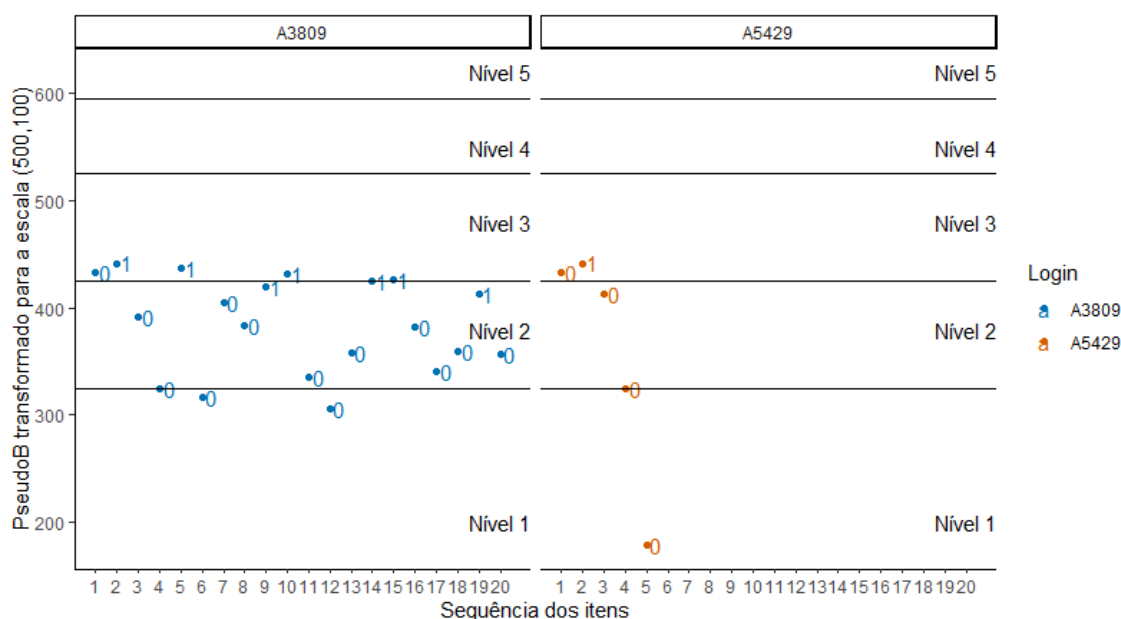


**Fonte:** Elaboração própria.

Em linhas gerais, os dados das proficiências parciais dos dois alunos são iguais até o item 4, a partir do próximo item existe uma diferença entre as estimativas e o aluno A5429 acaba tendo uma diminuição de mais de 50 pontos de proficiência e o intervalo de confiança fica completamente dentro dos limites do Nível 2, o que faz com que o teste pare. Ao mesmo tempo, o aluno A3809 teve as estimativas de proficiências parciais, em diversos momentos dos 20 itens, oscilando entre o Nível 2 e 3, terminando o TAI-Leitura 3.0 com a estimativa próxima de 400 pontos na escala de proficiência da PB, o que faz com que o teste pare no Nível 2, porém, quase 30 pontos ficam dentro do Nível 3. O Gráfico 27, pode nos ajudar a compreender a trajetória no teste desses dois alunos.



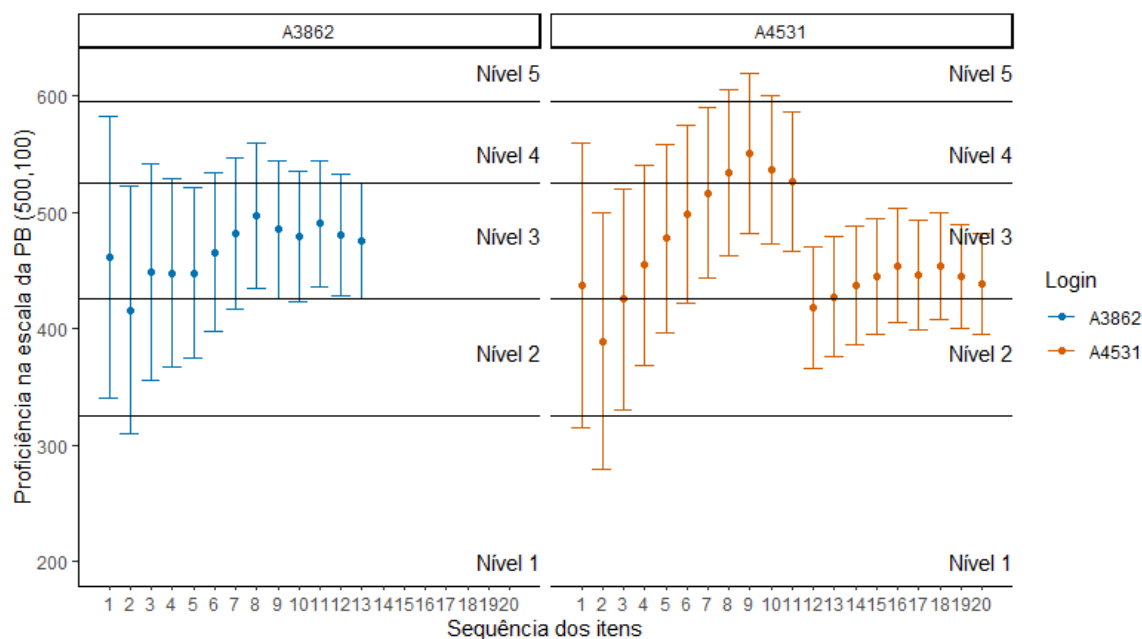
**Gráfico 27** – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 2. Abril 2023



**Fonte:** Elaboração própria.

Pelo fato de que os dois primeiros itens administrados sejam iguais esses alunos, podemos analisar quais foram as respostas, ou melhor, as alternativas que eles escolheram. No terceiro item, que foi igual, existe uma pequena diferença na posição do 0 no gráfico, o que indica que as alternativas assinaladas não foram as mesmas. No quarto item, eles apresentaram uma resposta “errada” com valor do pseudoB igual, o que significa que escolheram a mesma alternativa. O quinto item agiu como o “fiel da balança”, pois o aluno A5429 acaba “errando” e escolhendo uma opção de resposta com um valor muito mais baixo de pseudoB, o que fez com que a estimativa de proficiência e o intervalo de confiança atendessem ao critério de parada e o teste fosse encerrado. Já o aluno A3809 conseguiu “acertar” o item e a estimativa da proficiência parcial aumentou alguns pontos na escala, o que modificou completamente a sua trajetória no teste, ainda assim deixando os valores muito próximos do limite entre os Níveis 2 e 3. O Gráfico 28, apresenta os dados para dois alunos do Nível 3.

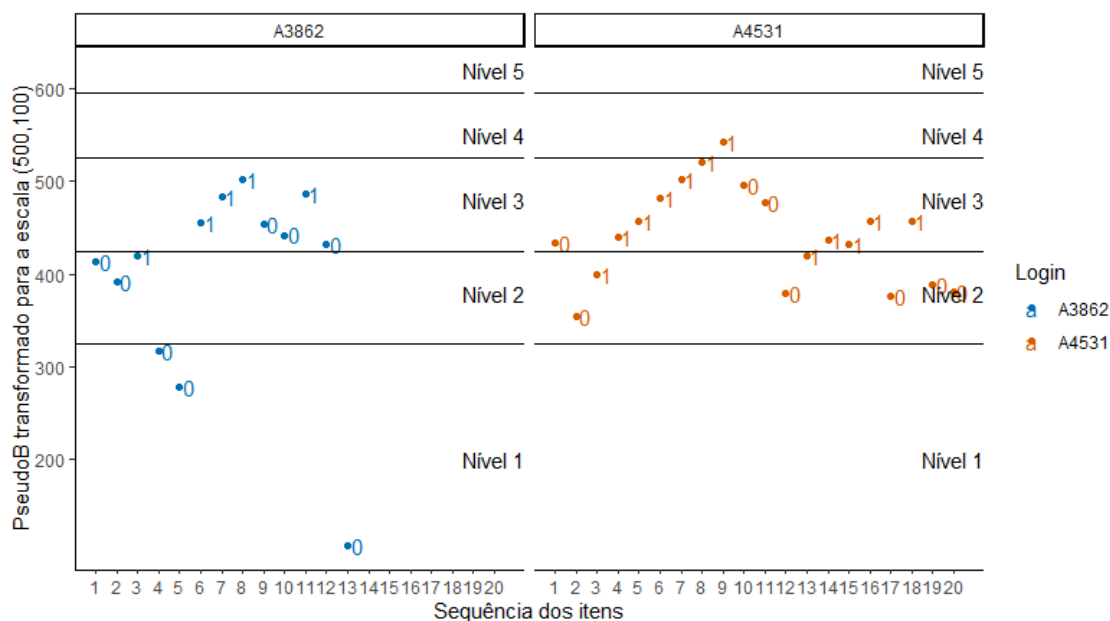
**Gráfico 28** – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 3. 2023



**Fonte:** Elaboração própria.

No Gráfico 28 temos alguns pontos que são diferentes dos dois casos anteriores, pois, além de termos mais itens administrados para o aluno que respondeu menos itens entre os dois, ocorre uma queda acentuada na estimativa da proficiência parcial do aluno A4531 entre os itens 11 e 12, algo entorno de 100 pontos. Além disso, o intervalo de confiança para este aluno fica entre os Níveis 2 e 3, mesmo com a estimativa de proficiência, no último item possível no teste, estando dentro dos limites do Nível 3. Para o aluno A3862, foram aplicados 13 itens, com um aumento dos valores de proficiência até o item 8 e uma estabilização das estimativas até o final do teste. No Gráfico 29 temos os pseudoB dos itens que foram administrados para os dois alunos.

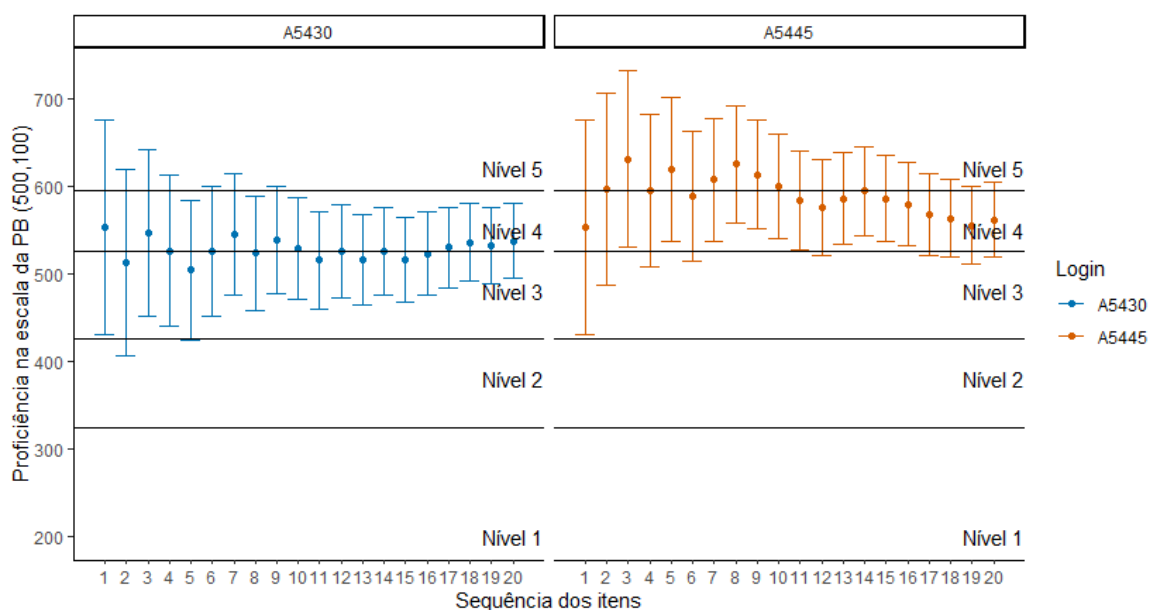
**Gráfico 29** – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 3. Abril 2023



**Fonte:** Elaboração própria.

É possível observar que existe um aumento considerável nos itens que os alunos “acertam” em comparação com os exemplos anteriores, o que indica uma maior capacidade dos dois alunos em enfrentar um teste de leitura, porém, dentro das possibilidades de quem está classificado no Nível 3. Ademais, podemos perceber os diversos momentos nos dois testes, o que ocorre com a sequência de acertos dos itens, em que as estimativas das proficiências parciais aumentam e os pseudoB acompanham. Isso indica quanto o TAI-Leitura 3.0 consegue se adaptar conforme a estimativa de proficiência aumenta e a exigência também. O Gráfico 30, nos traz a sequência de proficiência de dois alunos classificados no Nível 4.

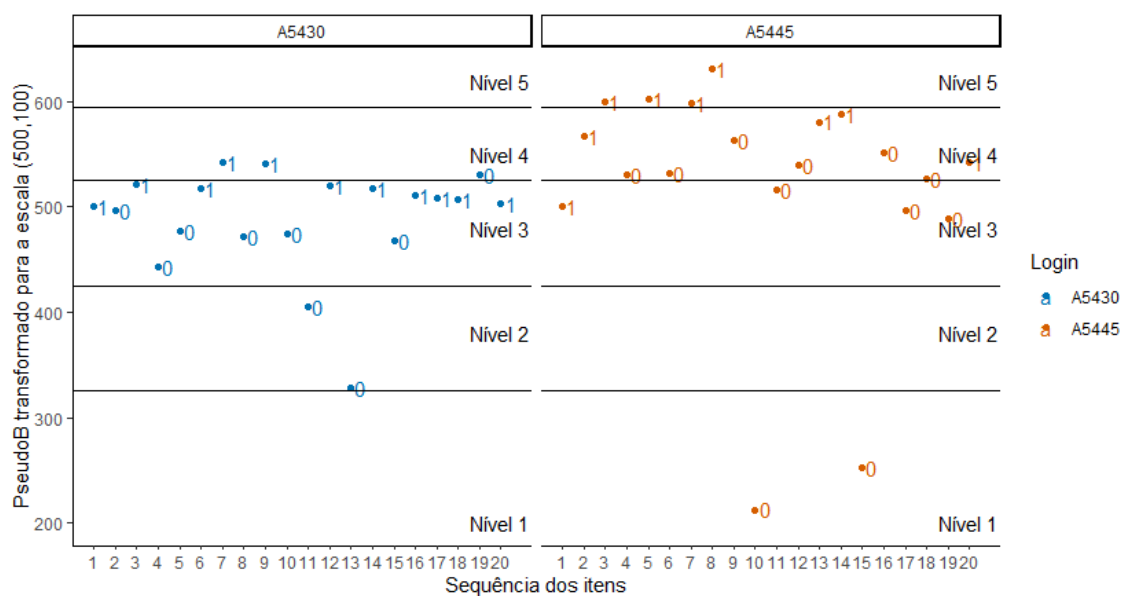
**Gráfico 30** – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 4. Abril 2023



**Fonte:** Elaboração própria.

O ponto principal que podemos identificar no Gráfico 30 é que nenhum dos dois alunos respondeu menos que 20 itens, algo que ocorreu para todos aqueles que foram classificados no Nível 4. O que nos leva a considerar o quanto as limitações de um BI podem trazer para um TAI. Da mesma forma, é interessante observar que o aluno A5445 tem, após a aplicação do item 20, a estimativa da proficiência com o intervalo de confiança muito próximo de ficar dentro dos limites do Nível 4. O Gráfico 31 apresenta os valores dos pseudoB e a indicação de “acerto” ou “erro” de cada um dos dois alunos do Nível 4.

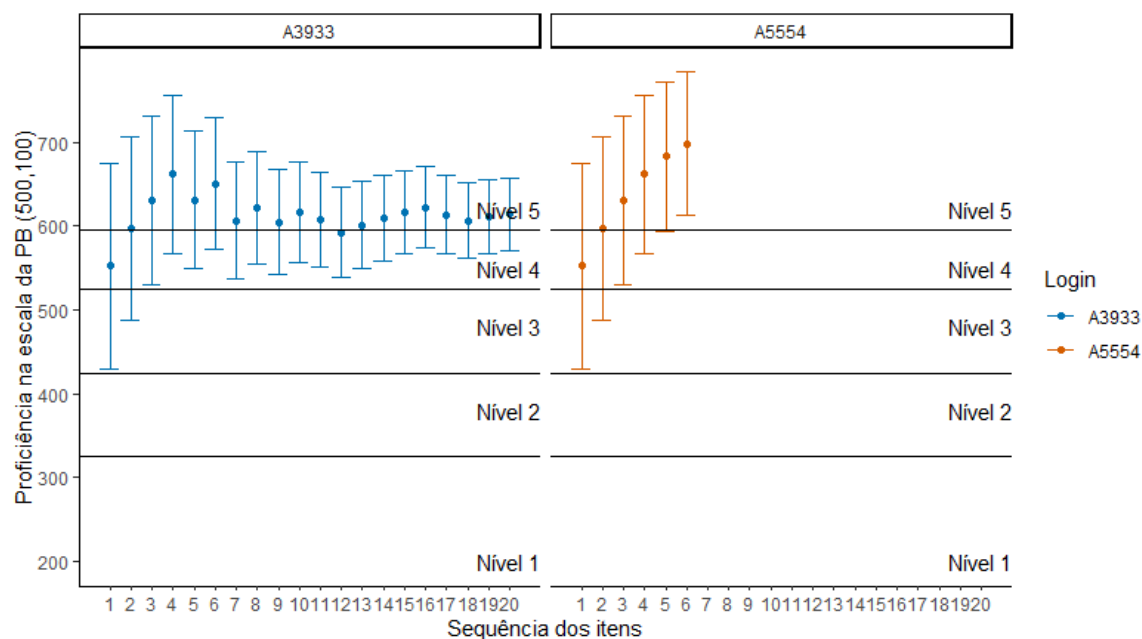
**Gráfico 31** – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 4. Abril 2023



**Fonte:** Elaboração própria.

Podemos identificar um novo aumento na quantidade de “acertos” dos dois alunos no Nível 4, como a maioria dos pseudoB estão localizados nos Níveis 3 e 4, o que mostra que os itens apresentam uma maior exigência. Além desse fato, é possível observar, na sequência de “acertos” entre os itens 16 e 18 do aluno A5430, uma diminuição gradativa do valor do pseudoB, algo diferente do que aconteceu no exemplo dos dois alunos do Nível 4. O Gráfico 32 nos traz os dados das proficiências de dois alunos que foram classificados no Nível 5.

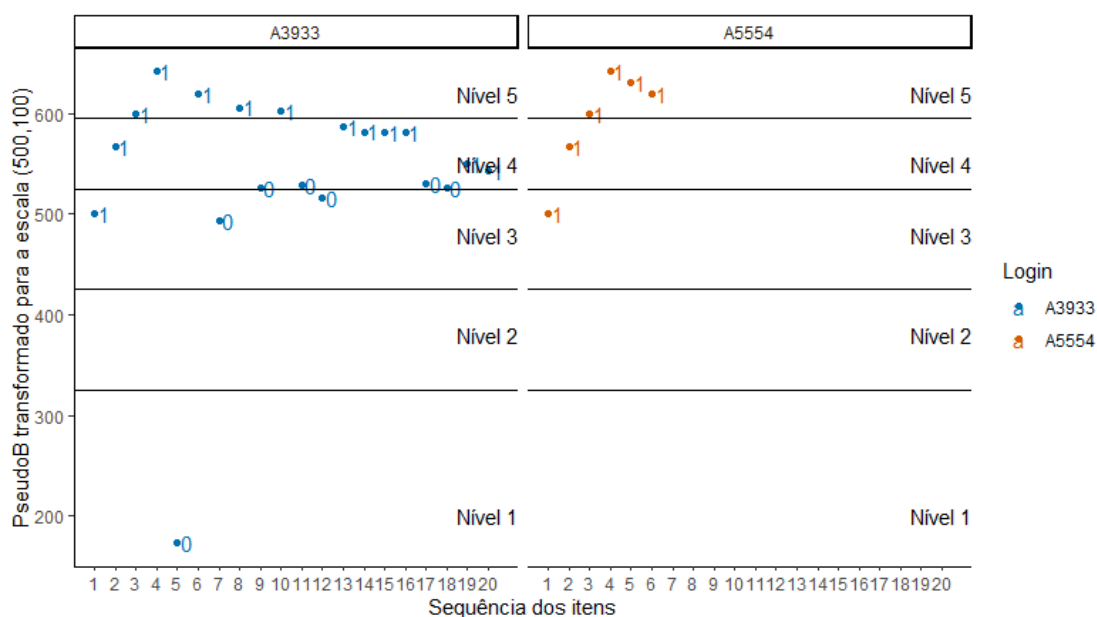
**Gráfico 32** – TAI-Leitura 3.0. Proficiência e intervalo de confiança para dois alunos com proficiências no Nível 5. Abril 2023



**Fonte:** Elaboração própria.

É possível observar um aumento rápido das estimativas de proficiência nos dois casos, em que no segundo item a medida já está no Nível 5, além de serem respondidos os mesmos itens por ambos os alunos até o item 4. Após isso, o aluno A3933 tem uma queda na estimativa de proficiência parcial e continua na sequência com valores próximos ao limite entre os Níveis 4 e 5. Já o aluno A5554 responde somente 6 itens até que a estimativa de proficiência mais o intervalo de confiança estivessem totalmente enquadrados acima do limite inferior do Nível 5. Além disso, podemos destacar o fato de não termos uma maior precisão para a medida deste último aluno, pois como rapidamente é alcançado o valor suficiente para atender o critério de parada, não conseguimos saber o quanto mais acima essa estimativa poderia estar, porém, se acrescentarmos a parte superior do intervalo de confiança, teremos um valor próximo de 800 pontos na escala. O Gráfico 33 nos mostra as informações de “acerto” e “erro” para cada item respondido pelos dois alunos e os valores para os respectivos pseudoB.

**Gráfico 33** – TAI-Leitura 3.0. PseudoB dos itens respondidos por dois alunos com proficiências no Nível 5. Abril 2023



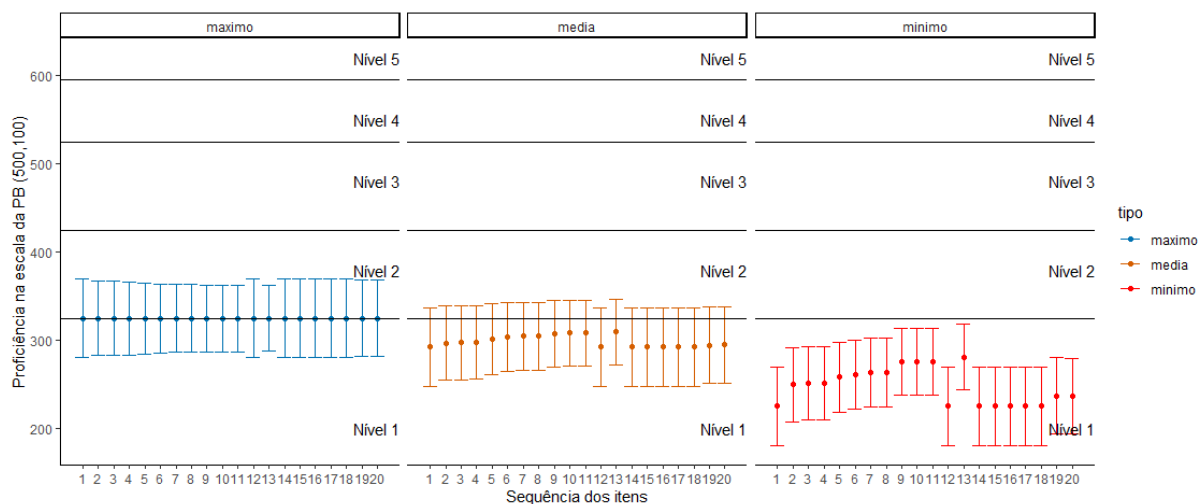
**Fonte:** Elaboração própria.

Temos um aumento na quantidade de acertos dos dois alunos, em que mesmo tendo que responder os 20 itens, o aluno A3933 consegue acertar 13 dos 20 itens. Ademais, conseguimos identificar um fenômeno que ocorre em ambas as aplicações: um aumento do pseudoB até o item 4, depois há uma queda, mesmo com um e o outro aluno acertando o quinto item. Provavelmente, isso ocorre devido à limitação superior do BI, onde o item com maior valor de parâmetro  $b$  foi utilizado no quarto item, e não havia, dentre os possíveis itens a serem utilizados, itens com parâmetros mais elevados. O que também acaba causando uma não diminuição do intervalo de confiança, mesmo para o aluno que responde a 20 itens no teste.

Longe de exaurir as possibilidades de análise, os exemplos para cada um dos níveis nos proporcionaram compreender como foi o funcionamento do TAI-Leitura 3.0, além de permitir identificar em quais pontos surgem os gargalos que são impostos, principalmente, pela quantidade e variabilidade dos parâmetros dos itens que compõem o BI. Nesse sentido, ampliamos as análises para verificarmos como se dão as trajetórias em cada nível quando consideramos a média de proficiência e os valores máximo e mínimo de proficiência em cada item do teste. Para tanto, consideramos como se todos os 20 itens fossem aplicados, pois, assim, conseguimos analisar todo o espectro de possibilidades do TAI-Leitura 3.0 em cada nível, além

de utilizarmos a média do erro padrão para construir os intervalos de confiança em cada item. O Gráfico 34 nos traz a distribuição das proficiências para cada agregação no Nível 1.

**Gráfico 34** – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 1. Abril 2023

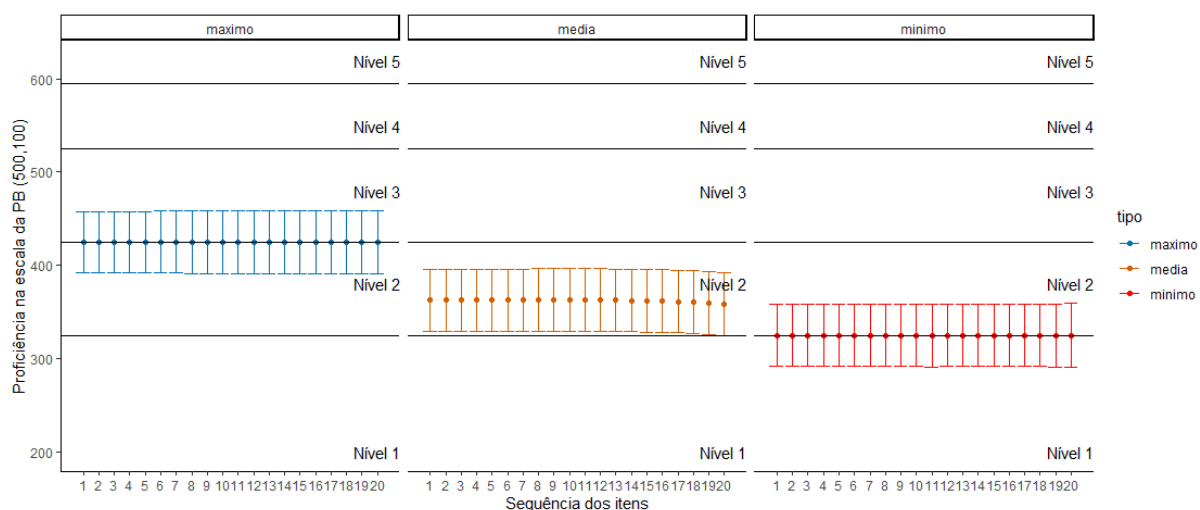


**Fonte:** Elaboração própria.

No Gráfico 34, podemos identificar que para as médias de proficiências parciais em cada item, temos valores muito próximos ao limite superior do Nível 1, mas com intervalos de confiança um pouco avançados para dentro do Nível 2. Para os valores máximos e mínimos de proficiências, temos valores constantes para o primeiro e uma certa oscilação para o segundo, com o destaque para o fato de as medidas para os valores mínimos estarem totalmente dentro do limite do Nível 1. Da mesma maneira, o Gráfico 35 nos mostra a mesma proposta de análise, mas para o Nível 2.



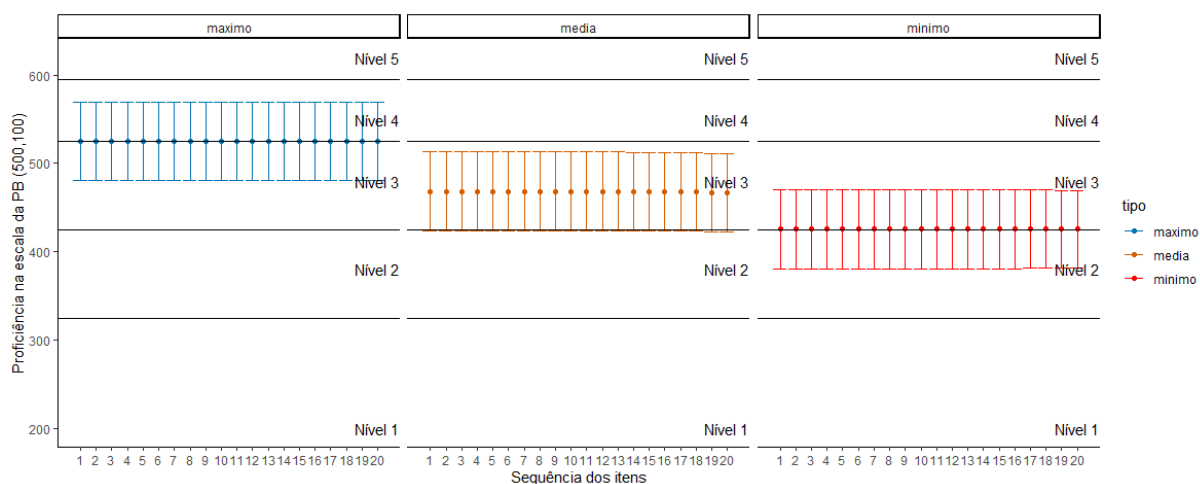
**Gráfico 35 – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 2. Abril 2023**



**Fonte:** Elaboração própria.

Diferente do que ocorreu nos dados do Nível 1, no Gráfico 35 conseguimos verificar uma diferença na forma como os dados se dispõem, pois temos os valores médios limitados pelas extremidades do nível e os máximos e mínimos com estimativas completamente nos limites do Nível 2. Um fato importante é que as médias, mesmo estando completamente dentro do intervalo, têm a sua concentração na parte inicial do nível, com os limites inferiores dos intervalos de confiança praticamente junto à divisa. O Gráfico 36 apresenta os valores para o Nível 3.

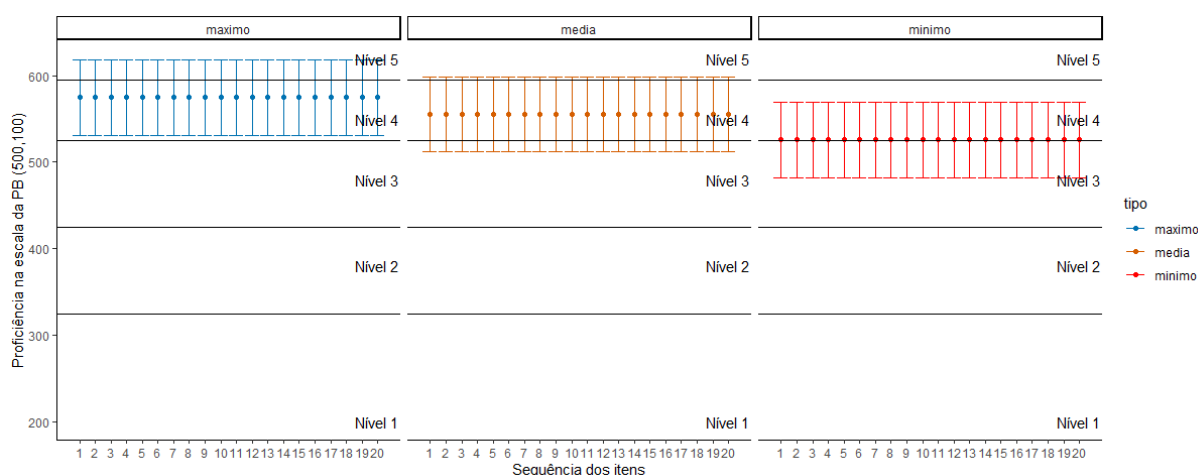
**Gráfico 36 – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 3. Abril 2023**



**Fonte:** Elaboração própria.

De maneira similar ao que ocorreu com os dados do Nível 2, temos um posicionamento dos valores máximo e mínimo das proficiências nos limiares do Nível 3 com os seus sucessor e antecessor; temos também, e isso é algo diferente das análises anteriores, um aumento do intervalo de confiança em cada item; logo, no caso dos valores médios, temos quase um preenchimento de todo o intervalo do Nível 3. O Gráfico 37 nos traz os dados para os alunos do Nível 4.

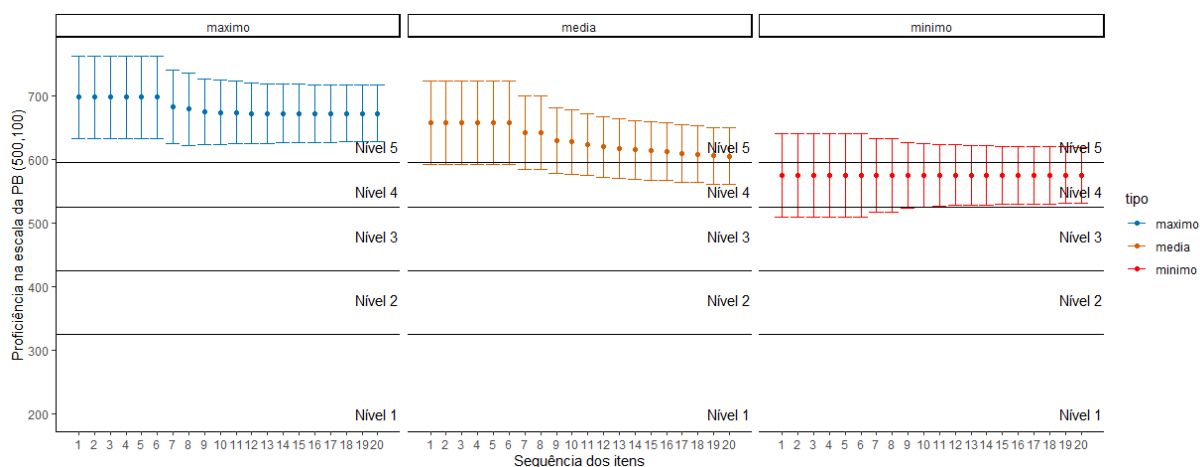
**Gráfico 37** – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 4. Abril 2023



**Fonte:** Elaboração própria.

O mesmo fenômeno que identificamos com os dados do Nível 3, ocorre com os valores médios do Nível 4, porém com o intervalo de confiança maior que todos os limites do nível, algo que faz com que o TAI-Leitura 3.0 tenha dificuldades de parar o teste como maior precisão neste nível. De todo modo, é perceptível o aumento do erro padrão, o que reforça essa incerteza em classificar os alunos neste nível, como também, em oferecer itens mais adequados para este grupo de alunos. Por fim, o Gráfico 38 nos mostra os valores máximo, mínimo e médio por item para os alunos classificados no Nível 5.

**Gráfico 38** – TAI-Leitura 3.0. Distribuição das proficiências e intervalo de confiança no máximo, na média e no mínimo de proficiência do Nível 5. Abril 2023



**Fonte:** Elaboração própria.

É possível observar uma mudança em relação aos dados que analisamos dos níveis anteriores, pois existe uma proximidade maior das medidas, além de termos os valores máximos contidos dentro do Nível 5. Todavia, é um fato já explorado neste trabalho, que o BI acabou limitando as medidas de proficiência na parte superior da escala e, temos, portanto, um teto que não permitiu que muitos dos valores máximos fossem maiores do que os apresentados. Além disso, conseguimos notar uma diminuição das estimativas de proficiências médias, o que também pode sinalizar a limitação superior do BI, pois, por termos uma falta de itens mais “difíceis”, o algoritmo começa a oferecer itens que estão abaixo da estimativa de proficiência parcial ao responder o item anterior, sendo assim, ao invés de termos o avanço das proficiências uma perda é observada.

De qualquer maneira, esse fato relacionado ao BI não se restringe ao TAI-Leitura 3.0, pelo contrário, qualquer algoritmo que foi implementado com o uso desse banco, encontrará as mesmas limitações. De todo modo, temos um TAI que consegue entregar estimativas de proficiência, classificar os alunos nos Níveis e possibilitar uma interpretação pedagógica que vai além do que a PB oferecia. Nesse sentido, no subcapítulo a seguir, iremos abordar uma possibilidade de realizarmos uma interpretação pedagógica dos resultados individuais, incluindo as escolhas das alternativas corretas e incorretas para cada item.

### 5.1 – Interpretação pedagógica: uma possibilidade de análise das alternativas em um TAI-Leitura

Para ilustrarmos o que estamos propondo como sendo uma possibilidade de interpretação pedagógica individualizada, ou seja, considerarmos a trajetória do aluno no TAI-Leitura 3.0 para incrementar a interpretação pedagógica do nível, vamos utilizar os dados dos dois alunos apresentados no exemplo do Nível 2 e, para tanto, vamos retomar alguns pontos importantes da interpretação pedagógica da PB. Como abordamos no capítulo 1, a PB trazia em seu material, disponibilizado para as professoras e os professores, um guia que informava como as “correções” das provas deveriam ser feitas, além de apresentar informações pedagógicas do que cada nível conseguia fazer em relação ao traço latente medido. Dessa maneira, se um aluno tivesse uma quantidade de acertos que o classificasse como sendo do Nível 2, por exemplo, a professora ou o professor podiam realizar a leitura da interpretação pedagógica desse nível, o que indicaria, muito provavelmente, o que o aluno conseguia fazer em relação à leitura. A Figura 13 nos traz a interpretação pedagógica para o Nível 2 da PB.

**Figura 13** – Provinha Brasil – Leitura. Interpretação pedagógica do Nível 2

Nível 2
de 4 a 9 acertos
<p>Os estudantes que se encontram neste nível, além de, provavelmente, já terem consolidado as habilidades do nível anterior, geralmente já podem:</p> <ul style="list-style-type: none"> <li>• reconhecer palavras de formação silábica canônica escritas de diferentes formas;</li> <li>• estabelecer relação entre grafemas e fonemas, identificando, por exemplo, a letra ou a sílaba inicial de uma palavra;</li> <li>• ler palavras formadas por sílabas canônicas e não canônicas.</li> </ul>

**Fonte:** Inep (2016, p. 26).

Com base nessas informações, se retomarmos os resultados dos alunos A5429 e A3809 apresentados no Gráfico 26, vamos relembrar que os valores dos pseudoB para os dois primeiros itens foram iguais e para o terceiro houve uma pequena diferença, com um valor um pouco maior para o pseudoB da alternativa escolhida pelo aluno A5429. Porém, para o item 4, temos o valor igual para os dois, o que nos leva a considerar, sem consultarmos se de fato é o que ocorreu, que os itens são iguais e ambos escolheram a mesma alternativa. Isso fez com que o item 5 fosse exatamente o mesmo para eles, contudo o aluno A3809 escolhe a alternativa “correta” e recebe um item diferente, pois o aluno A5429 acabou escolhendo uma alternativa

considerada “incorreta” para o item. Desse ponto em diante houve uma grande diferenciação entre os dois testes, pois um continuou recebendo itens até o vigésimo e o outro, o TAI-Leitura 3.0 parou após o sexto item.

O que queremos nesta parte do trabalho não é esgotar as possibilidades de como poderia ser feita a interpretação pedagógica individual, pelo contrário, estamos propondo o início de análise que toma como ponto de apoio a possibilidade que o TAI-Leitura 3.0 tem de individualizar os resultados, ao mesmo tempo, diferenciando as potencialidades e dificuldades de cada aluno. Nesse sentido, a Tabela 22 nos mostra a informação dos 5 primeiros itens que esses dois alunos que estamos analisando responderam.

**Tabela 22** – TAI-Leitura 3.0. Informações pedagógicas dos 5 primeiros itens dos testes dos alunos A5429 e A3809. Abril 2023

Habilidade do item	A5429	A3809
Reconhecer o assunto de um texto com base em seu título.	Consegue localizar uma informação explícita no texto, pois localiza a palavra	
Identificar uma palavra que se repete em sequência de quatro ocorrências de nomes próprios, representada graficamente com letras de diferentes tipos.	Consegue identificar o mesmo nome repetido quatro vezes com tipos de letras diferentes	
Identificar a frase que corresponde à imagem	Não consegue identificar a frase que corresponde à imagem, pois troca a palavra LIVRO por LITRO e a ação que está sendo executada na imagem.	
Reconhecer, a partir de uma imagem, o valor sonoro de uma sílaba.	Não conseguem distinguir o valor sonoro da sílaba "TA" de "LA".	
Localizar informação explícita em textos	Não consegue localizar informação explícita em uma tirinha, identificando um dos elementos da tirinha como a informação procurada.	Consegue localizar informação explícita em uma tirinha.

**Fonte:** Elaboração própria.

É importante destacar, antes de analisarmos as informações da Tabela 22, que esses dois alunos responderam exatamente os cinco primeiros itens, sendo assim, conseguimos identificar os pontos nos quais eles se diferenciam pedagogicamente. Podemos observar que para os quatro primeiros itens temos uma agregação da nossa interpretação pedagógica do que eles conseguiram responder nos itens apresentados, pois as escolhas foram iguais ou muito semelhantes. Dessa maneira, quando consideramos o primeiro item, que pode ser observado na Figura 3, temos um poema com três estrofes e a primeira alternativa com uma palavra grafada que aparece no primeiro verso e, sendo assim, esses dois alunos conseguem localizar uma palavra em um texto, pois fazem isso ao localizar a palavra “MACACO”, porém ainda não conseguem reconhecer o assunto do texto por meio do título.

Por terem escolhido essa alternativa, associada à menor proficiência do item, o item que foi apresentado na sequência solicita que os alunos consigam identificar uma sequência de quatro ocorrências de nomes próprio e que se apresentem com tipos de letras diferentes. Assim, esses alunos conseguem identificar o mesmo nome próprio que se repete de maneiras diferentes e, logo, acertam o item. Isso faz com que a complexidade do próximo item aumente, pois a exigência deixa de ser somente palavra. Dessa maneira, é solicitado que os alunos identifiquem a frase que corresponde à imagem, o que exige que eles leiam as frases que se apresentam nas alternativas. Os dois alunos escolhem alternativas diferentes, porém, ambas apresentam a mesma palavra “LITRO” no lugar de “LIVRO”.

Como a diferença entre as alternativas é mínima, o item apresentado na sequência foi o mesmo. Era solicitado que o aluno soubesse reconhecer, a partir de uma imagem, o valor sonoro de uma sílaba e ambos escolheram a mesma alternativa, trocando a sílaba “TA” por “LA”. O que pode corroborar com a troca de letras que ocorreu no item anterior. De qualquer maneira, por terem respondido a mesma alternativa, o quinto item do teste foi igual, no qual era solicitada a localização de informação explícita em textos, o que nos remete ao primeiro item, pois eles conseguiram localizar uma palavra que estava na primeira alternativa e no primeiro verso. Todavia, o aluno A5429 não consegue localizar a informação explícita em uma tirinha pequena, pois ele identifica o nome da representação de um objeto bastante utilizado na cozinha, o “FOGÃO”, mas não conseguiu compreender o que estava de fato ocorrendo no quadrinho.

A ideia não é trocarmos a interpretação pedagógica da escala de proficiência por algo individual, pois isso não tem nenhum sentido de existir como tal, porém é sabido que a

utilização de meios digitais nos possibilitam avançar para processos que com o papel seriam custosos ou praticamente impossíveis. Dessa maneira, ao realizarmos um trabalho que busca analisar as escolhas dos alunos em cada item respondido, abrimos uma possibilidade de começarmos a agregar informações individuais, nos aproximando de algo mais processual, em que o resultado tem a sua importância, porém a trajetória de cada um ao longo do teste também deve ser considerada. Assim, o que queremos é iniciar as discussões sobre as possibilidades de interpretar pedagogicamente os “acertos” e “erros” de cada aluno que responda o TAI-Leitura 3.0.

## Considerações finais e perspectivas futuras

Nesta Dissertação, tendo em vista o problema de pesquisa e os passos dados para enfrentá-lo, apresentamos o desenvolvimento de um TAI-Leitura, em sua versão TAI-Leitura 3.0, com resultados processados estatisticamente pela TRI, como um instrumento para estimar a proficiência em leitura, condição para sua avaliação, considerando, para isso, todas as alternativas que compõem um item de múltipla escolha, tipo de item indicado para a tarefa de dimensionar e avaliar essa proficiência. Isso faz com que se possa incorporar as respostas em todas as alternativas de cada item no processamento em um TAI, inclusive para que se possa considerar as alternativas em seus processos de escolha do próximo item, de parada do teste e, sobretudo, de estimativa de proficiências com qualidade superior em relação aos testes em papel tradicionalmente aplicados e a outros TAI.

Nesse sentido, a partir dos dados das aplicações da Provinha Brasil na RME-SP, no período de 2012 a 2015, além dos dados do pré-teste de itens em 2014, definimos o escopo de dados com os quais iríamos analisar e qual modelo de tratamento de respostas poderia contribuir com o objetivo deste trabalho. O primeiro passo foi a definição dos modelos que possibilitam a análise das alternativas pela TRI e, sendo assim, o levantamento realizado, dentre alguns modelos disponíveis, permitiu-nos identificar os modelos Nominal e Aninhado (*Nested Logit*) como sendo os que mais poderiam contribuir com o nosso objetivo.

A base de dados da primeira aplicação de 2015 da PB na RME-SP foi utilizada para os processamentos pela TRI e os resultados dos modelos Rasch, 2PL, Nominal e Aninhado foram cotejados entre si, para que pudéssemos verificar se os ajustes dos modelos politômicos eram melhores, ou pelo menos, similares com os dicotômicos. Utilizando o Critério de Informação de Akaike (AIC) e o Critério Bayesiano de Schwarz (BIC), encontramos valores um pouco melhores para o modelo Rasch, porém, por conta da quantidade de parâmetros que é maior para os modelos politômicos, a diferença foi considerada como desprezível ou muito pequena, o que nos fez continuar com as análises envolvendo somente o modelo Aninhado, pois este permite que a alternativa correta seja indicada, o que mantém a proximidade com os modelos logísticos da TRI.

Com base nos modelos da TRI selecionados, estimamos os parâmetros dos 120 itens e utilizamos para compor o Banco de Itens das simulações dos TAI-Leitura 3.0. Esses, por sua vez, foram elaborados com base na utilização do modelo Aninhado no algoritmo de estimativa



da proficiência parciais dos alunos. Todavia, como forma de ampliarmos a potencialidade do uso das alternativas, elaboramos uma maneira de considerar a “dificuldade” de cada uma delas. Para tanto, cotejamos as CCI de cada uma das alternativas para identificar o ponto de intersecção entre aquelas que consideramos como incorretas e a correta, ponto esse que definimos como sendo o pseudoB. Dessa maneira, conseguimos desenvolver um algoritmo que considera os parâmetros pelo modelo Aninhado na estimativa das proficiências e outro que estima as proficiências da mesma maneira, porém escolhe os itens com base no pseudoB.

As simulações contemplaram os algoritmos dos dois modelos de TAI-Leitura 3.0 e, além disso, realizamos, ao longo das simulações, o cotejamento do algoritmo de estimativa das proficiências desenvolvido neste trabalho com os resultados proveniente da utilização da função *fscore* do pacote *mirt*. Dessa maneira, obtivemos uma maior certeza de que os algoritmos desenvolvidos conseguiram atender aos requisitos de precisão e confiabilidade que um dos pacotes mais utilizados em linguagem R para análise pela TRI pode proporcionar.

Definidas as estruturas do TAI-Leitura 3.0, realizamos a implementação no sistema do TAI-Leitura que está hospedado no serviço de nuvem da USP. Após esse processo de ajustes do sistema para funcionar conforme a lógica do novo TAI, o sistema foi posto à prova com a aplicação do pré-teste para o 3º, 4º e 5º ano de uma escola. A análise dos dados dessa aplicação nos mostrou que o TAI-Leitura 3.0 poderia ser utilizado como um instrumento para levantar dados da competência leitura no Ensino Fundamental, principalmente, por conseguir distinguir o desempenho dos alunos nos níveis da PB.

Com base nestes resultados do pré-teste, a aplicação geral para todos os anos escolares de três escolas parceiras do projeto no bojo do Gepave contribuiu para que tivéssemos dados de mais de 2 mil alunos e, assim, fosse possível ampliar os testes de capacidade de respostas dos algoritmos, mas também mostrou as limitações que ainda precisam ser superadas, principalmente, com a amplitude e variabilidade do Banco de Itens.

Dessa maneira, temos como desafios para serem superados: a ampliação da quantidade de itens no BI; a variabilidade de itens em relação a escala de proficiência, pois isso demonstrou afetar diretamente as estimativas de proficiência, como também, as possibilidades de interpretação pedagógica de cada aluno; a melhoria da metodologia para considerarmos a interpretação pedagógica individualizada, algo que se distancia do que é comumente praticado no universo

das avaliações externas em larga escala, considerando que essa individualização, ao nosso ver, ganha potencialidade com a implementação de um TAI que considera as alternativas.

Portanto, temos diversas possibilidades para continuidade desta pesquisa, além do próprio aprimoramento dos algoritmos do TAI-Leitura 3.0 e dos meios de interpretação pedagógica do desempenho dos alunos. Processo que poderia ser expandido para outras competências, sobretudo destacando a disponibilidade do teste de Matemática da PB, incluindo as análises das alternativas. De maneira geral, os resultados gerais dessa pesquisa demonstram que o TAI pode ser uma ferramenta para a avaliação da aprendizagem realizada no dia a dia da escola, principalmente, como um auspicioso instrumento de mensuração no âmbito da avaliação educacional, notadamente porque deu mostras do potencial de se adaptar a cada aluno e identificar sua proficiência, uma condição para o sucesso escolar.

## Referências

- ALAVARSE, Ocimar Munhoz; CATALANI, Érica Toledo; MENEGHETTI, Douglas de Rizzo; TRAVITZKI, Rodrigo. Teste adaptativo informatizado como recurso tecnológico para alfabetização inicial. In: **MEMORIAS de la Séptima Conferencia Iberoamericana de Complejidad y Cibernética (CICIC 2017)**. Winter Garden, FL: International Institute of Informatics and Systemics (IIS), March 2017. p. 68-78.
- ALAVARSE, Ocimar Munhoz; CATALANI, Érica Maria Toledo; SANTOS, Ailton Carlos; COSTA, Thiago Fernando Ferreira. **A formação de professores para avaliação da leitura no processo de alfabetização: uma experiência a partir da construção de um teste adaptativo informatizado (TAI)**. Trabalho apresentado na X Reunião da Abave “Avaliação educacional no Brasil: o desafio da qualidade. São Paulo, 28 a 30 ago. 2019.
- ALAVARSE, Ocimar Munhoz ; MELO, Wolney Candido de. Avaliação educacional e testes adaptativos informatizados (TAI): desafios presentes e futuros, In: COMITÊ GESTOR DA INTERNET – CGI.br. **Pesquisa sobre o uso das tecnologias de informação e comunicação no Brasil – TIC Educação 2012**, Coord. Alexandre F. Barbosa, São Paulo: Comitê Gestor da Internet no Brasil, 2013, p. 103-112.
- ANDRADE, Dalton Francisco de; TAVARES, Héilton Ribeiro; VALLE, Raquel da Cunha. **Teoria da Resposta ao Item: conceitos e aplicações**. São Paulo: Associação Brasileira de Estatística, 2000.
- BARRADA, Juan Ramón. Tests Adaptativos Informatizados: una perspectiva general. **Anales de Psicología**, v. 28, n. 1, p. 289-302, ene. 2012.
- BENNETT, Randy Elliot et al. **Toward a framework for constructed-response items**. Princeton, NJ: Education Testing Service, June 1990.
- BOCK, R. Darrell. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. **Psychometrika**, v. 37, n. 1, p. 29-51, Mar. 1972.
- BOLT, Daniel M.; SUH, Youngsuk. Nested logit models for multiple-choice item response data. **Psychometrika**, v. 37, n. 1, p. 457, Sept. 2010.
- BONAMINO, Alicia; SOUSA, Sandra Zákia. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. **Educação e Pesquisa**, São Paulo, v. 38, n. 2, p. 373-388, abr./jun. 2012.
- BRASIL. Portaria Normativa nº 10, de 24 de abril de 2007. Institui a Avaliação de Alfabetização “Provinha Brasil”. **Diário Oficial da União**, Brasília, Seção 1, 26 abr. 2007.
- BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Nacionais Anísio Teixeira (Inep). **Guia de interpretação de resultados: Provinha Brasil – Leitura, Teste 2, 2009**. Brasília: MEC; Inep, 2009.
- BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Nacionais Anísio Teixeira (Inep). **Guia de aplicação: Provinha Brasil – Leitura, Teste 2, 2011**. Brasília: MEC; Inep, 2011.
- BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Nacionais Anísio Teixeira (Inep). **Guia de elaboração de itens: Provinha Brasil, 2012**. Brasília: MEC; Inep, 2012.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Nacionais Anísio Teixeira (Inep). **Guia de aplicação**: Provinha Brasil – Leitura, Teste 2, 2015. Brasília: MEC; Inep, 2015.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Nacionais Anísio Teixeira (Inep). **Guia de interpretação de resultados**: Provinha Brasil – Leitura, Teste 2, 2016. Brasília: MEC; Inep, 2016.

CATALANI, Érica Maria Toledo. **Teste Adaptativo Informatizado da Provinha Brasil**: a construção de um instrumento de apoio para professores(as) e gestores(as) de escolas. 282 p. Tese (Doutorado em Educação) – Programa de Pós-Graduação em Educação, Faculdade de Educação, Universidade de São Paulo. São Paulo, 2019.

CHALMERS, R. P. mirt: A Multidimensional Item Response Theory Package for the R Environment. **Journal of Statistical Software**, v. 48, p. 1-29, May 2012.

CHANG, Hua-Hua. Psychometrics behind computerized adaptive testing. **Psychometrika**, p. 1-20, 6 Feb. 2014.

COSTA, Denise Reis. **Métodos estatísticos em testes adaptativos informatizados**. 120 p. Dissertação (Mestrado) – Instituto de Matemática, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2009.

GUBA, Egon G.; LINCOLN, Yvonna S. **Avaliação de quarta geração**. Tradução de Beth Honorato. Campinas, SP: Editora da Unicamp, 2011.

GUTTMAN, Louis. Measurement as structural theory. **Psychometrika**, v. 36, n. 4, p. 329-347, Dec. 1971.

HAIR JR, Joseph F. et al. **Análise multivariada de dados**. 6. ed. Tradução de Adonai Schlup Sant’Anna. Porto Alegre: Bookman, 2009. 688 p. [Original 2006]

HAIR JR, Joseph F. et al. **Multivariate data analysis**. 8<sup>th</sup> ed. Hampshire: Cengage Learning, 2019.

KLEIN, Ruben. Utilização da Teoria da Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (Saeb). **Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 11, n. 40, p. 283-296, jul./set. 2003.

LORD, Frederic M.; NOVICK, Melvin R. **Statistical theories of mental test scores**. Reading, MA: Addison-Wesley, 1968. (The Addison-Wesley Series in Behavioral Science. Quantitative Methods).

LORD, Frederic M. A broad-range tailored test of verbal ability. **Applied Psychological Measurement**, v. 1, n. 1, p. 95-100, Winter 1977.

LORD, Frederic M. **Applications of item response theory to practical testing problems**. Hillsdale: Lawrence Erlbaum, 1980.

MAGIS, David; YAN, Duanli; VON DAVIER, Alina A. **Computerized Adaptive and Multistage Testing with R**: using Packages catR and mstR. Cham: Springer, 2017.

MENARBINI, Andreia; GOMES, M. Tavares. O direito das crianças à alfabetização: um desafio na educação brasileira. **Revista de Estudos Aplicados em Educação**, v. 4, n. 7, p. 3-18, jan./jun. 2019.

MORTATTI, Maria do Rosário. Longo. Um balanço crítico da "Década da Alfabetização" no Brasil. **Cadernos Cedes**, v. 33, n. 89, p. 15-34, jan. 2013.

MUÑIZ, José. **Introducción a la Teoría de Respuesta a los Ítems**. Madrid: Pirámide, 1997. (Psicología).

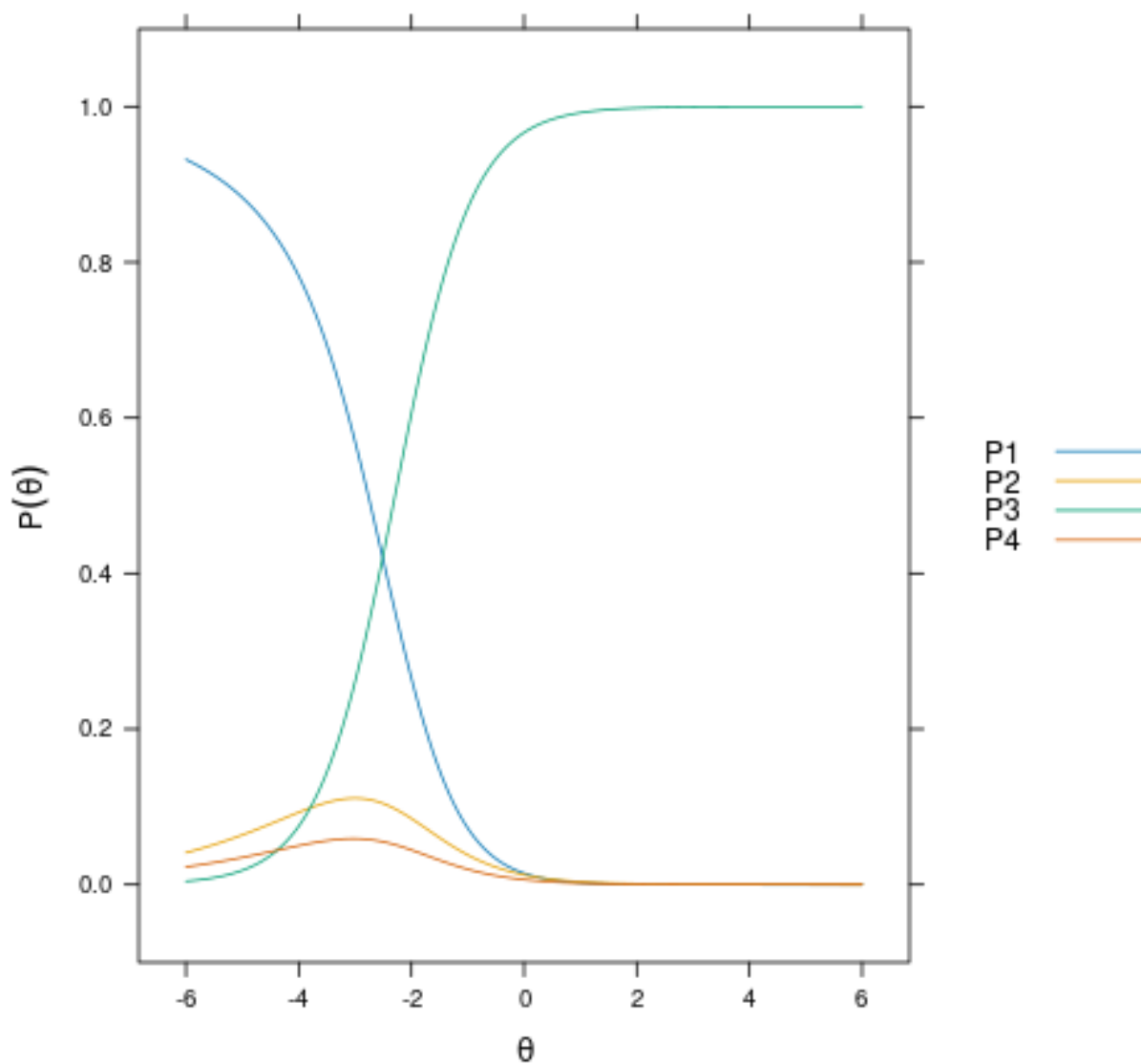
NEVO, David. **Evaluación basada en el centro: un diálogo para la mejora educativa**. Traducción de Maria Serrano Bericat. Bilbao: Mensajero, 1997.

OLEA, Julio; PONSODA, Vicente; PRIETO, Gerardo (Ed.). **Tests informatizados: fundamentos y aplicaciones**. Madrid: Pirámide, 1999. (Psicología).

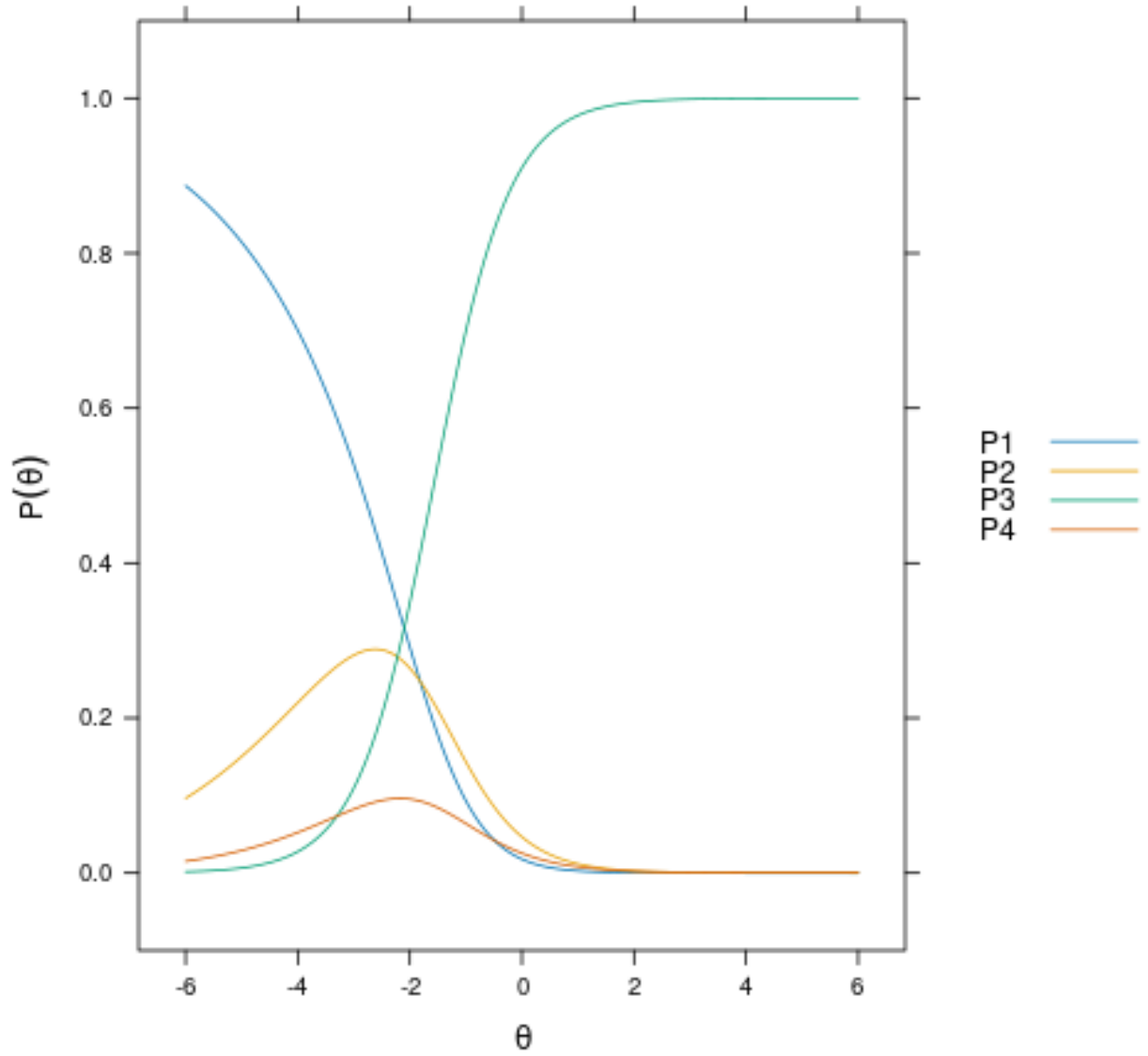
OLEA, Julio; PONSODA, Vicente. **Tests adaptativos informatizados**. Madrid: edición de los autores, 2002. Disponível em:  
[https://www.researchgate.net/profile/Julio\\_Diaz5/publication/265040034\\_TEST\\_ADAPTATIVOS\\_INFORMATIZADOS/links/54e481630cf2dbf60696bbb5/TEST-ADAPTATIVOS-INFORMATIZADOS.pdf](https://www.researchgate.net/profile/Julio_Diaz5/publication/265040034_TEST_ADAPTATIVOS_INFORMATIZADOS/links/54e481630cf2dbf60696bbb5/TEST-ADAPTATIVOS-INFORMATIZADOS.pdf). Acesso em: 24 set. 2021.

PASQUALI, Luiz. **Psicometria: teoria dos testes na psicologia e na educação**. 5. ed. Petrópolis, RJ: Vozes, 2013.

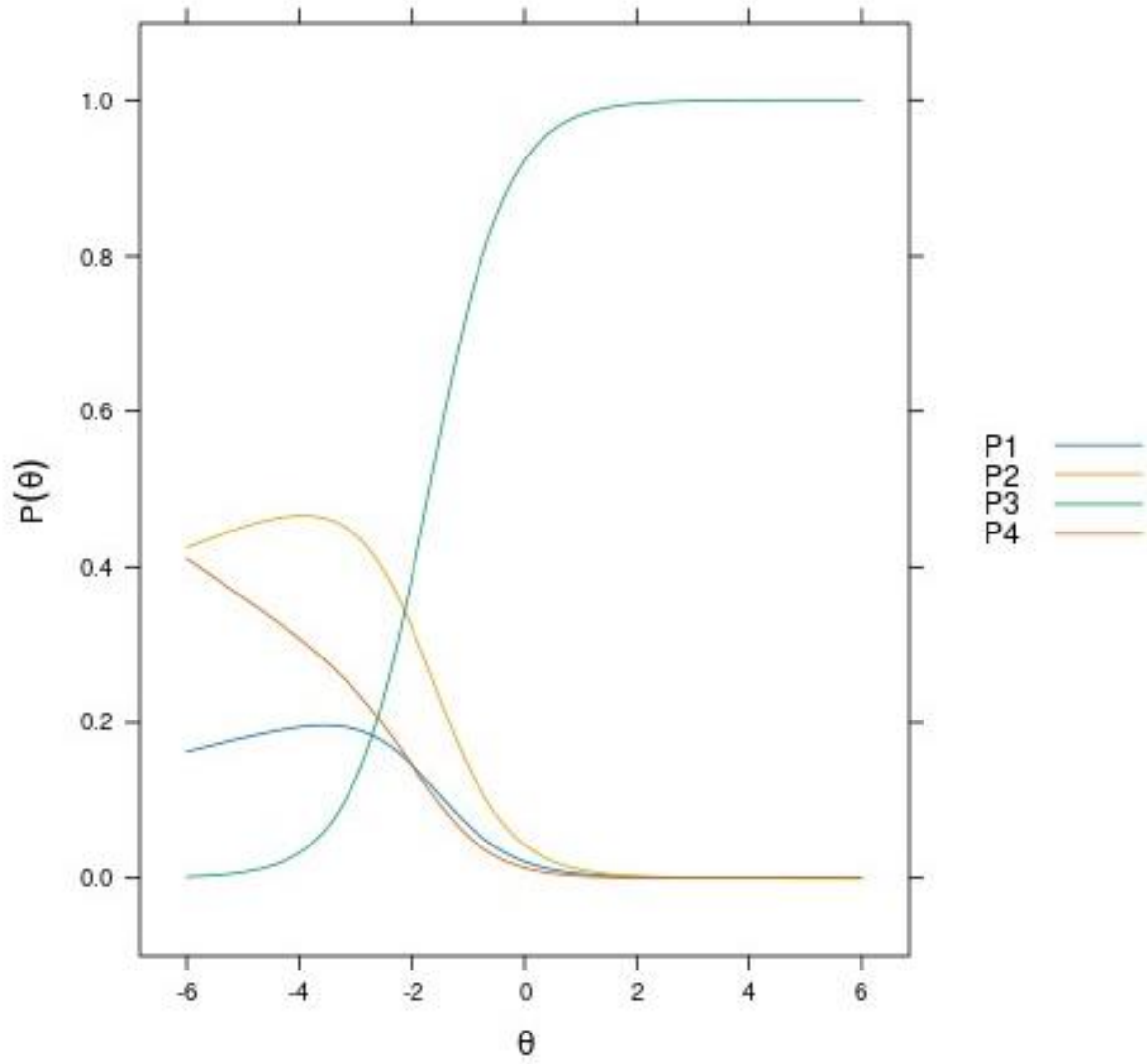
PRIMI, Ricardo. **Psicometria: fundamentos matemáticos da Teoria Clássica dos Testes**. **Avaliação Psicológica**, São Paulo, v. 11, n. 2, p. 297-307, 2012.

**Anexo 1 – Gráficos de Função de Probabilidade dos itens****Gráfico 39 – Função de Probabilidade – Item 1**

**Fonte:** Elaborado própria

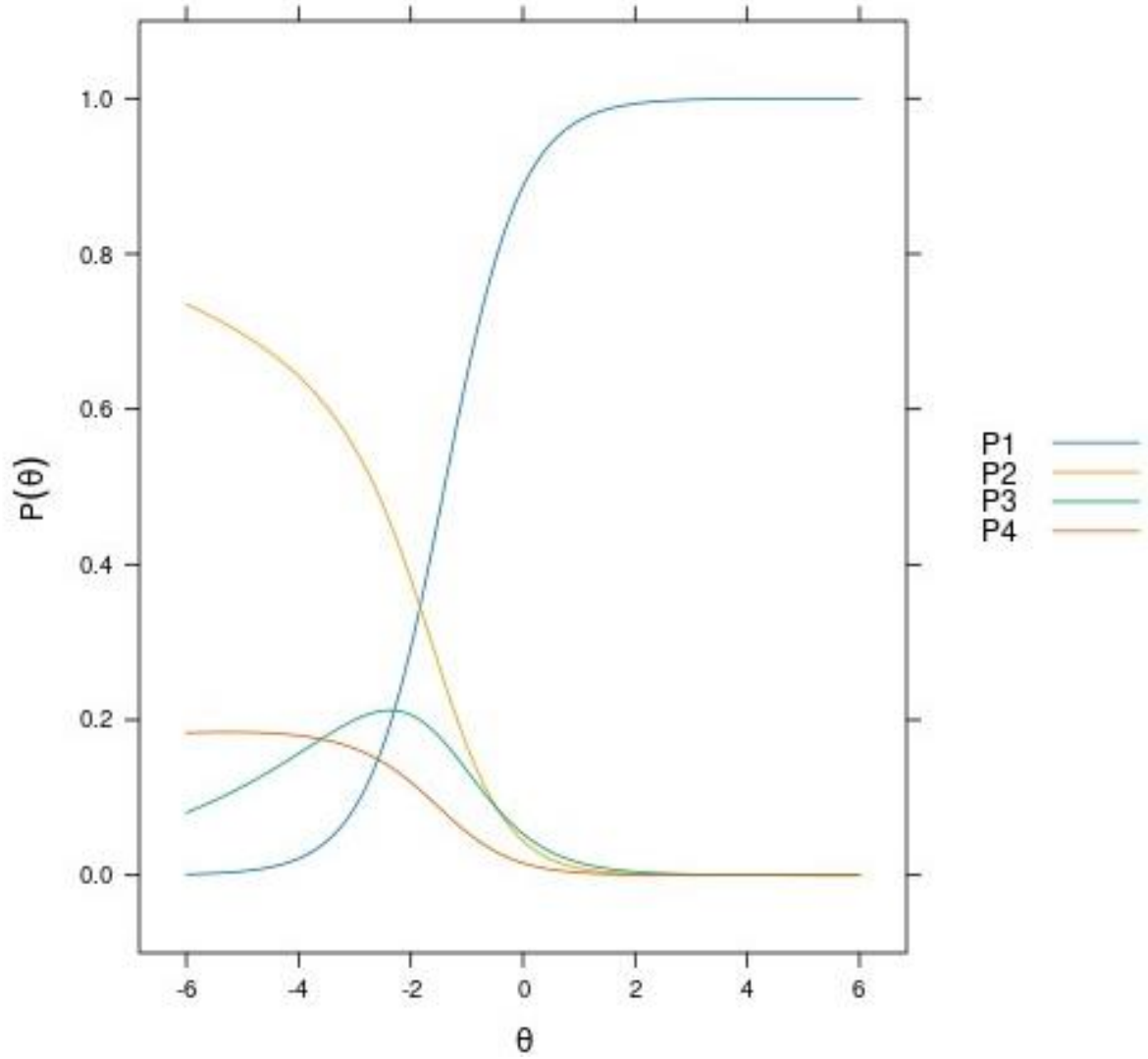
**Gráfico 40** – Função de Probabilidade – Item 2

**Fonte:** Elaborado própria

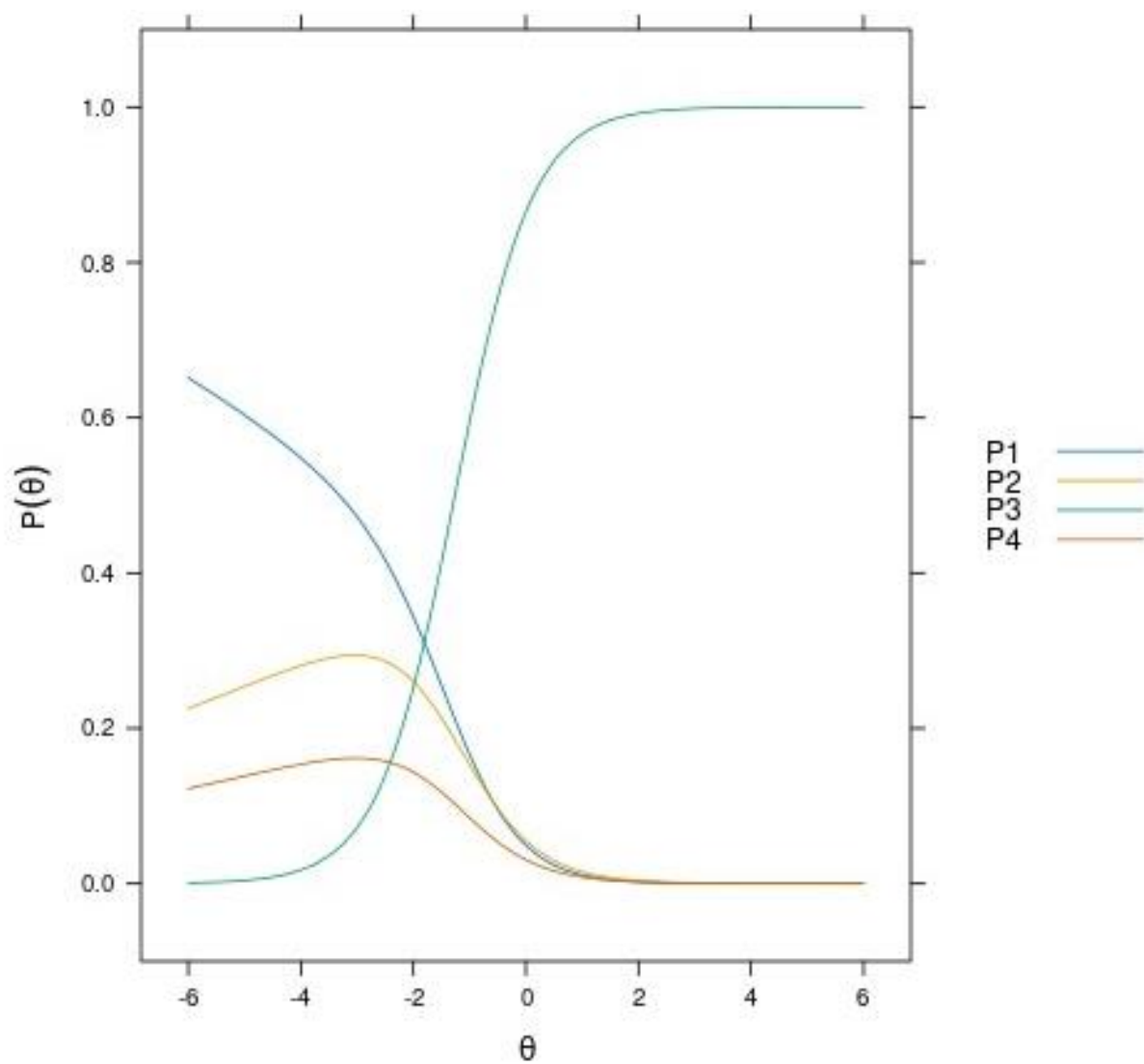
**Gráfico 41** – Função de Probabilidade – Item 03

Fonte: Elaborado própria

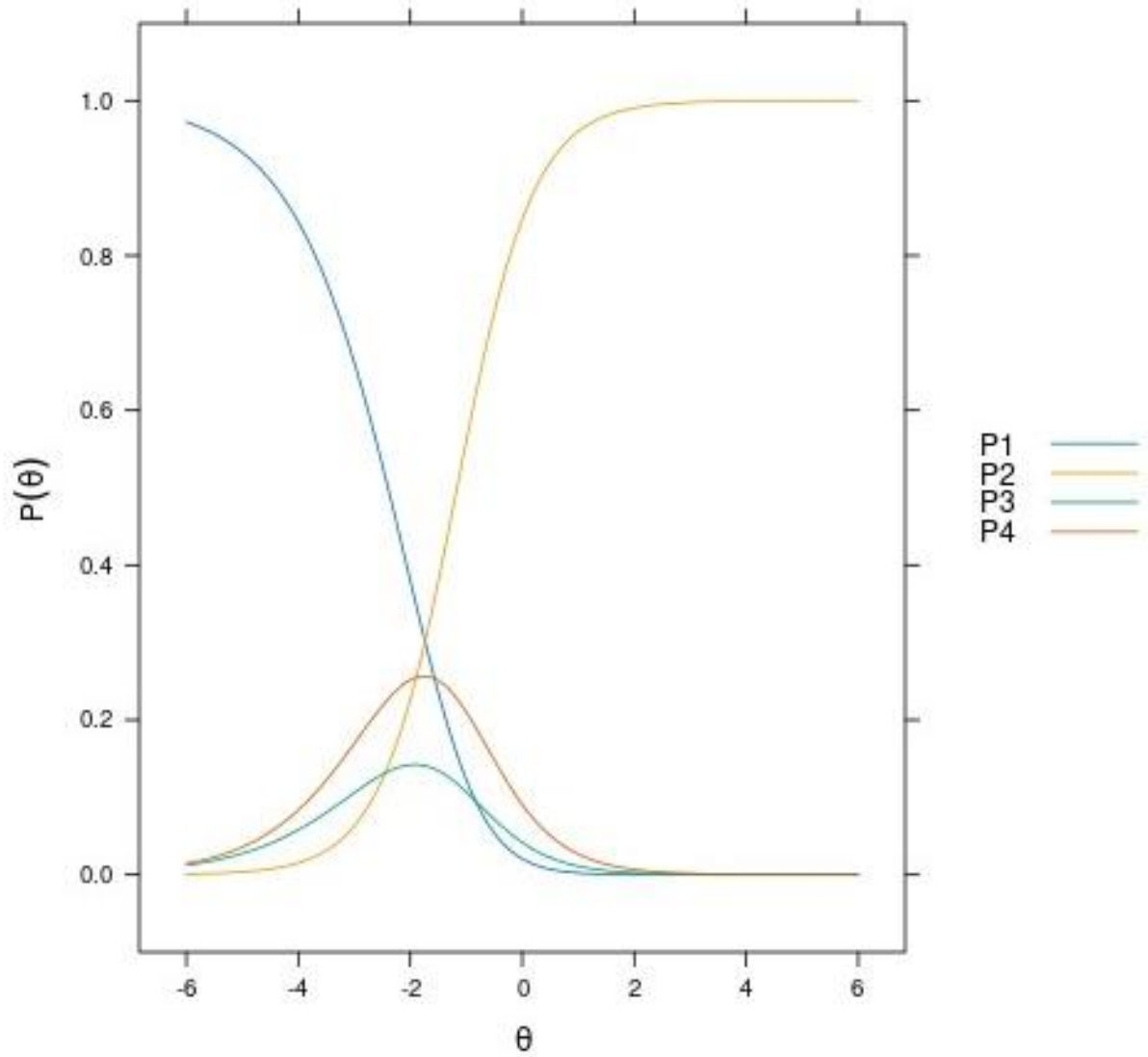


**Gráfico 42 – Função de Probabilidade – Item 04**

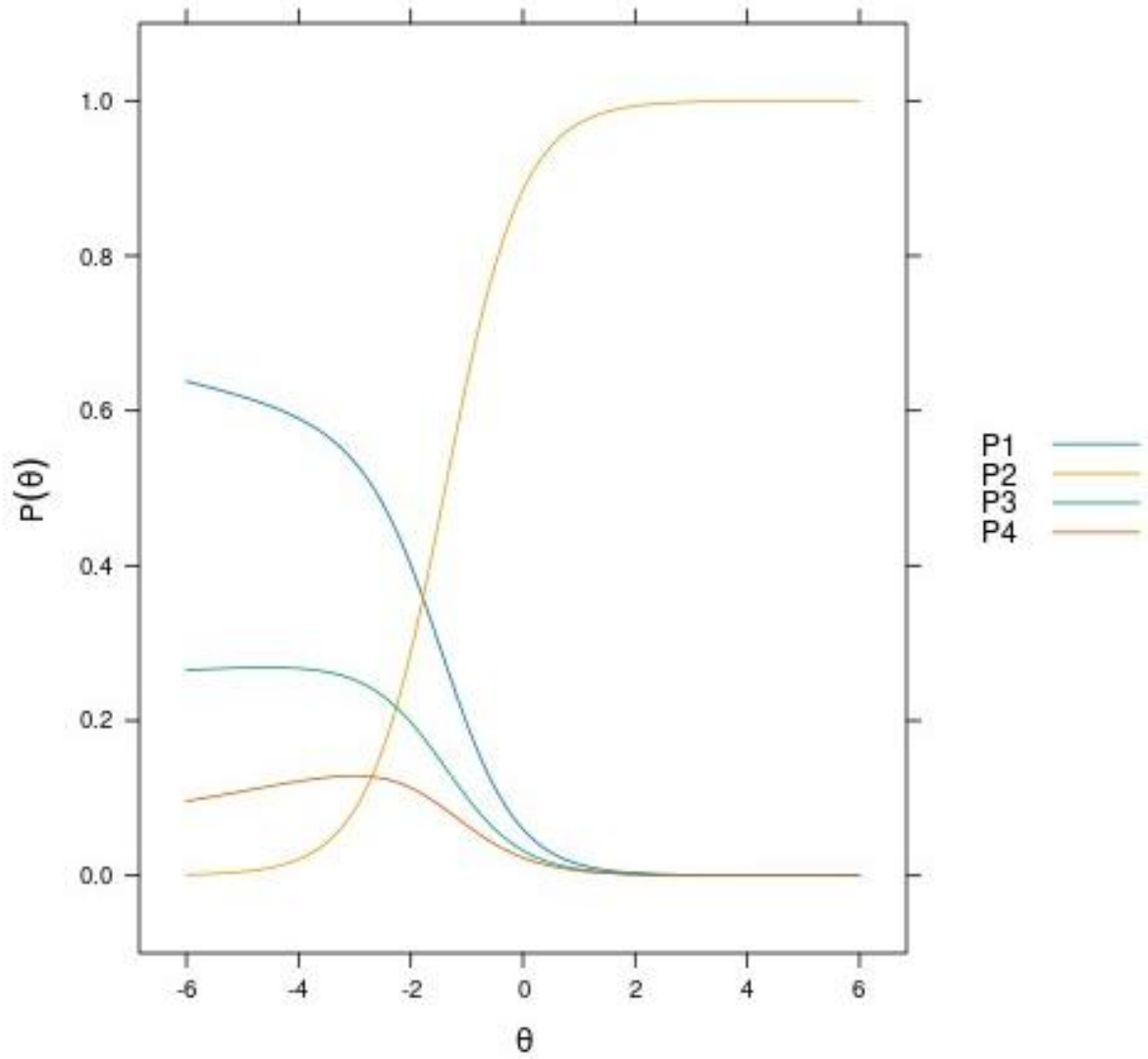
**Fonte:** Elaborado própria

**Gráfico 43** – Função de Probabilidade – Item 05

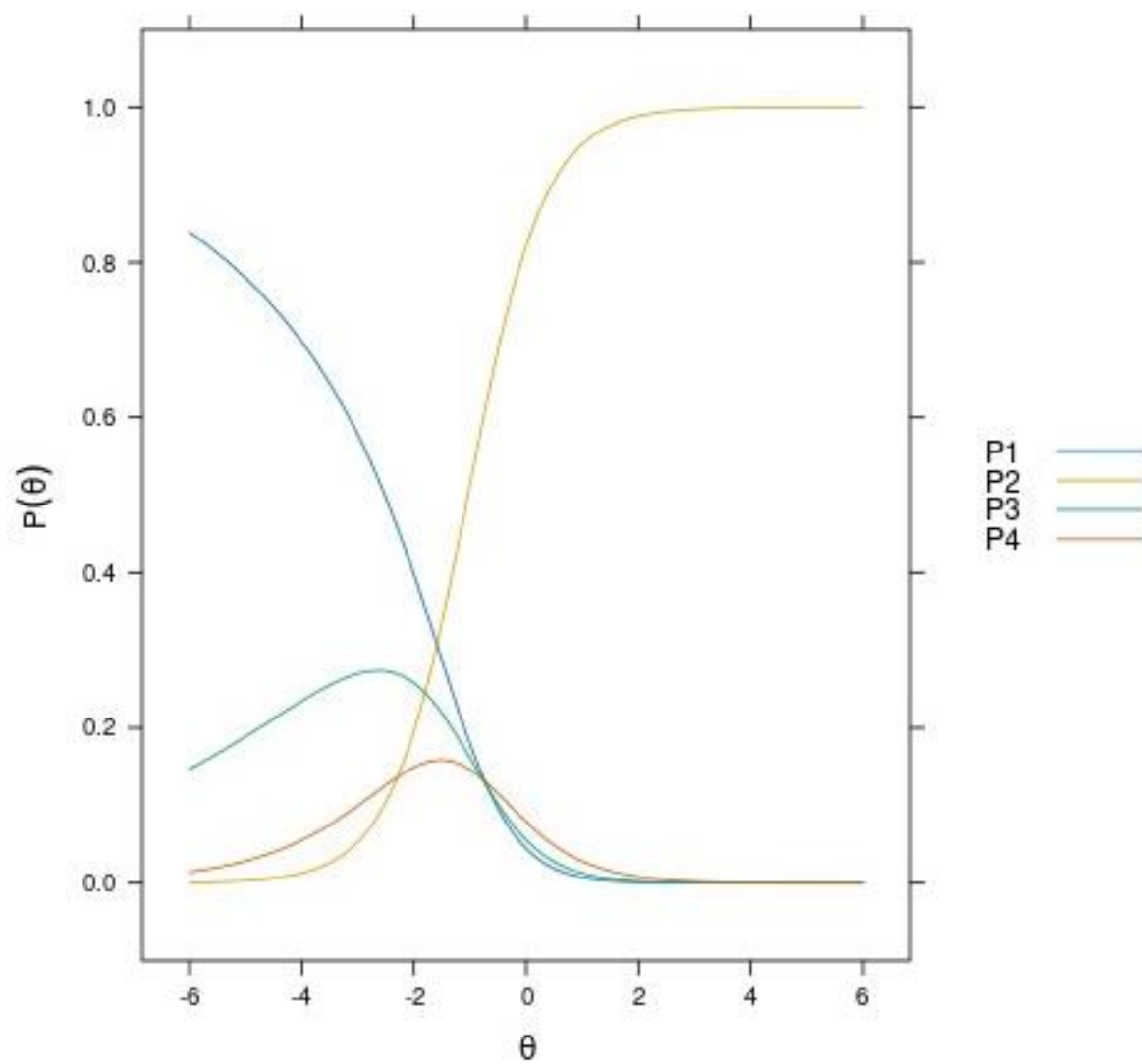
Fonte: Elaborado própria

**Gráfico 44** – Função de Probabilidade – Item 06

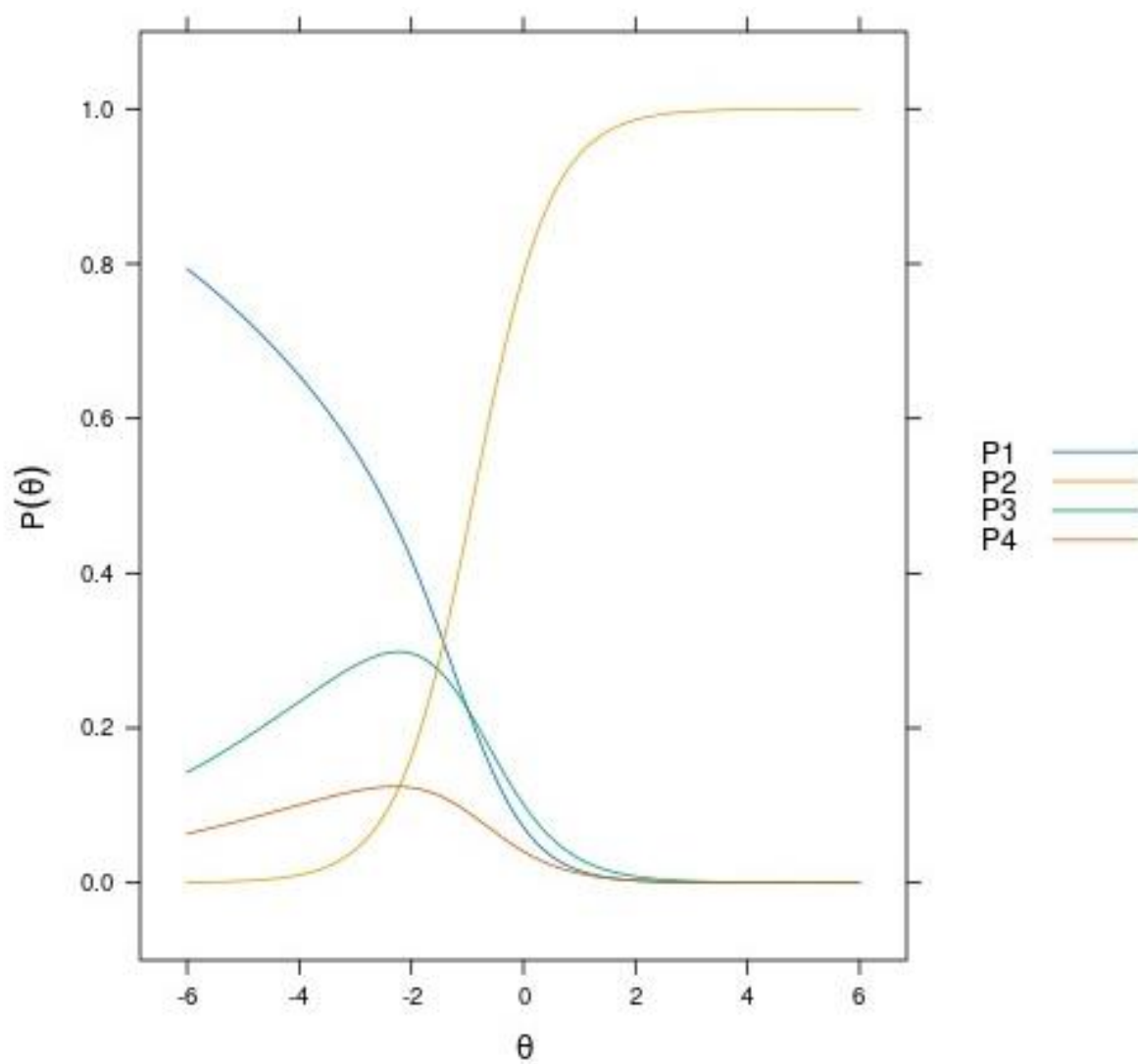
Fonte: Elaborado própria

**Gráfico 45** – Função de Probabilidade – Item 07

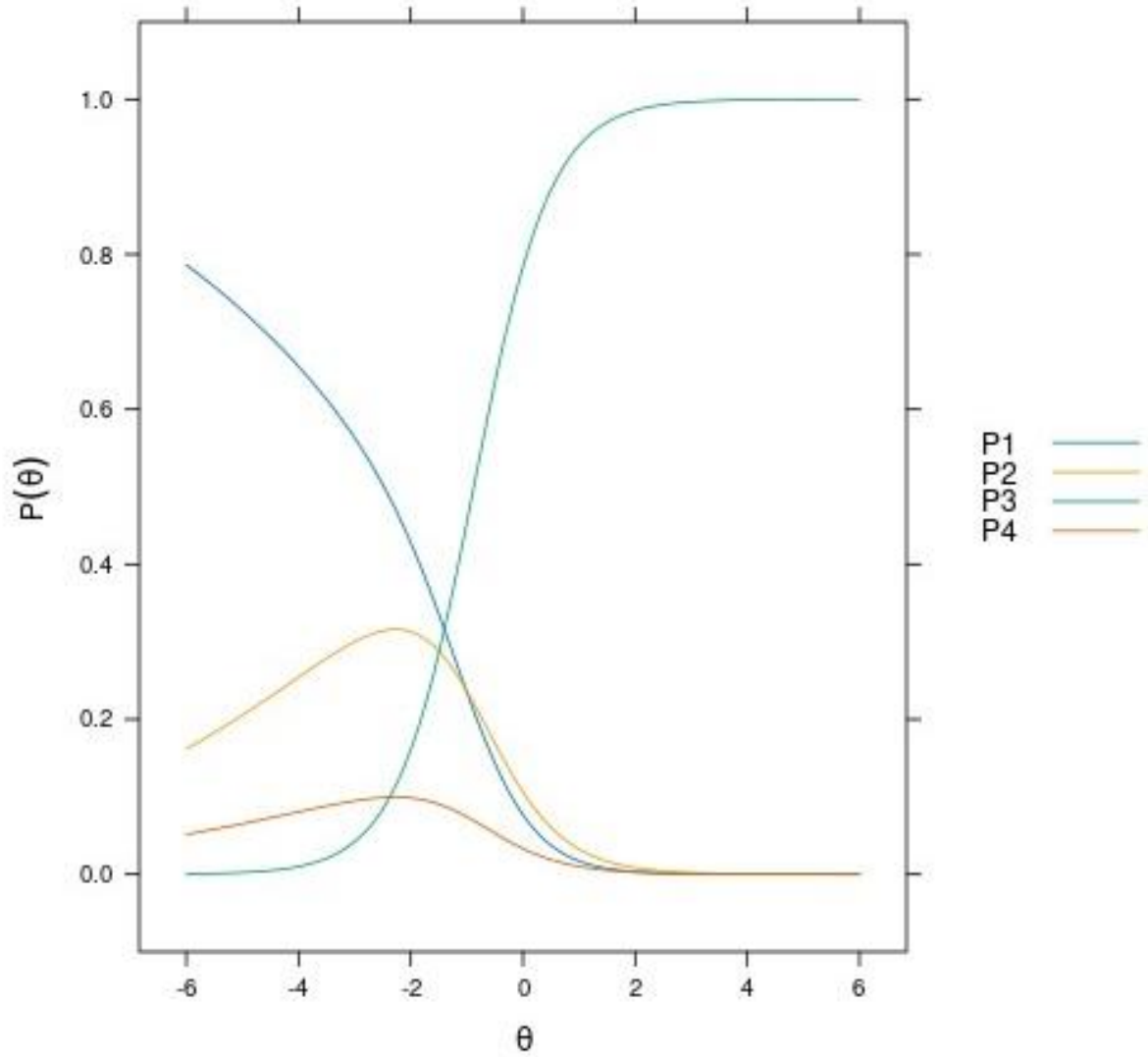
Fonte: Elaborado própria

**Gráfico 46** – Função de Probabilidade – Item 08

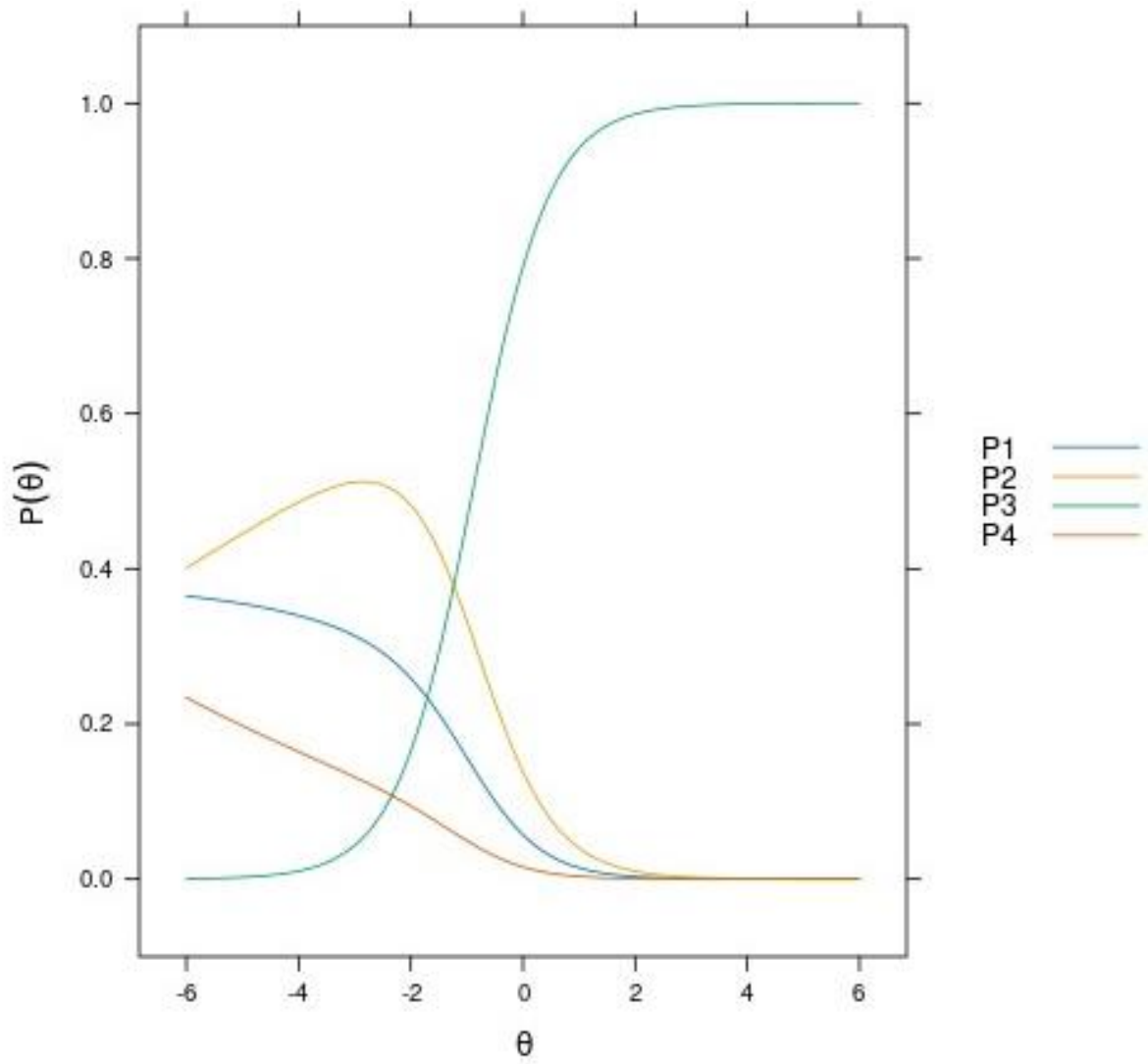
Fonte: Elaborado própria

**Gráfico 47** – Função de Probabilidade – Item 09

Fonte: Elaborado própria

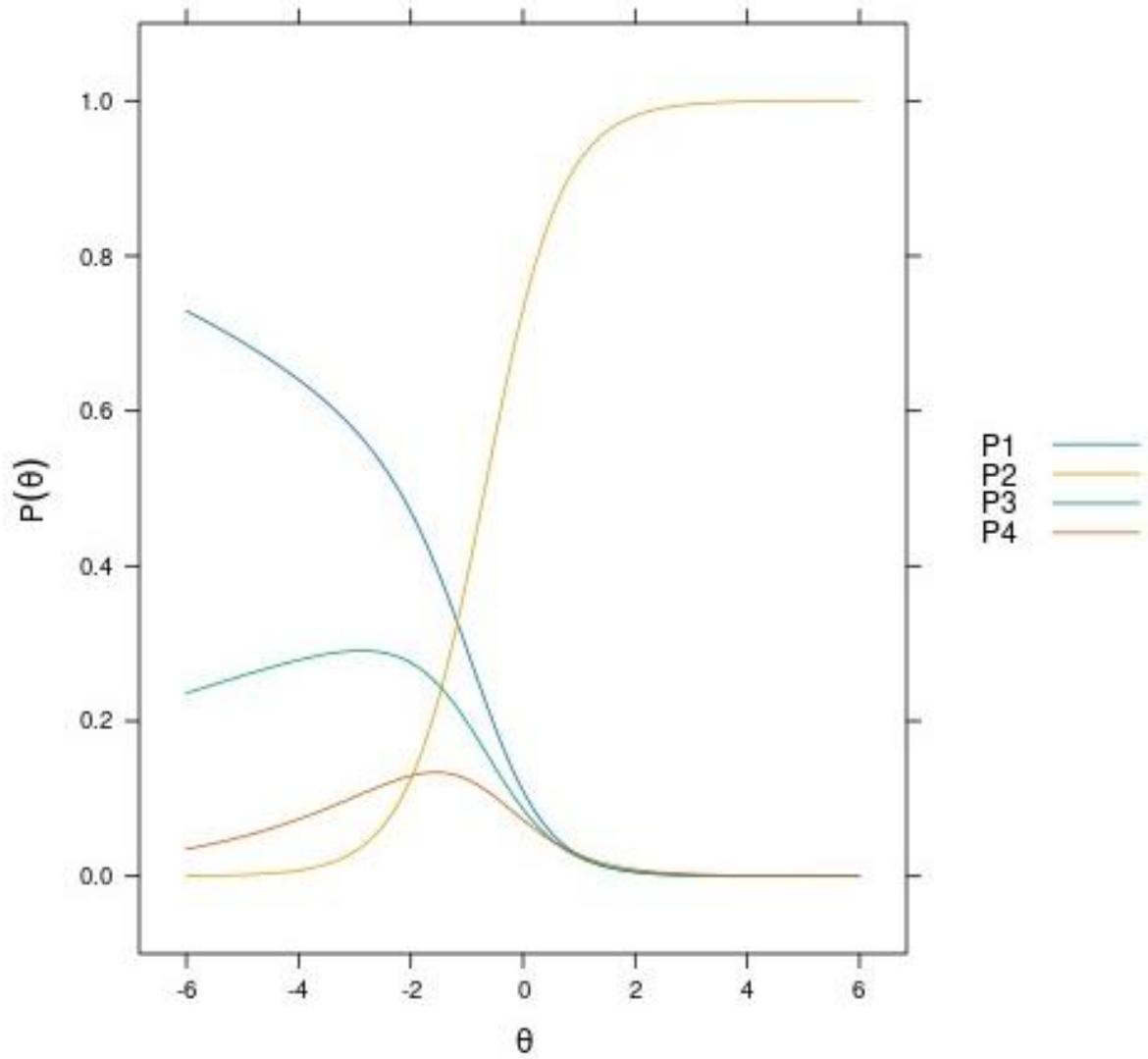
**Gráfico 48** – Função de Probabilidade – Item 10

**Fonte:** Elaborado própria

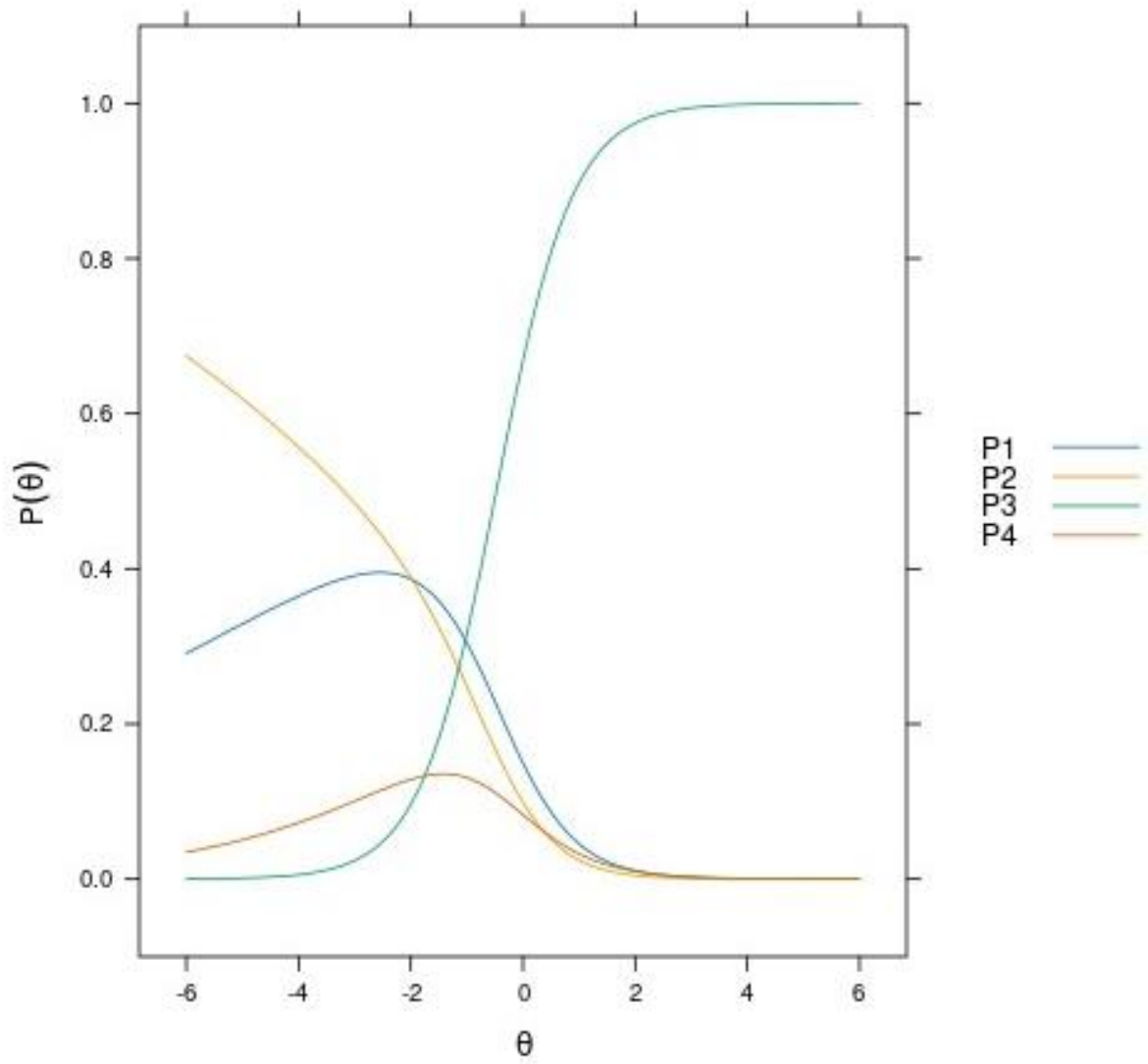
**Gráfico 49** – Função de Probabilidade – Item 11

Fonte: Elaborado própria

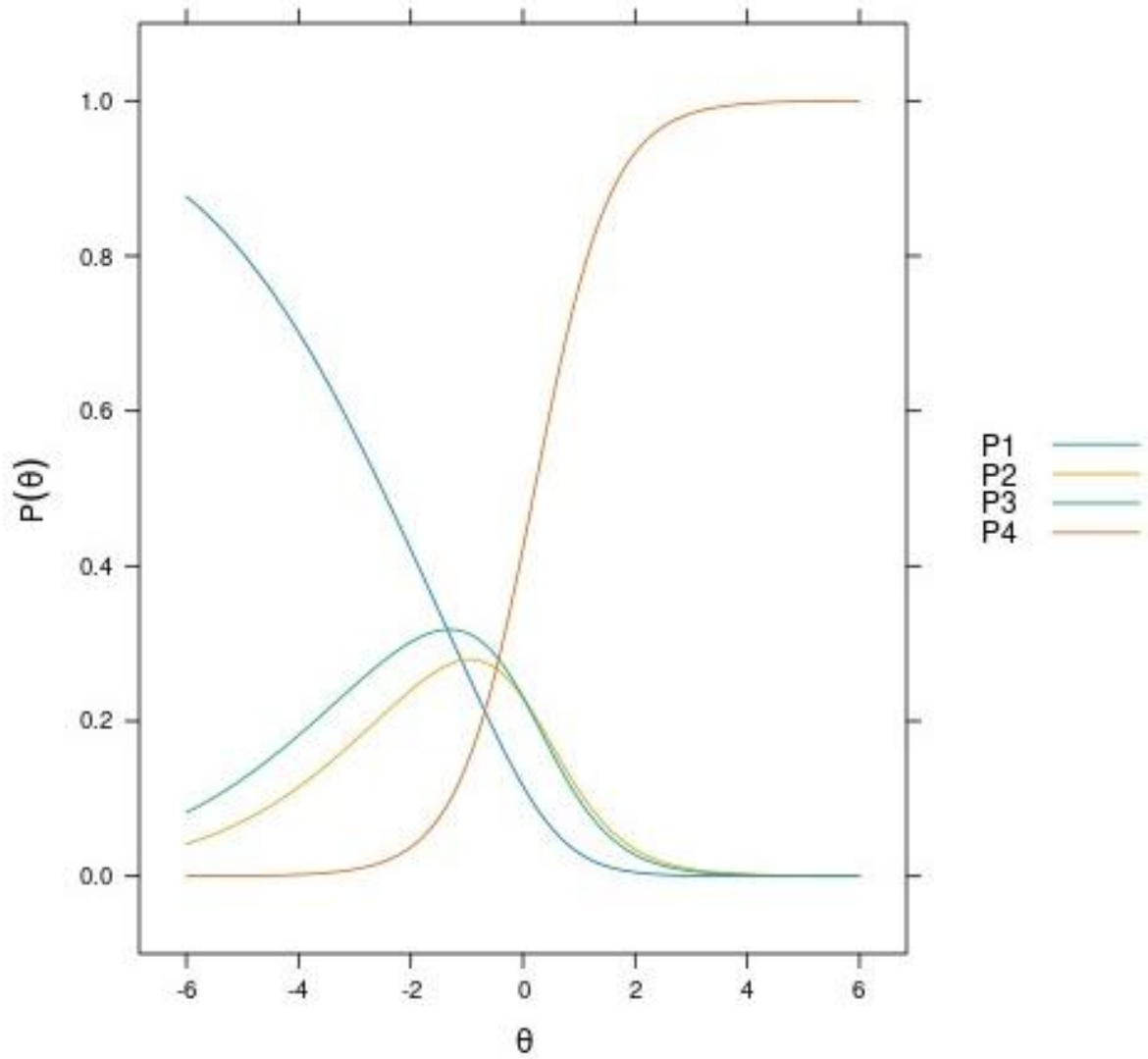


**Gráfico 50** – Função de Probabilidade – Item 12

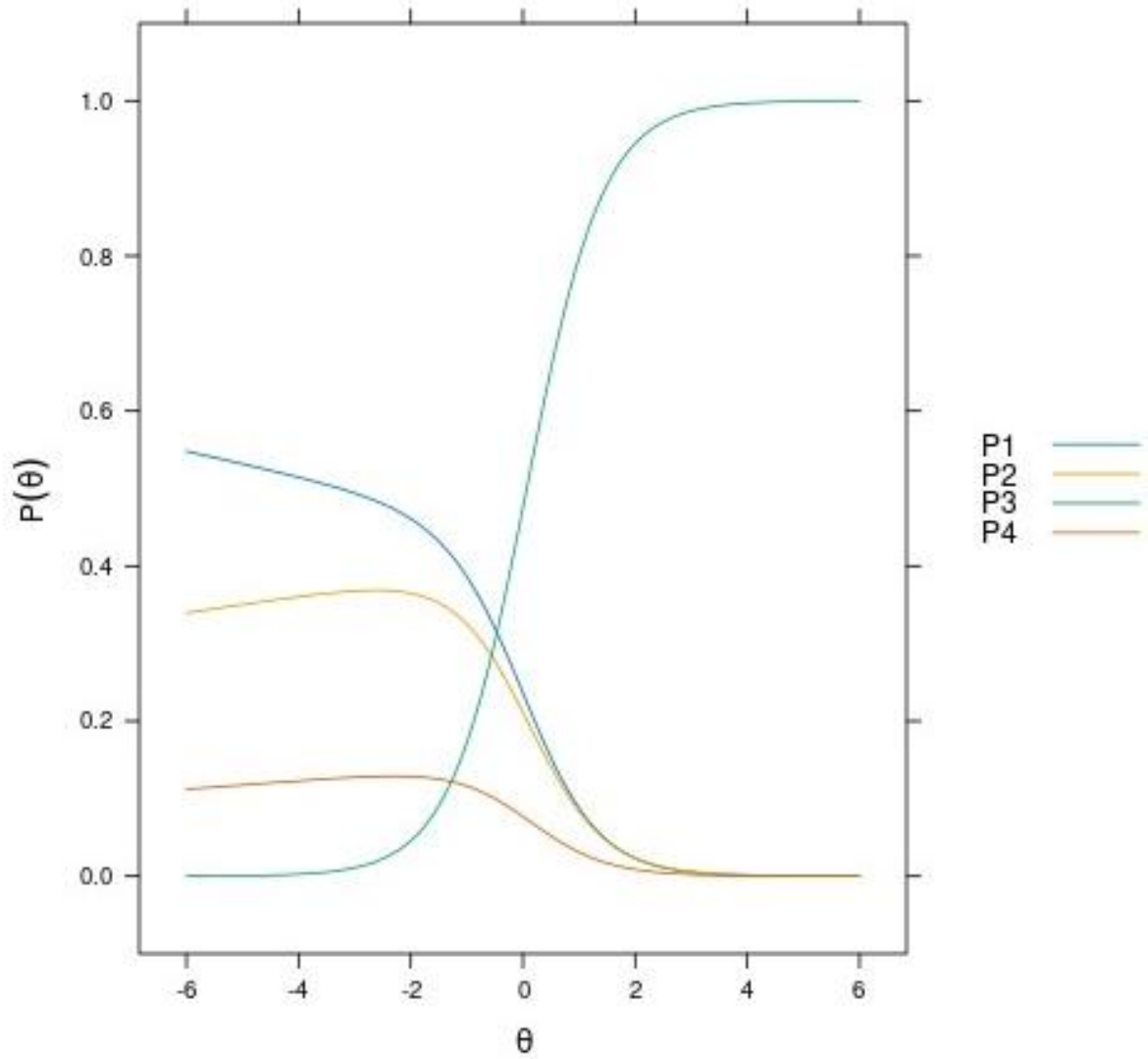
Fonte: Elaborado própria

**Gráfico 51** – Função de Probabilidade – Item 13

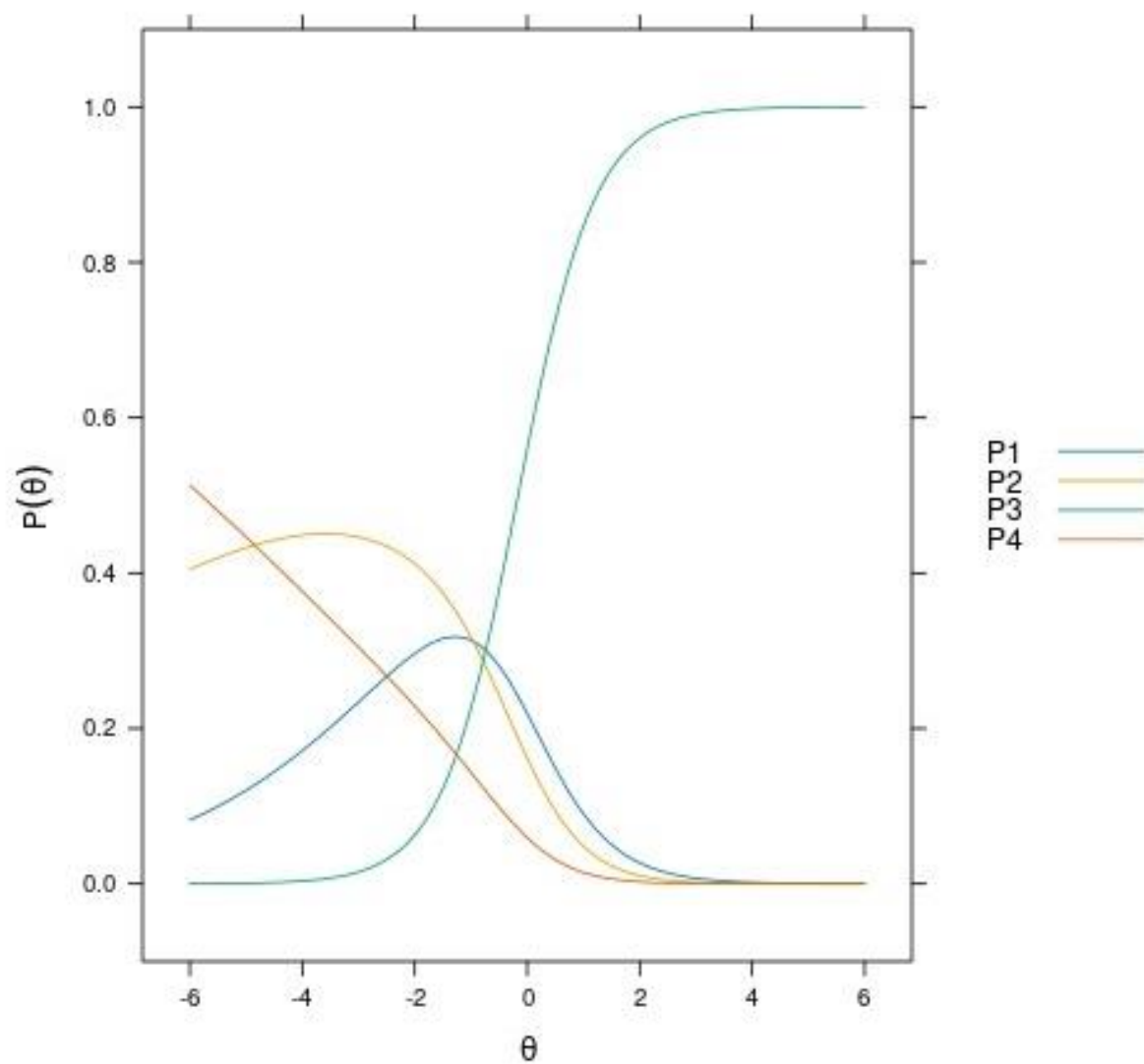
**Fonte:** Elaborado própria

**Gráfico 52 – Função de Probabilidade – Item 14**

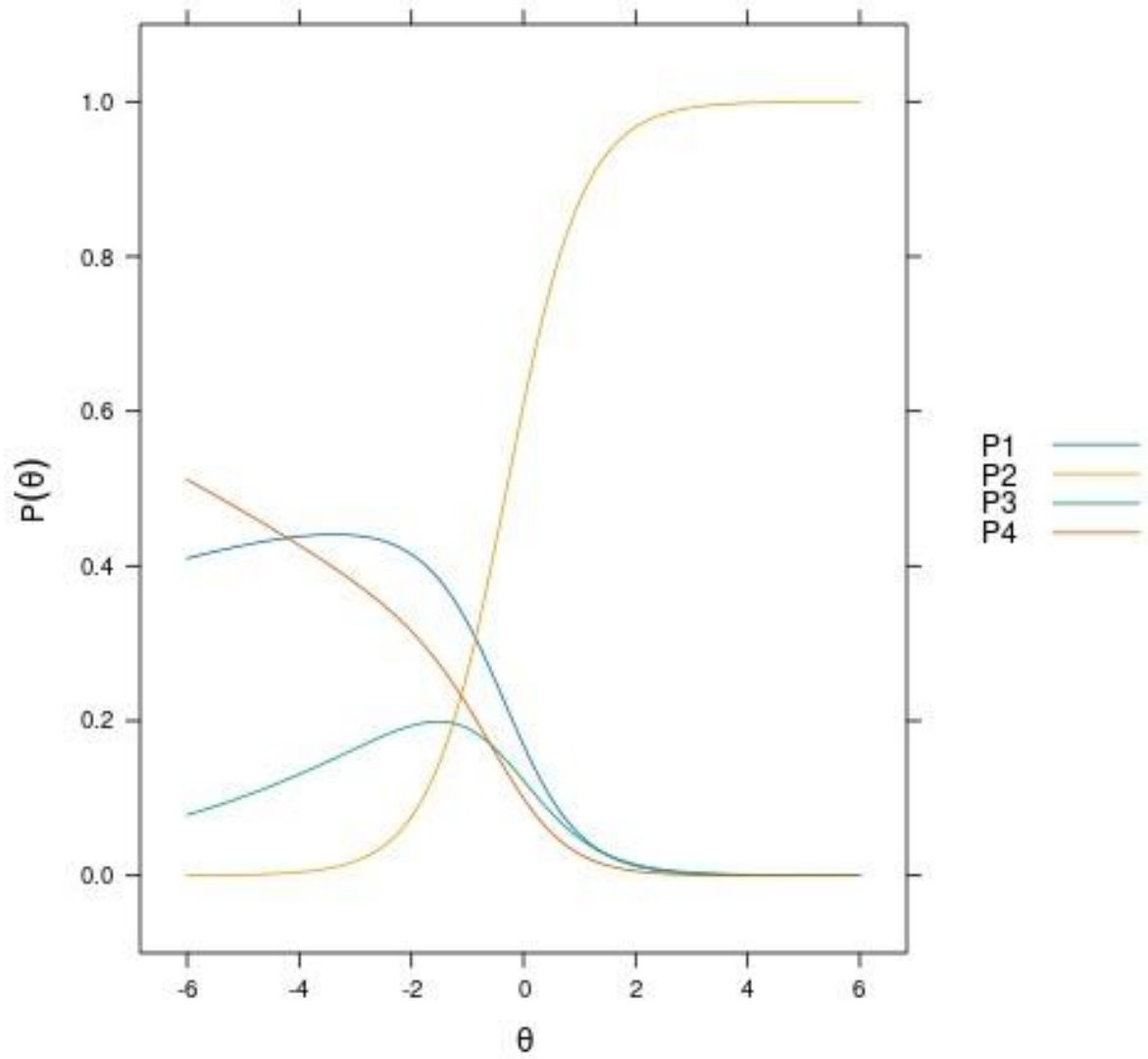
Fonte: Elaborado própria

**Gráfico 53** – Função de Probabilidade – Item 15

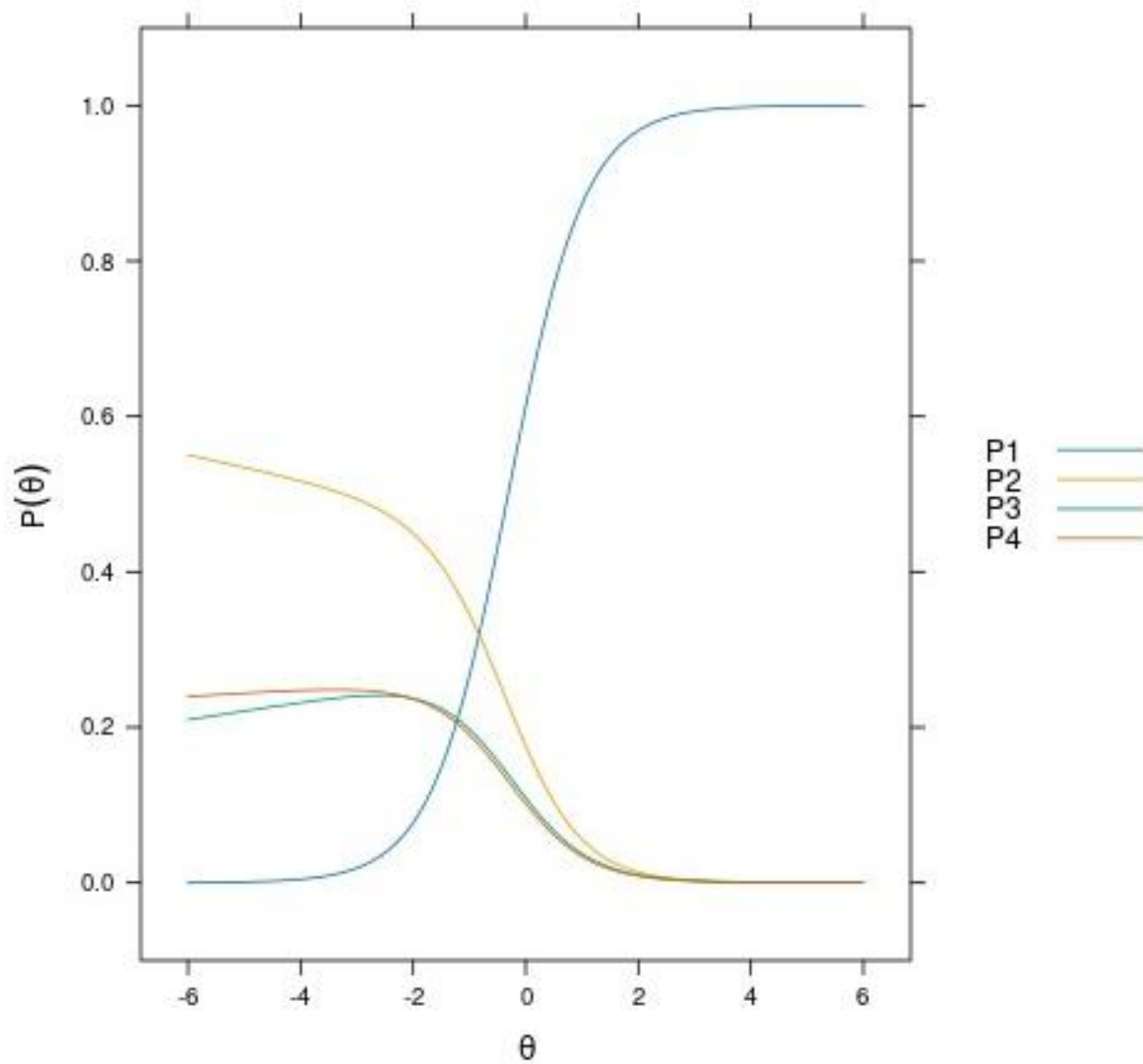
Fonte: Elaborado própria

**Gráfico 54** – Função de Probabilidade – Item 16

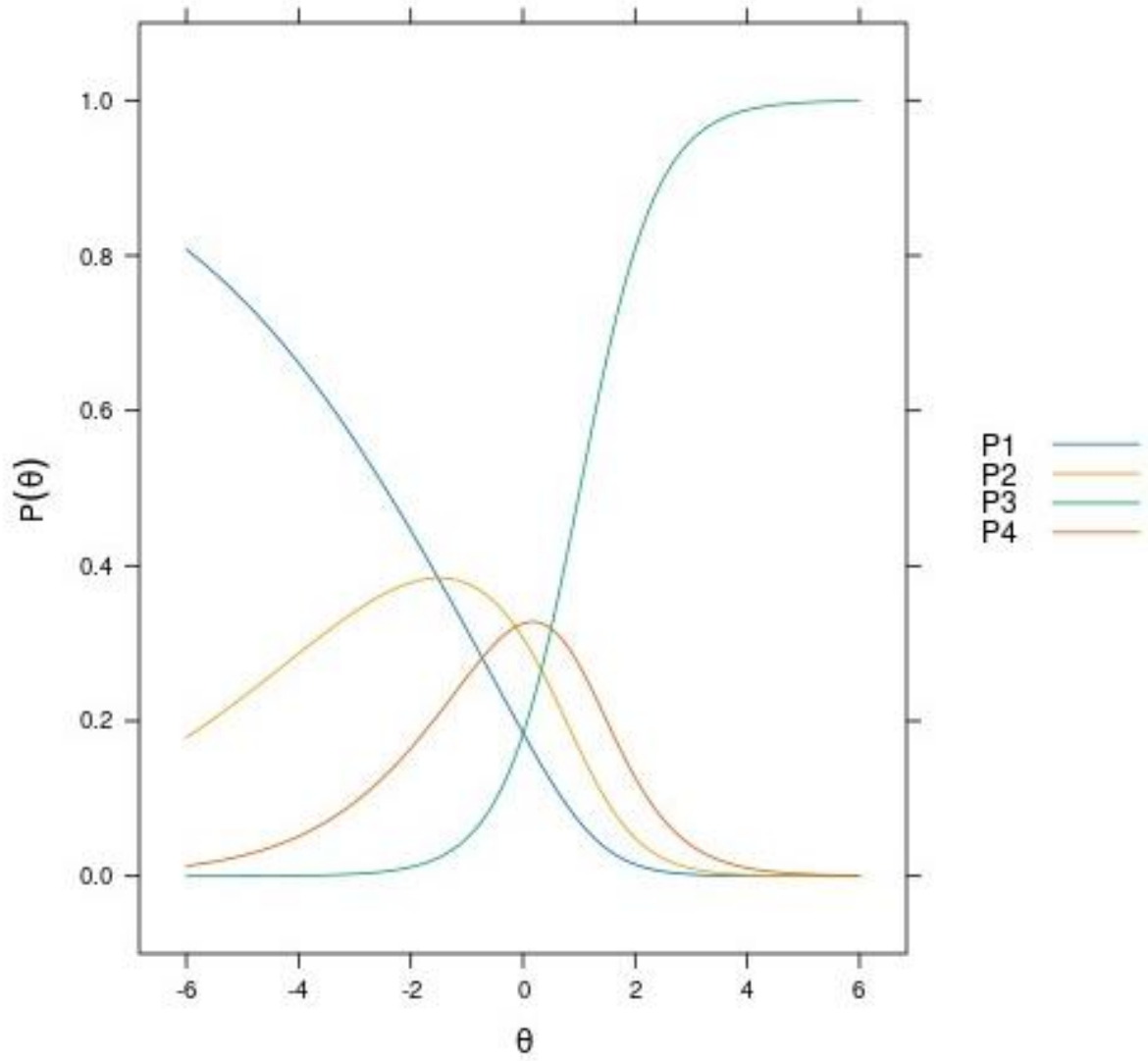
**Fonte:** Elaborado própria

**Gráfico 55** – Função de Probabilidade – Item 17

**Fonte:** Elaborado própria

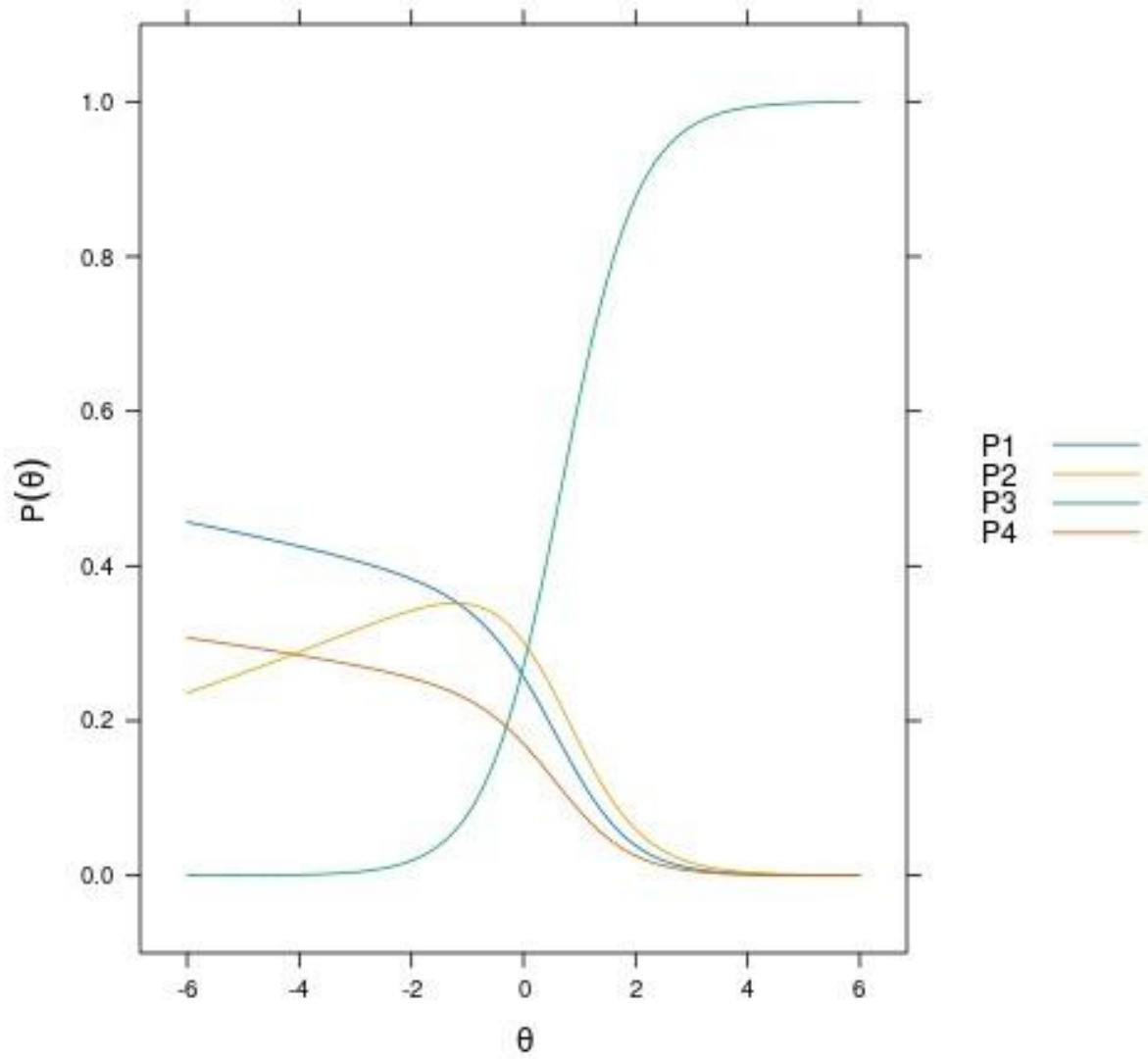
**Gráfico 56** – Função de Probabilidade – Item 18

**Fonte:** Elaborado própria

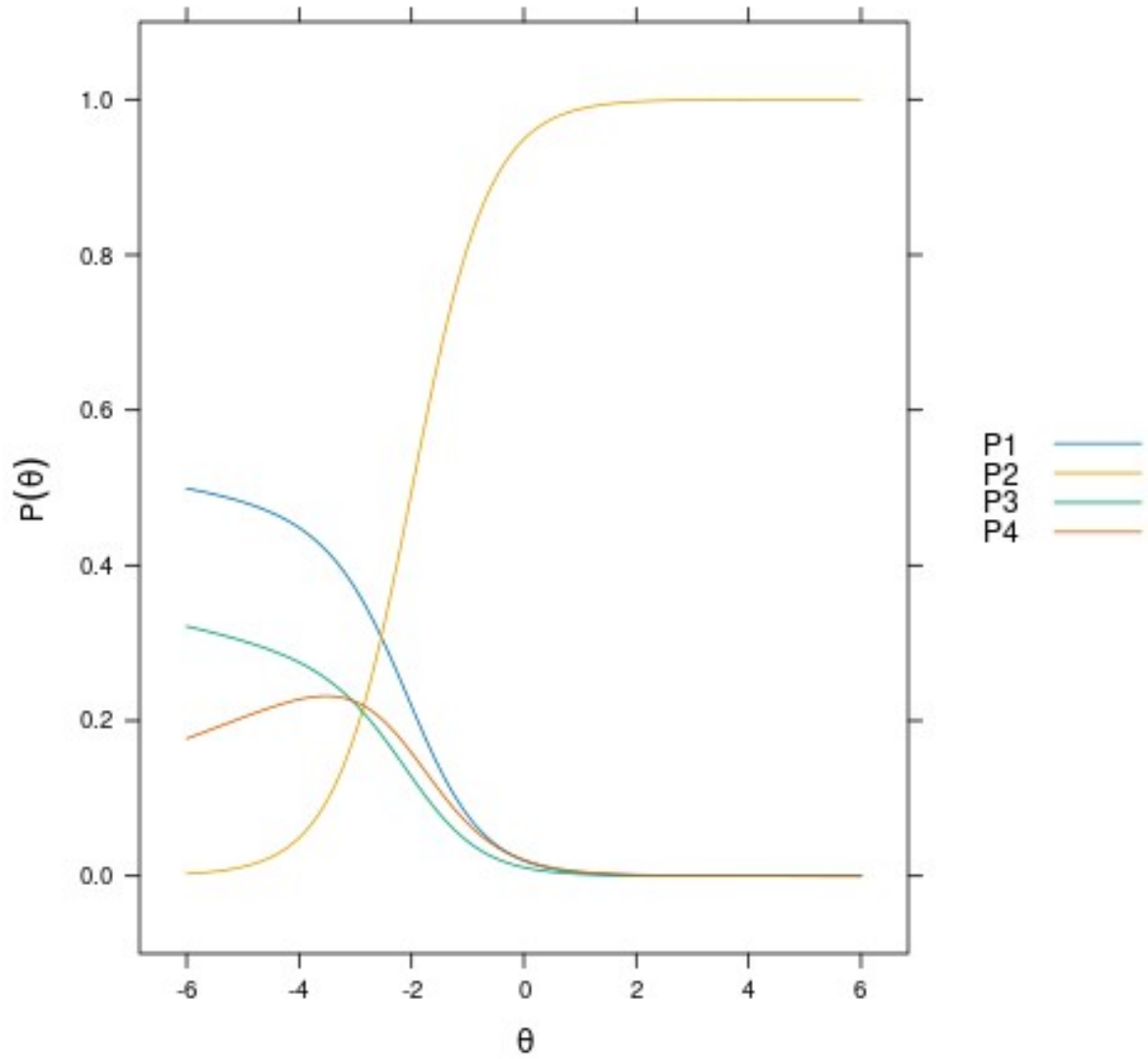
**Gráfico 57** – Função de Probabilidade – Item 19

**Fonte:** Elaborado própria

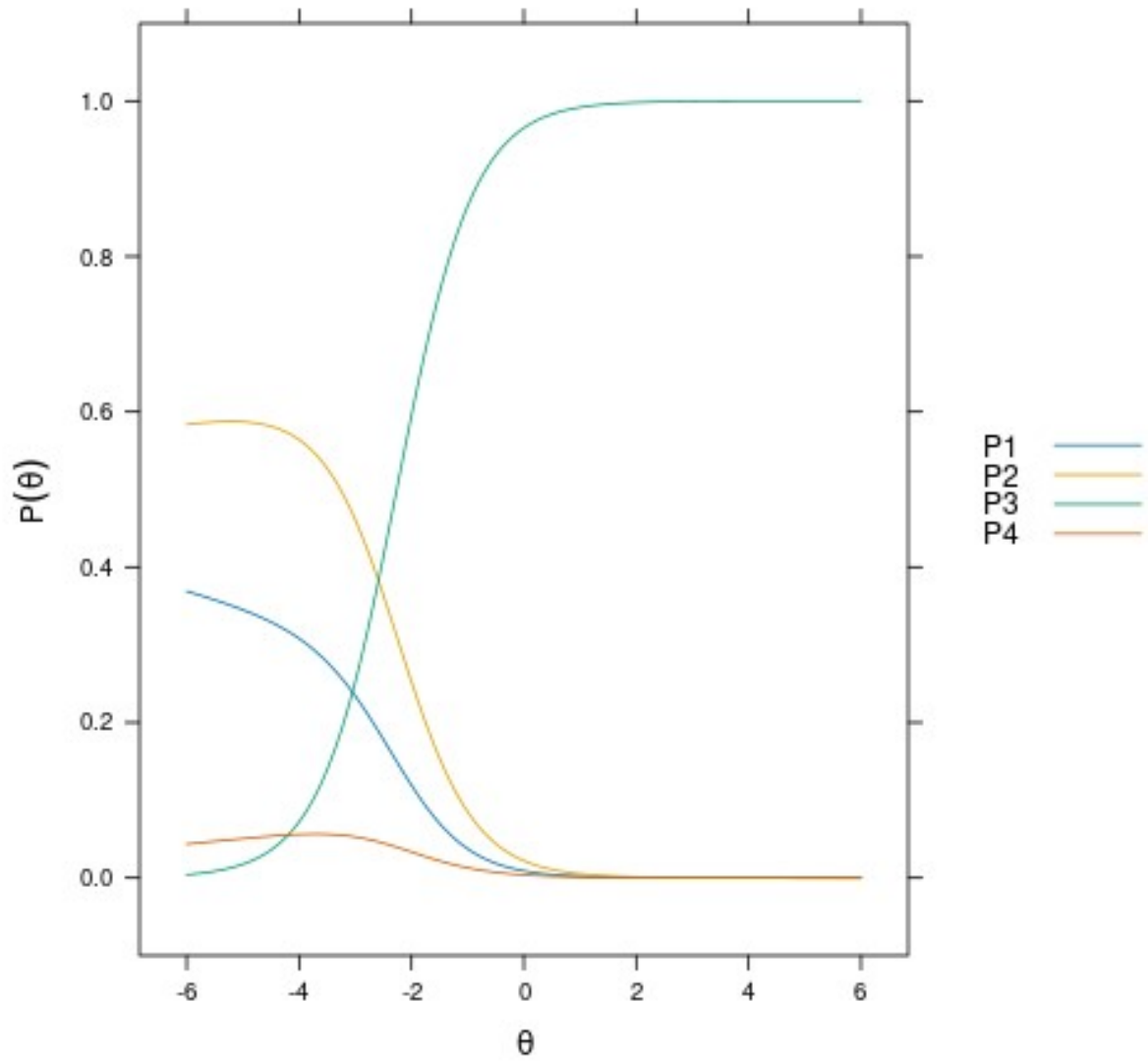


**Gráfico 58** – Função de Probabilidade – Item 20

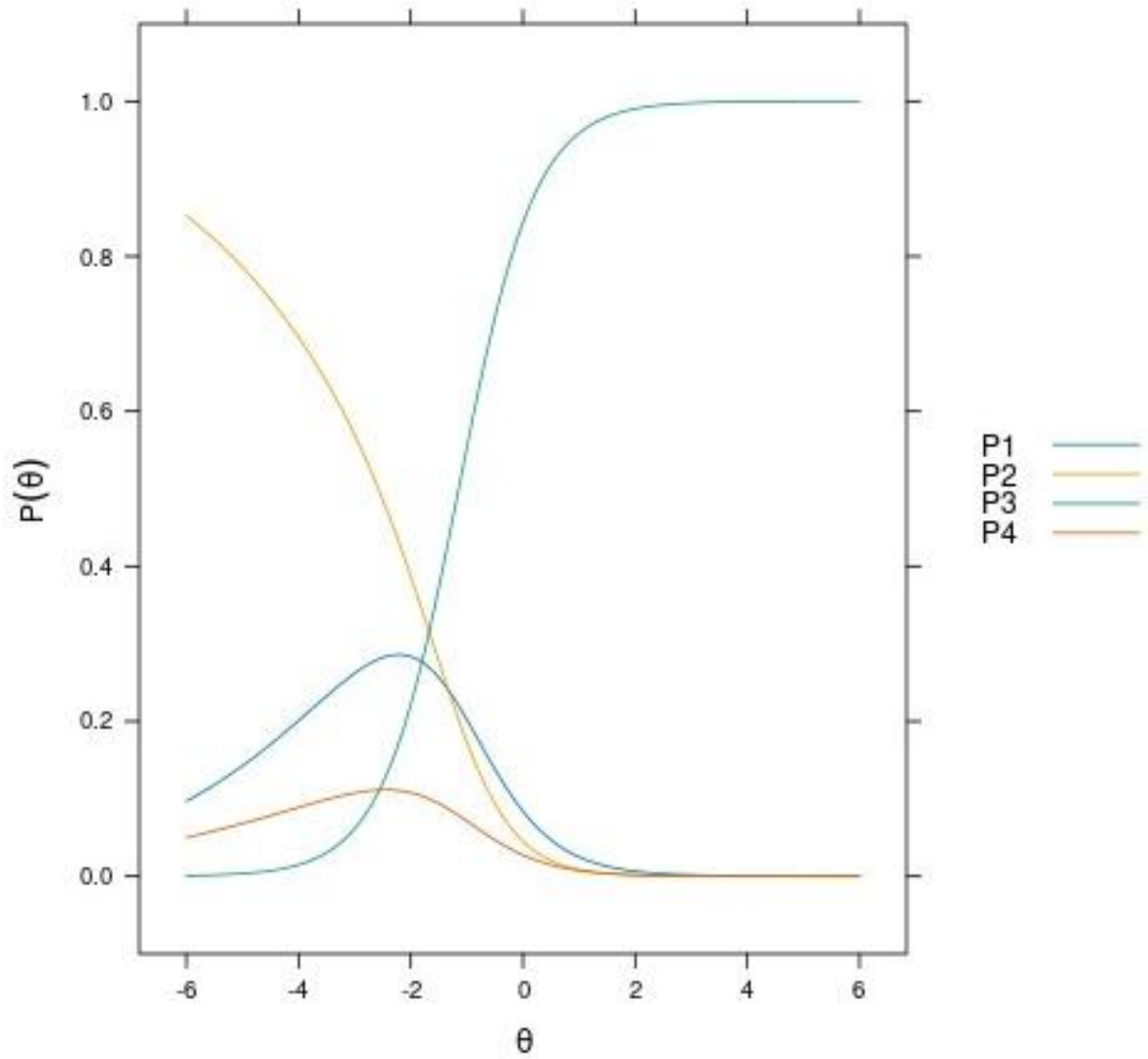
**Fonte:** Elaborado própria

**Gráfico 59** – Função de Probabilidade – Item 21

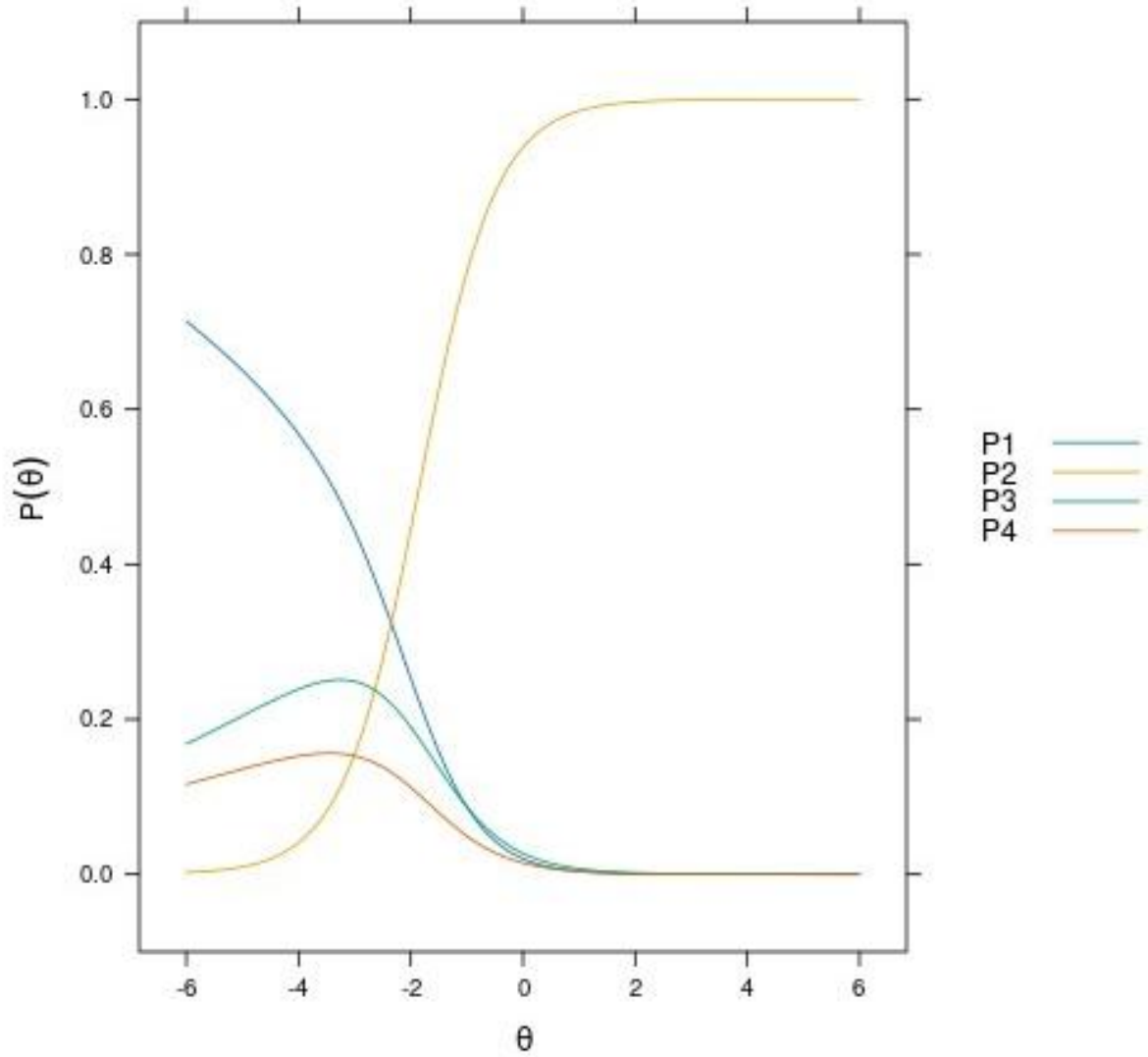
**Fonte:** Elaborado própria

**Gráfico 60** – Função de Probabilidade – Item 22

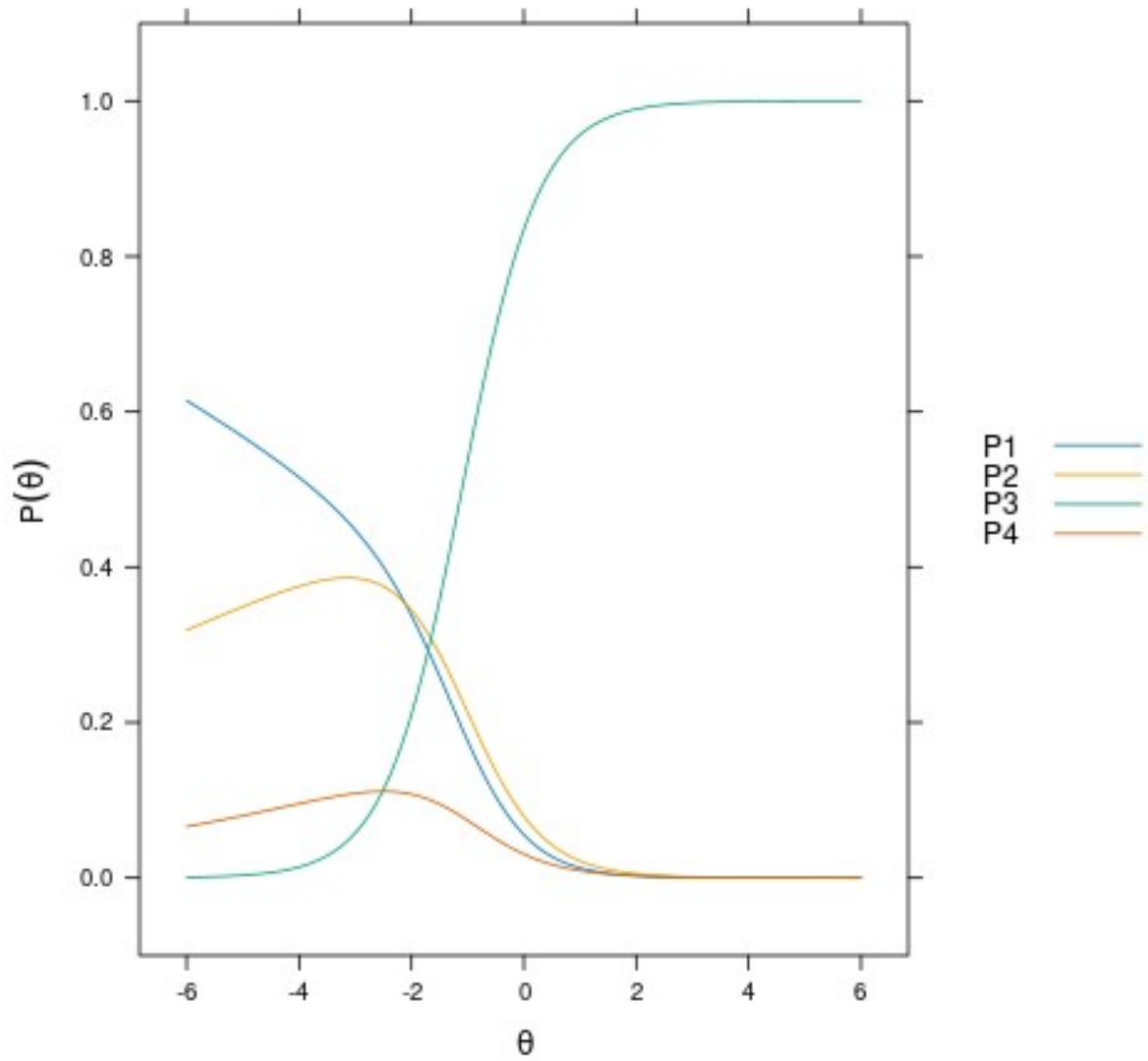
Fonte: Elaborado própria

**Gráfico 61** – Função de Probabilidade – Item 23

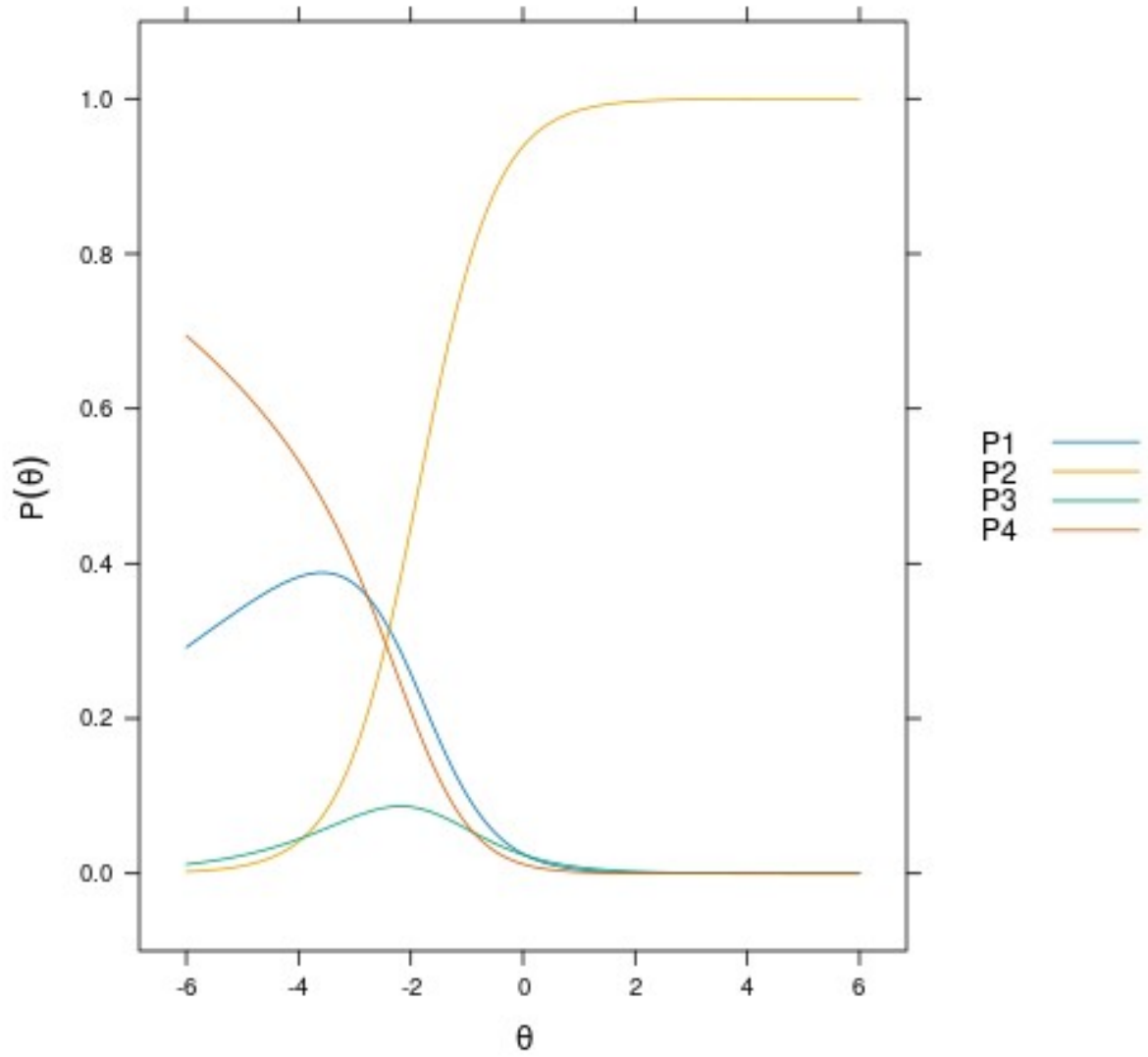
**Fonte:** Elaborado própria

**Gráfico 62** – Função de Probabilidade – Item 24

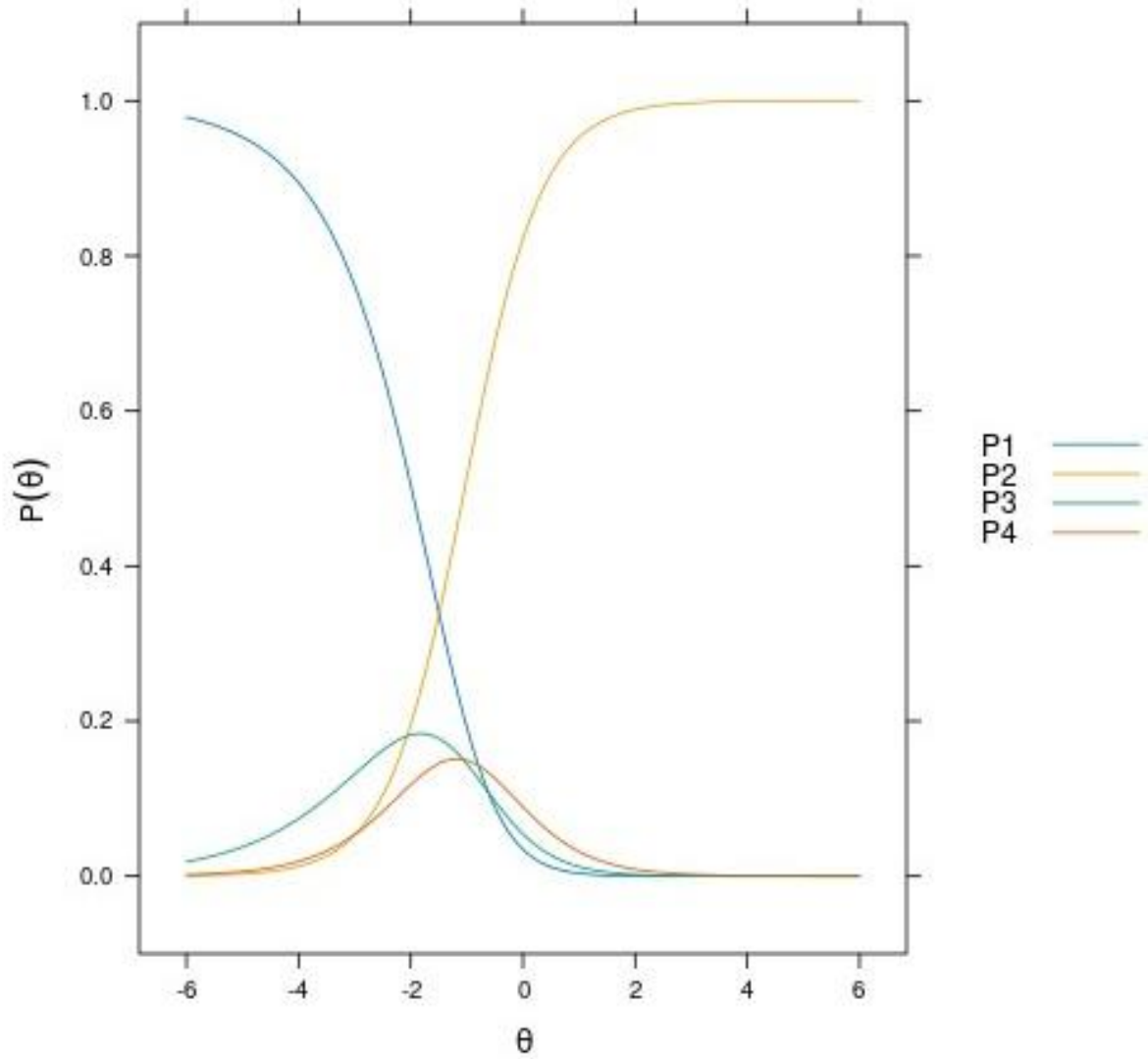
**Fonte:** Elaborado própria

**Gráfico 63** – Função de Probabilidade – Item 25

Fonte: Elaborado própria

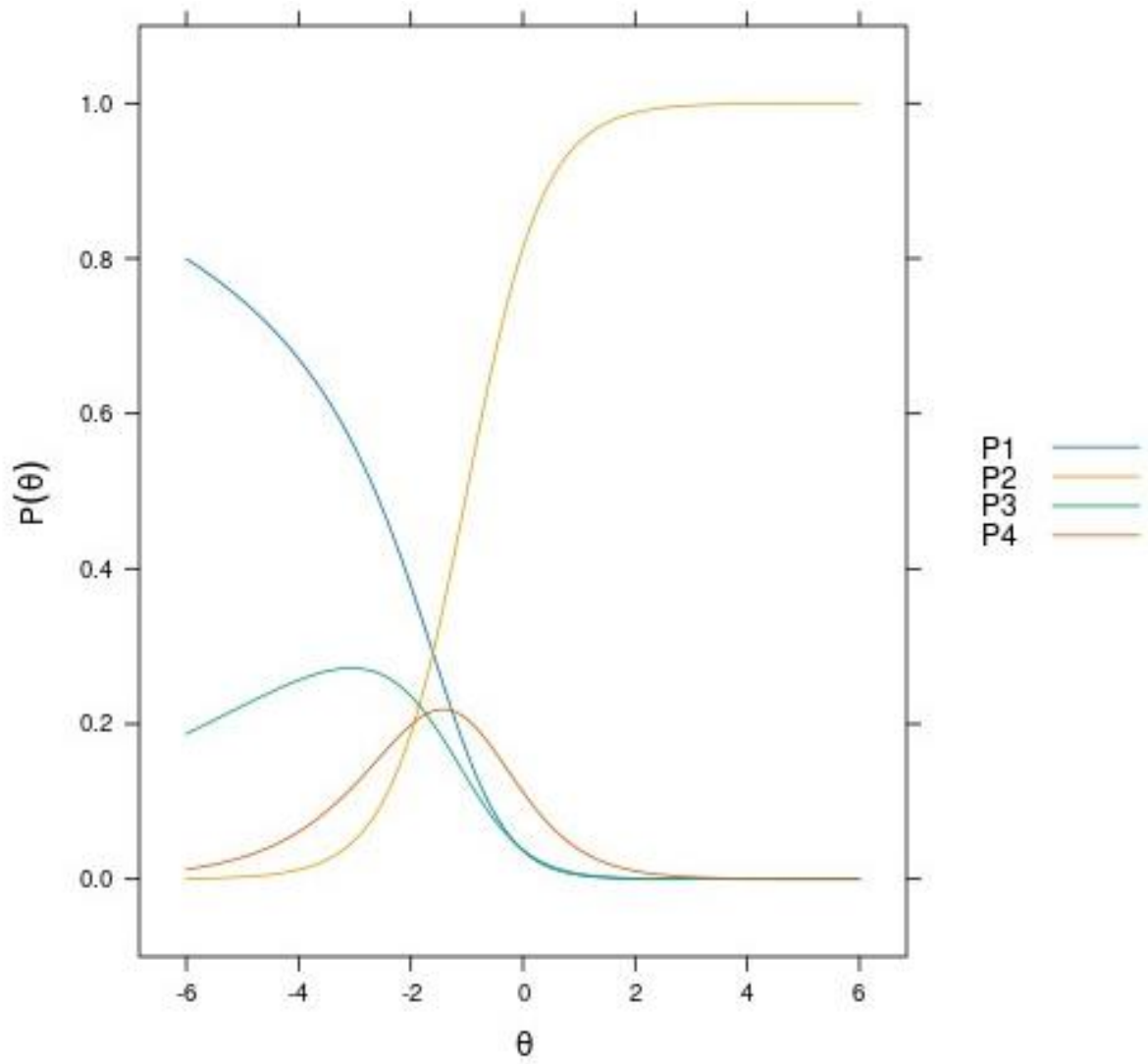
**Gráfico 64** – Função de Probabilidade – Item 26

**Fonte:** Elaborado própria

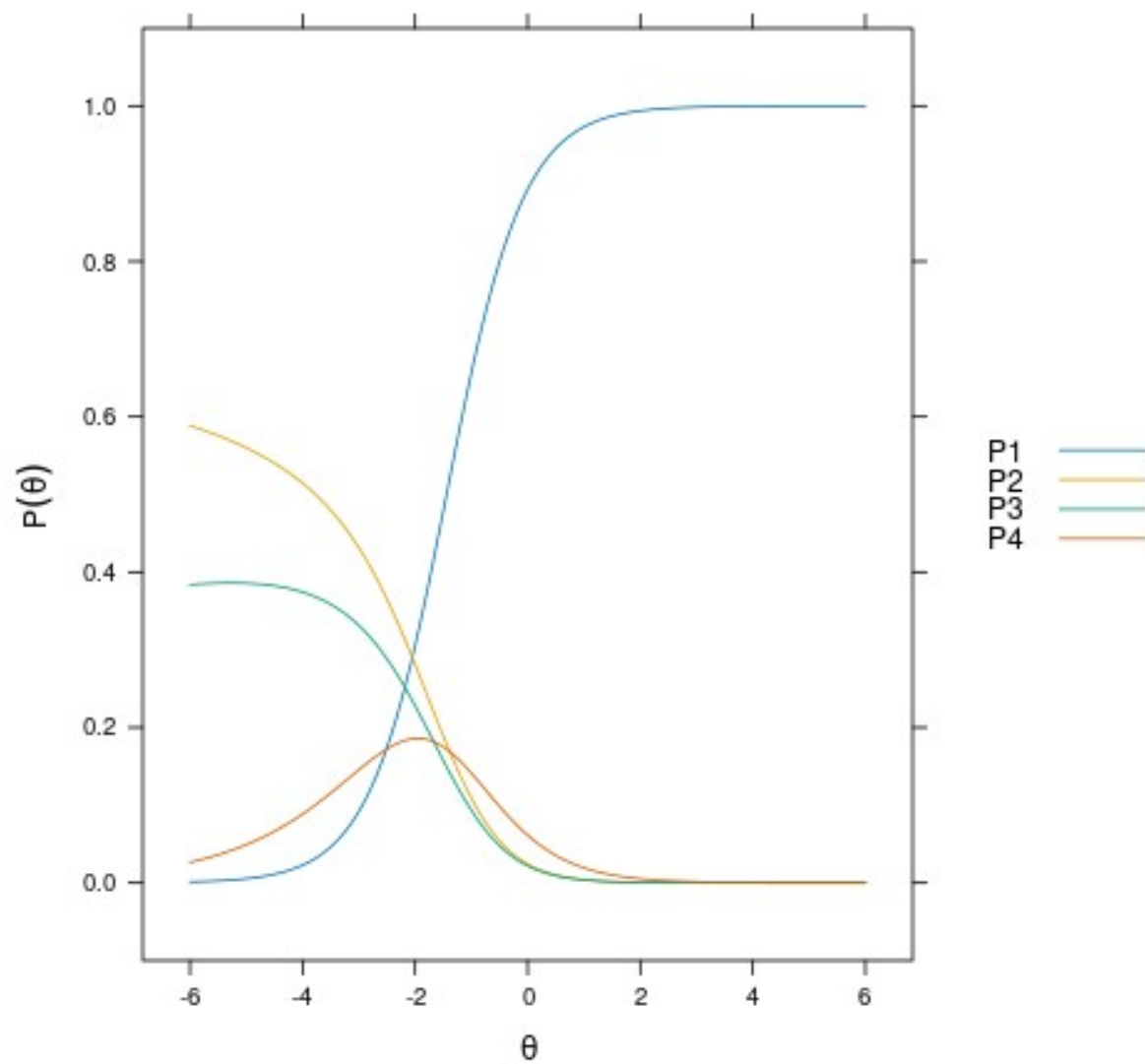
**Gráfico 65** – Função de Probabilidade – Item 27

Fonte: Elaborado própria

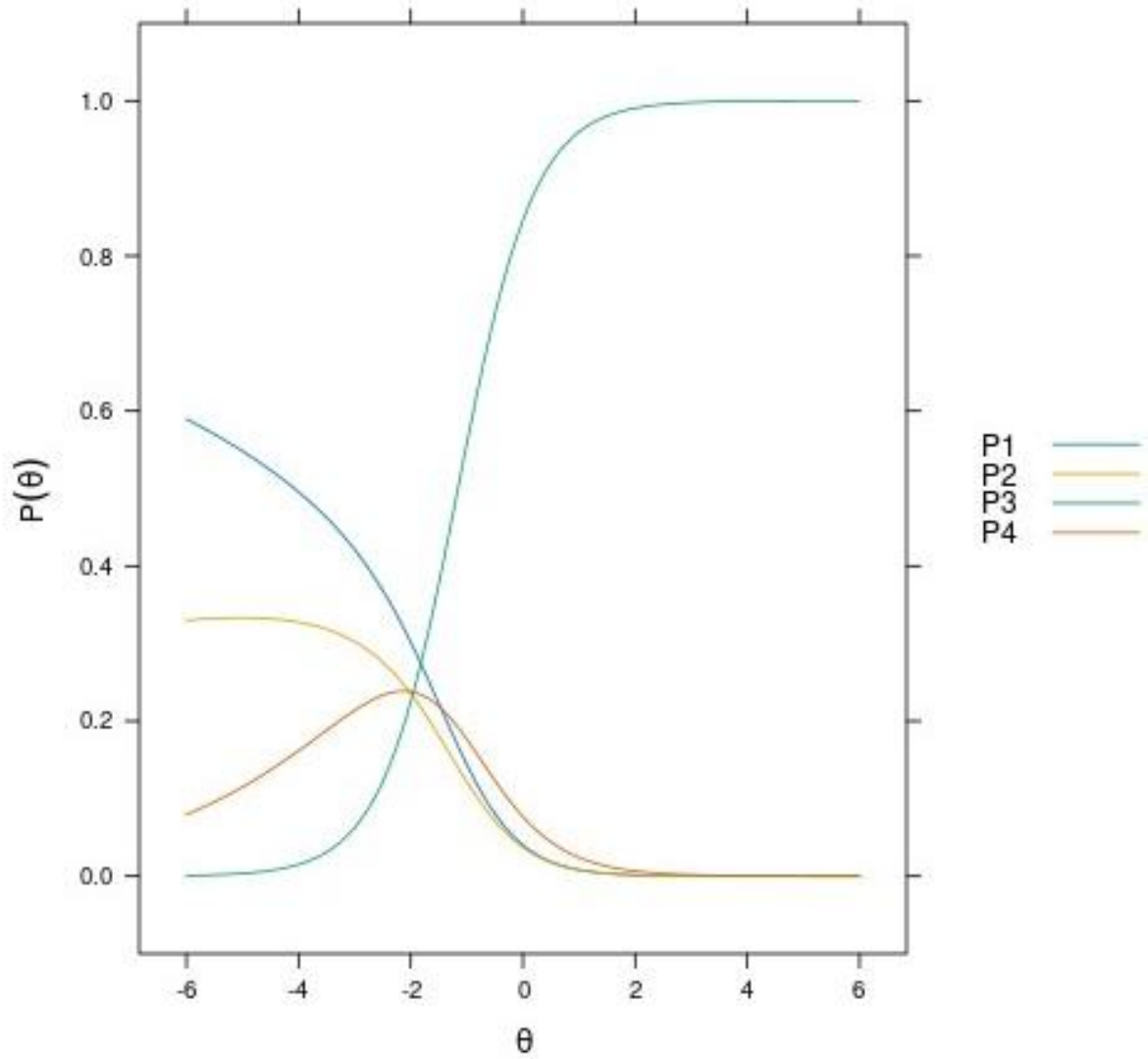


**Gráfico 66** – Função de Probabilidade – Item 28

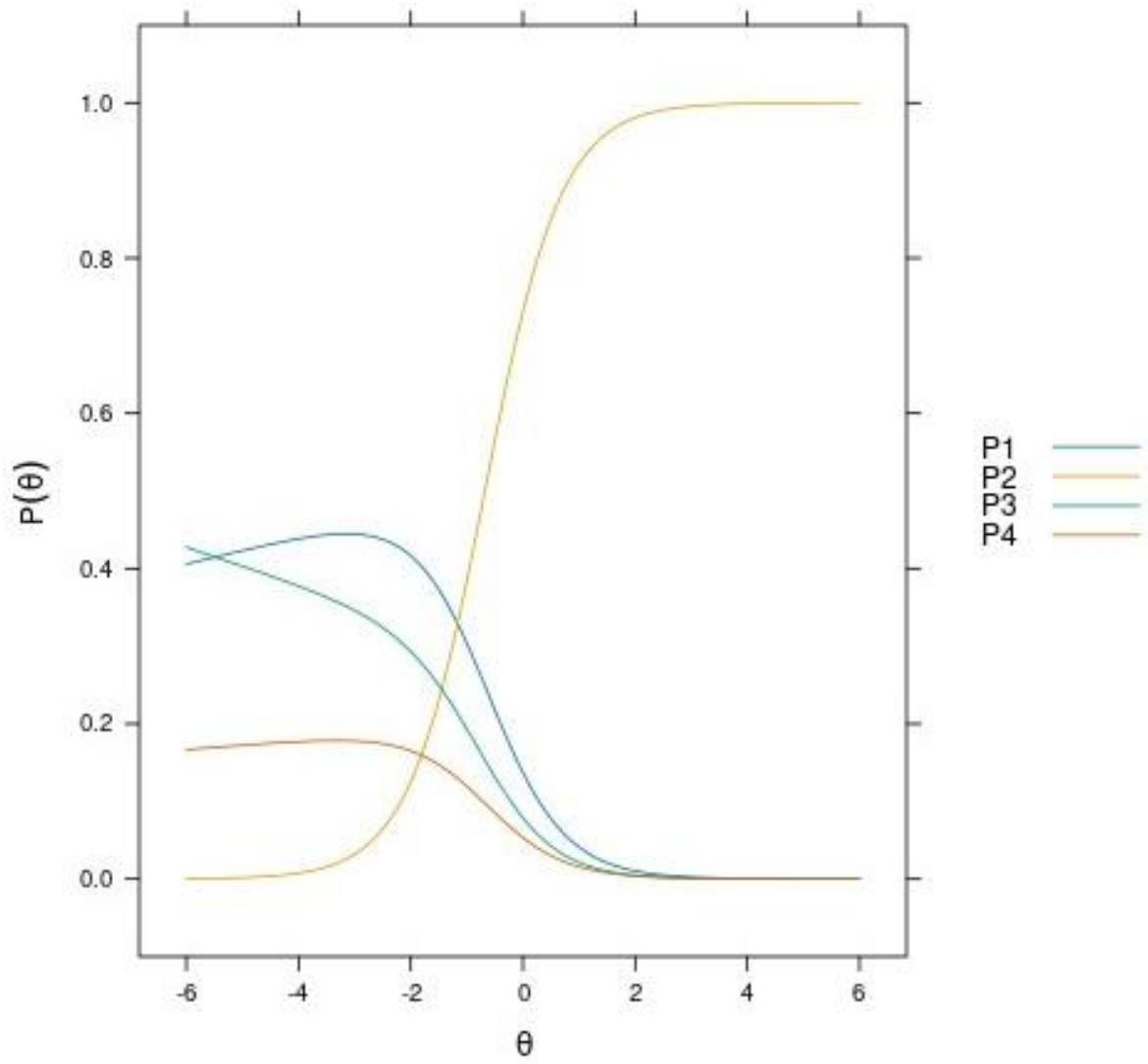
Fonte: Elaborado própria

**Gráfico 67** – Função de Probabilidade – Item 29

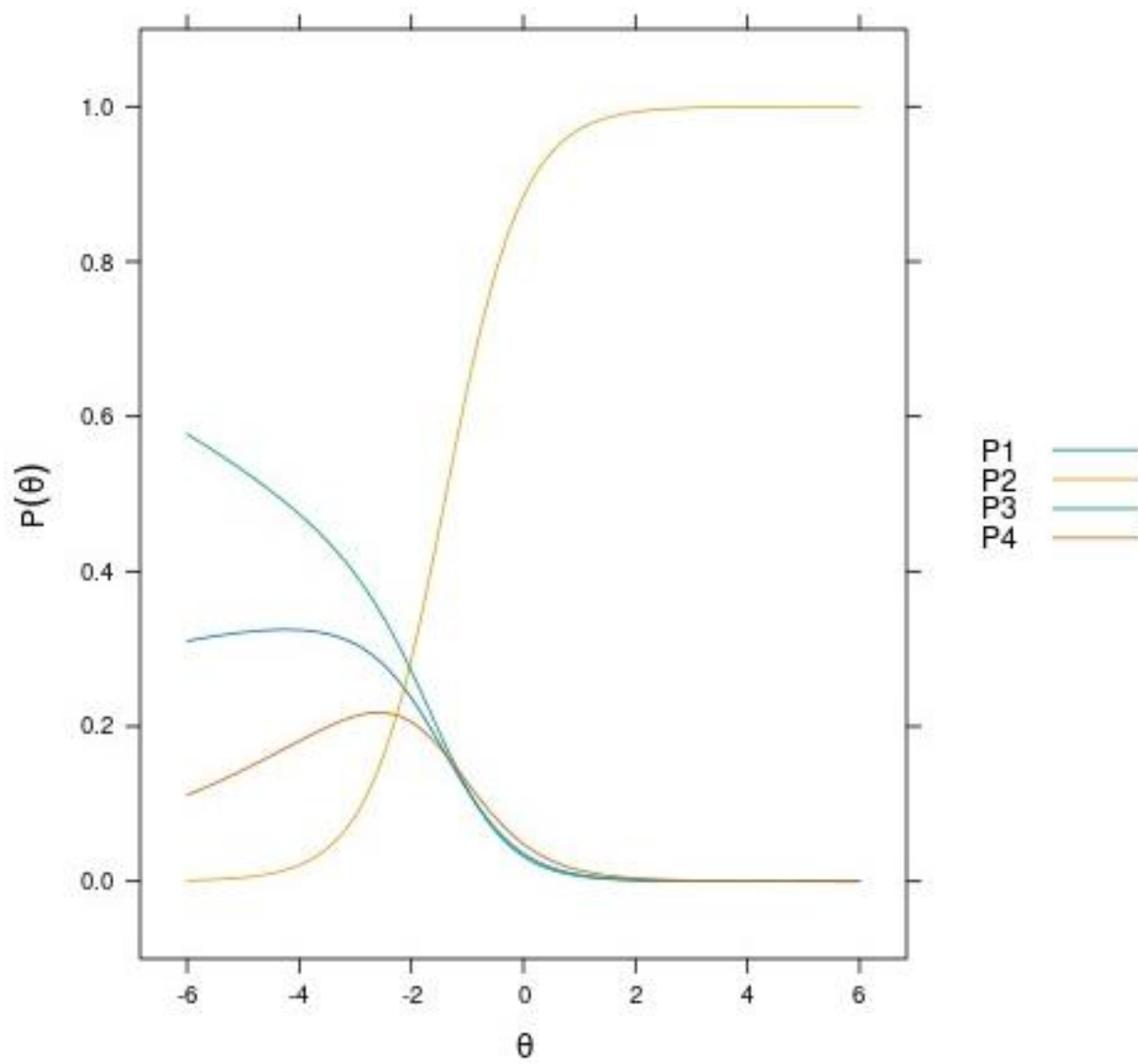
**Fonte:** Elaborado própria

**Gráfico 68** – Função de Probabilidade – Item 30

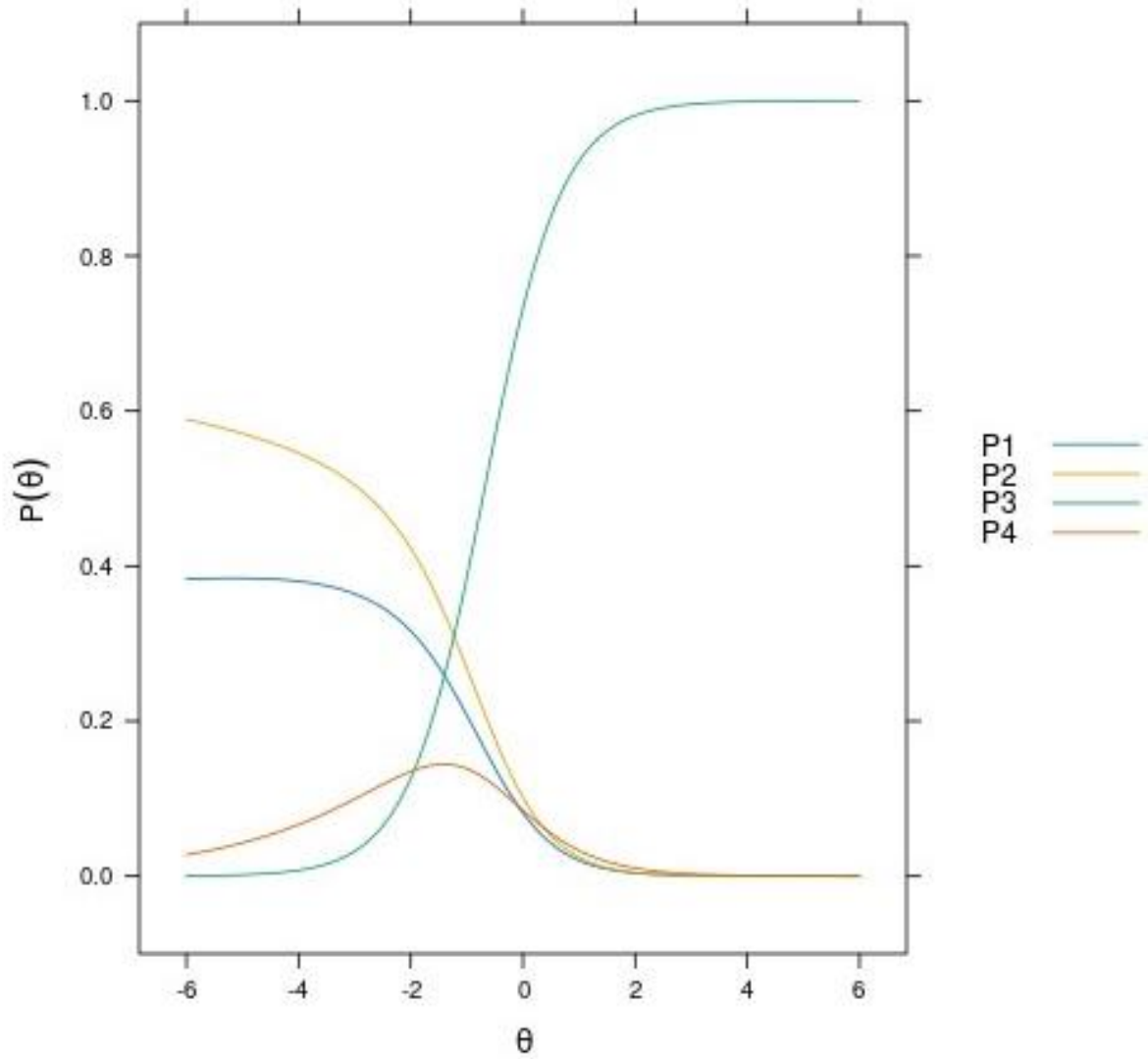
Fonte: Elaborado própria

**Gráfico 69** – Função de Probabilidade – Item 31

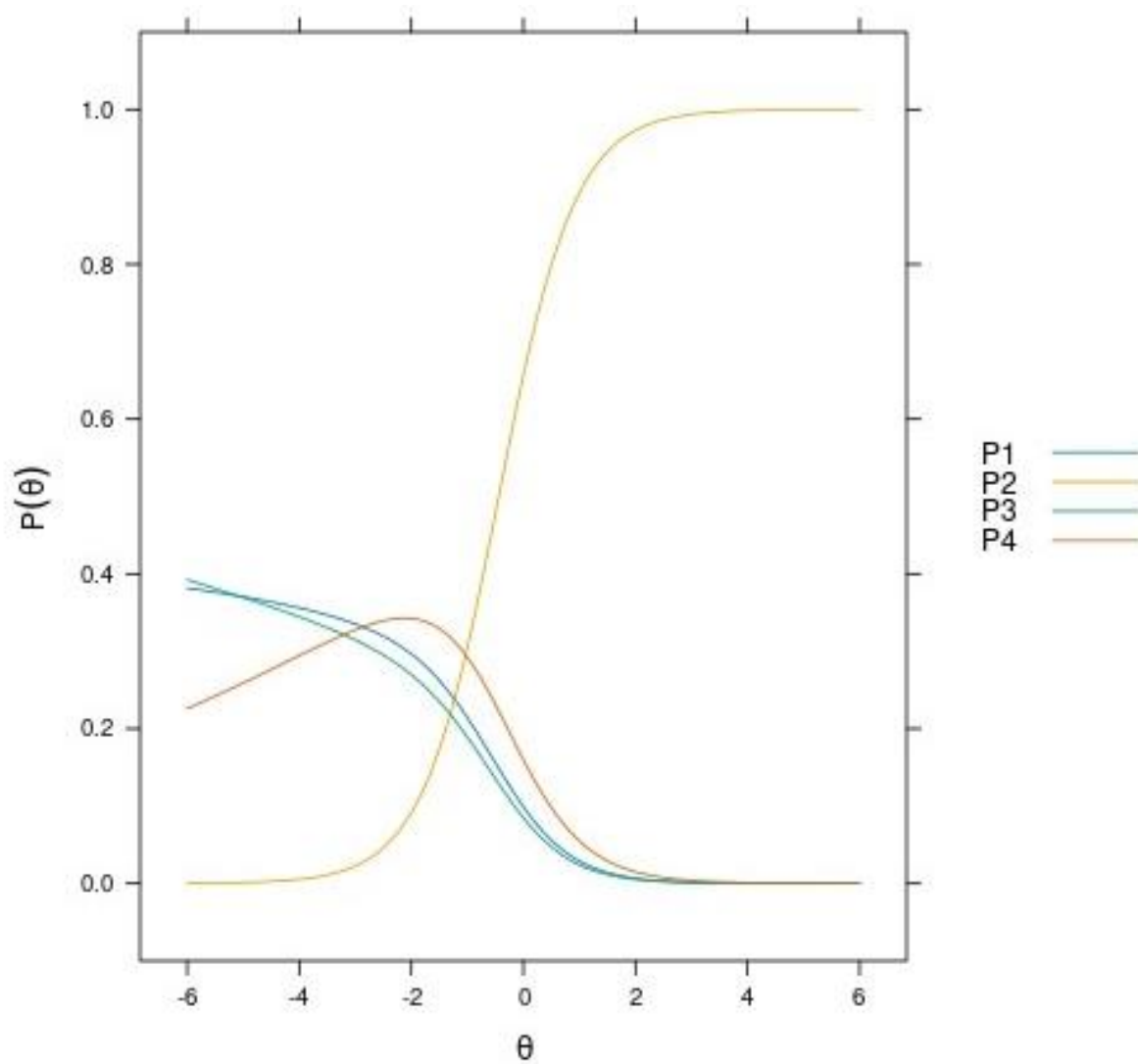
**Fonte:** Elaborado própria

**Gráfico 70** – Função de Probabilidade – Item 32

**Fonte:** Elaborado própria

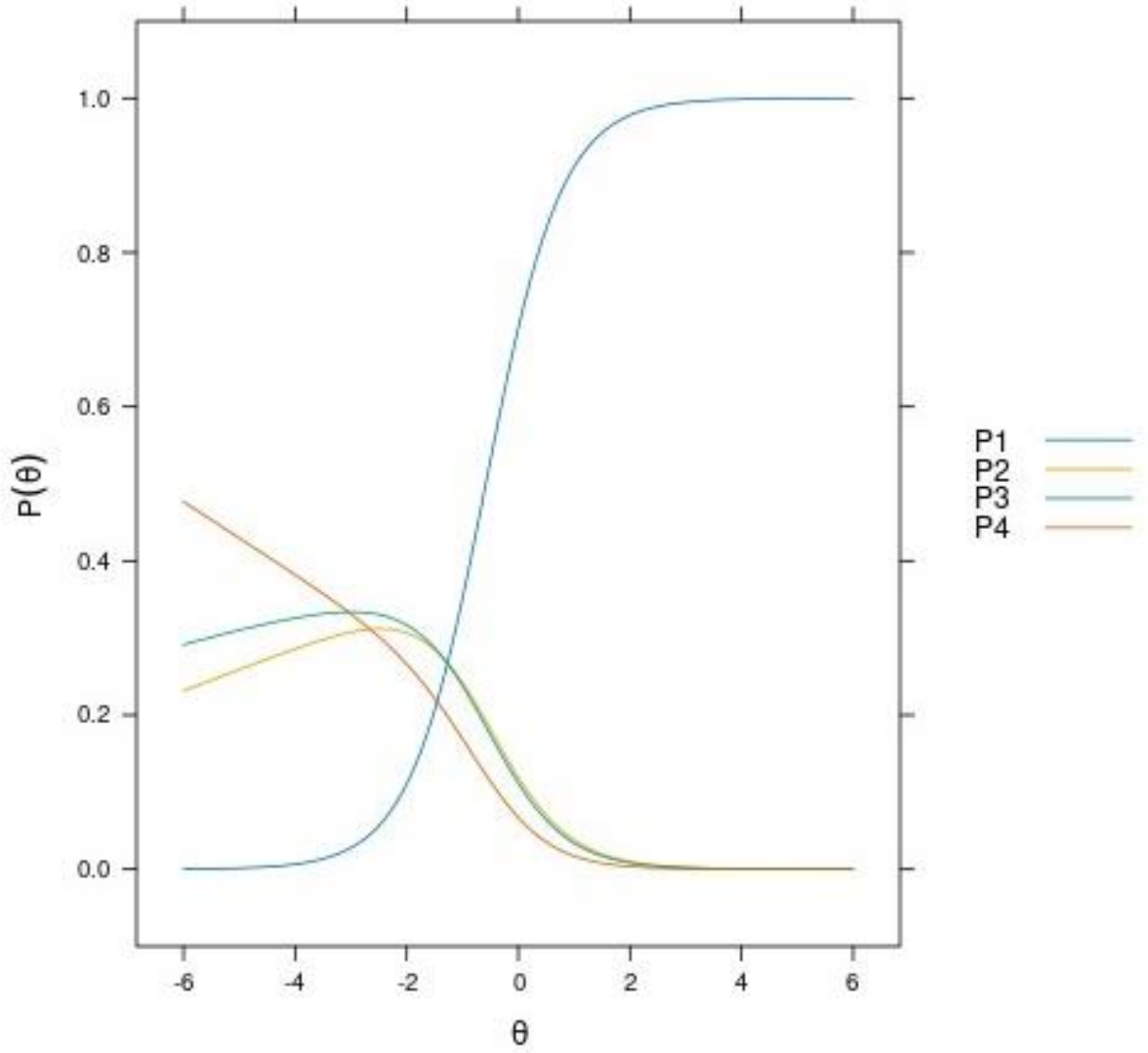
**Gráfico 71 – Função de Probabilidade – Item 33**

Fonte: Elaborado própria

**Gráfico 72 – Função de Probabilidade – Item 34**

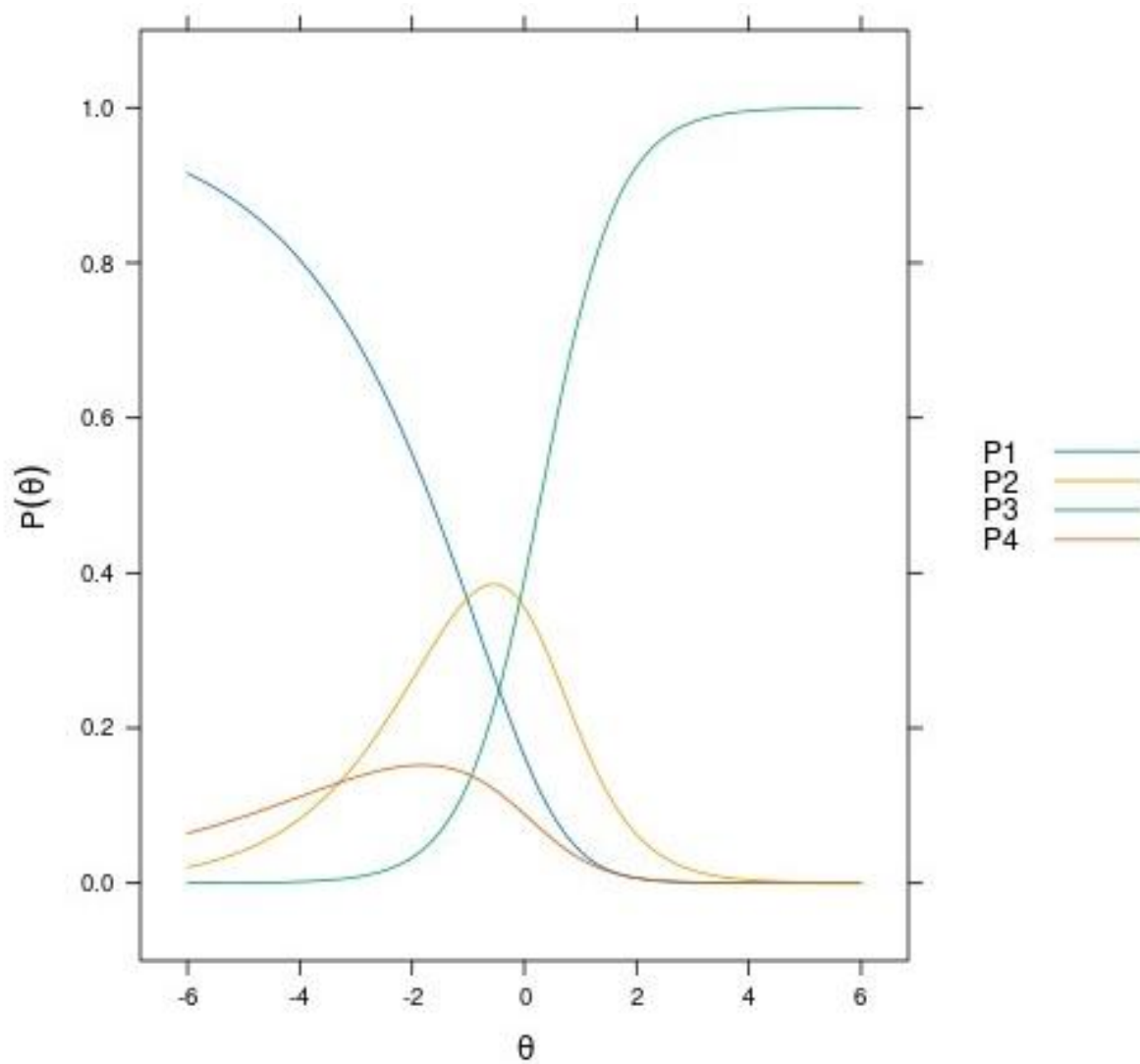
Fonte: Elaborado própria

**Gráfico 73** – Função de Probabilidade – Item 35

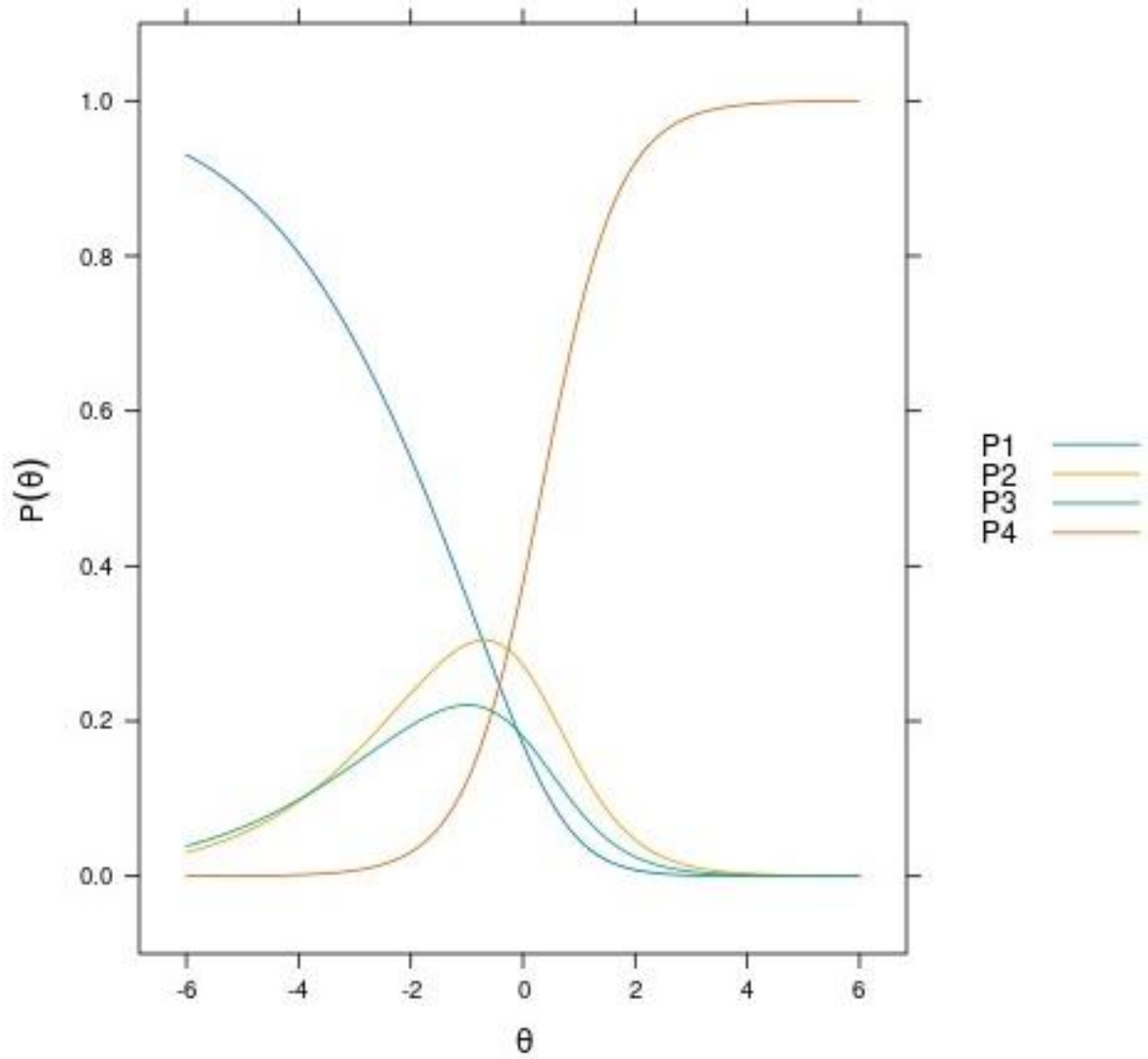


**Fonte:** Elaborado própria

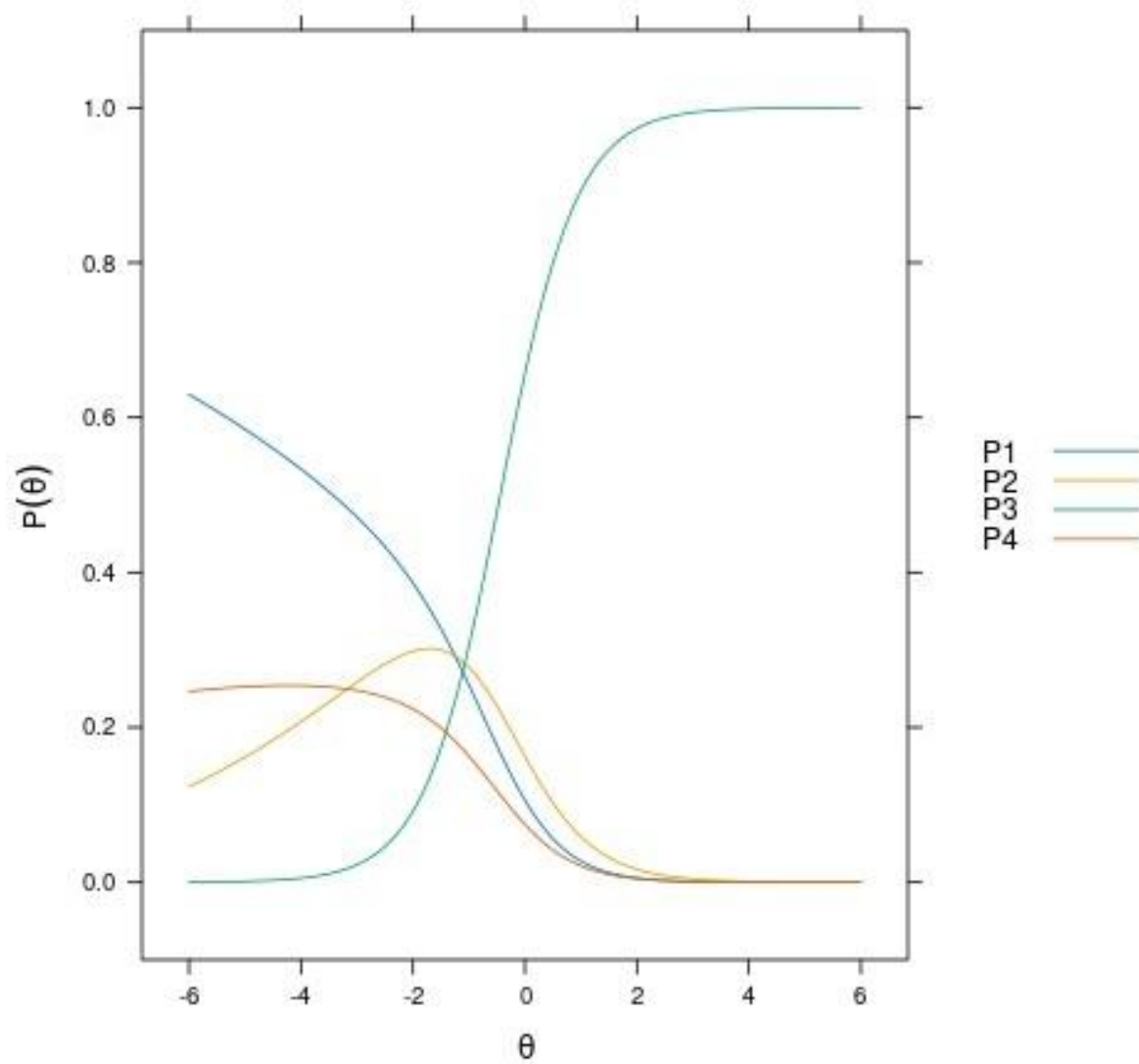


**Gráfico 74** – Função de Probabilidade – Item 36

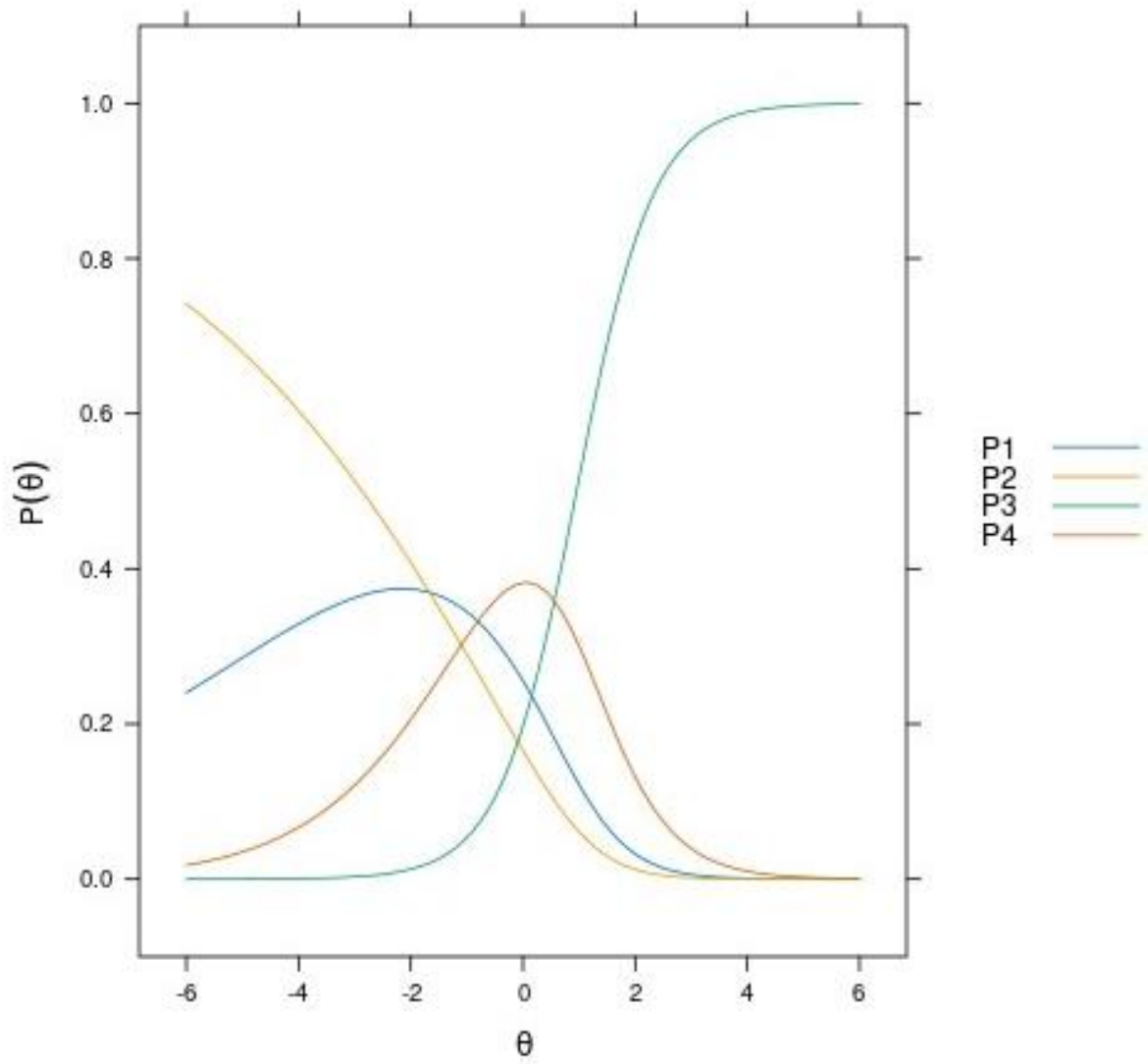
Fonte: Elaborado própria

**Gráfico 75** – Função de Probabilidade – Item 37

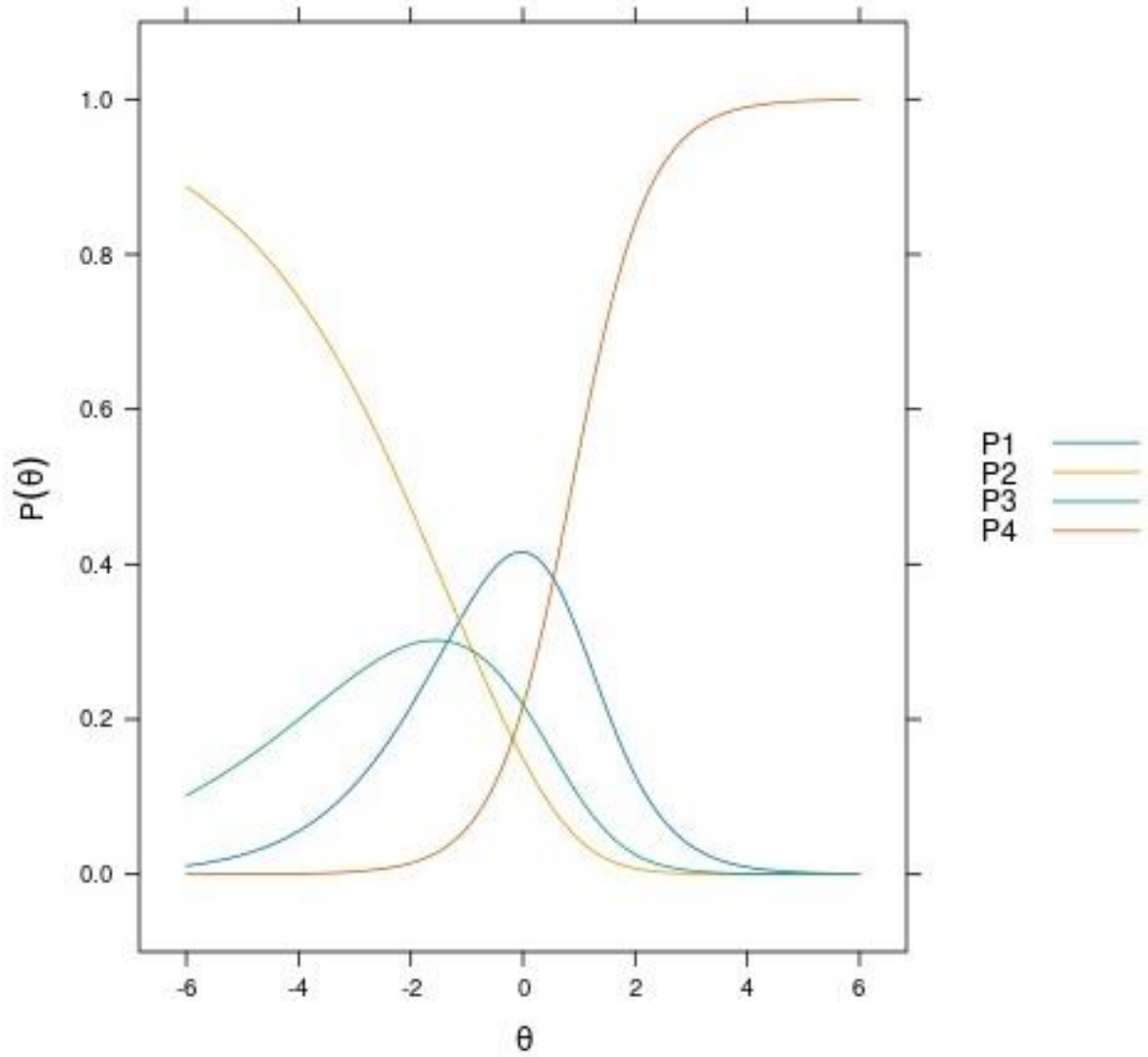
Fonte: Elaborado própria

**Gráfico 76** – Função de Probabilidade – Item 38

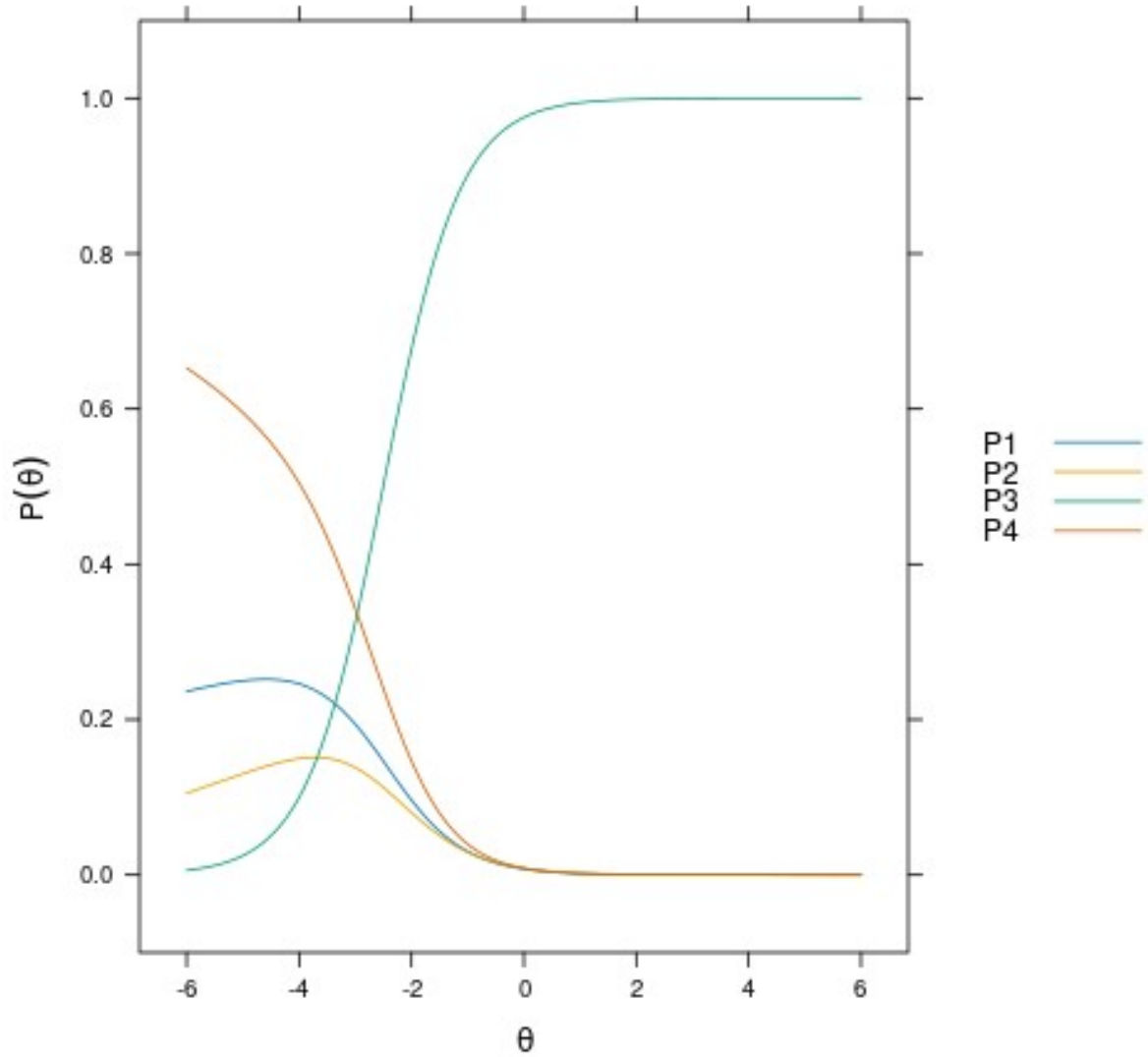
**Fonte:** Elaborado própria

**Gráfico 77** – Função de Probabilidade – Item 39

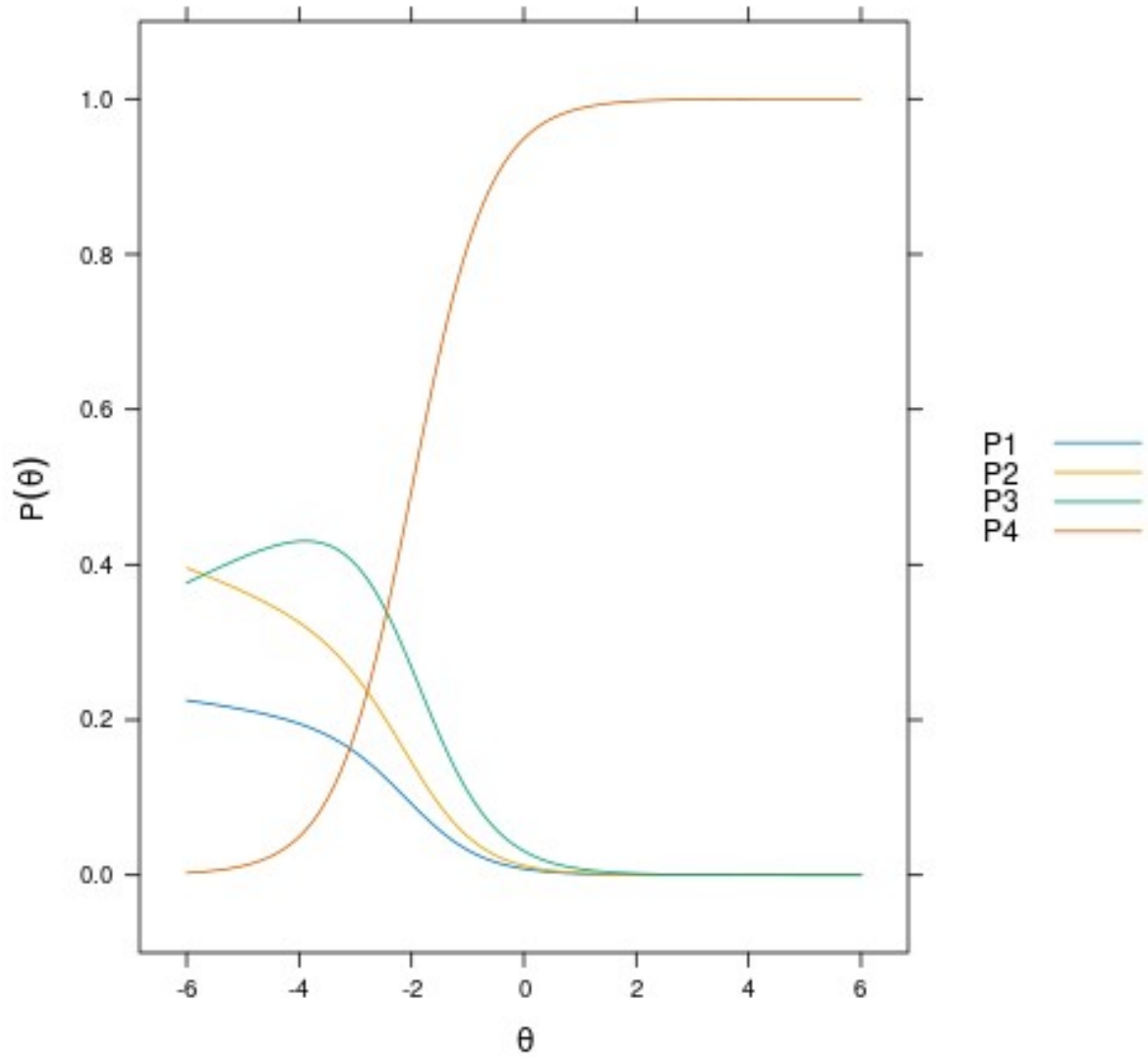
**Fonte:** Elaborado própria

**Gráfico 78** – Função de Probabilidade – Item 40

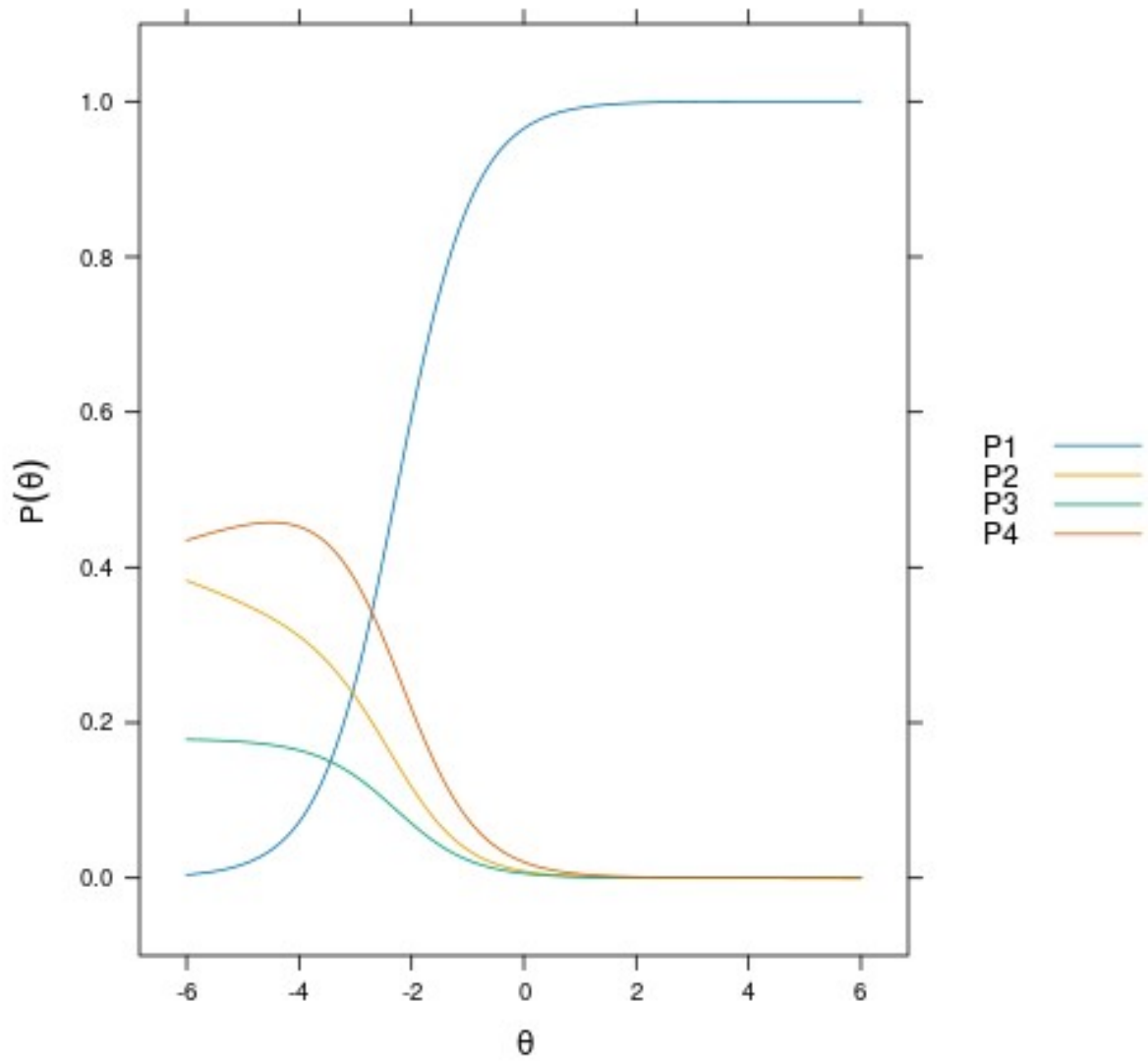
Fonte: Elaborado própria

**Gráfico 79 – Função de Probabilidade – Item 41****Probability Function for Item 41**

Fonte: Elaborado própria

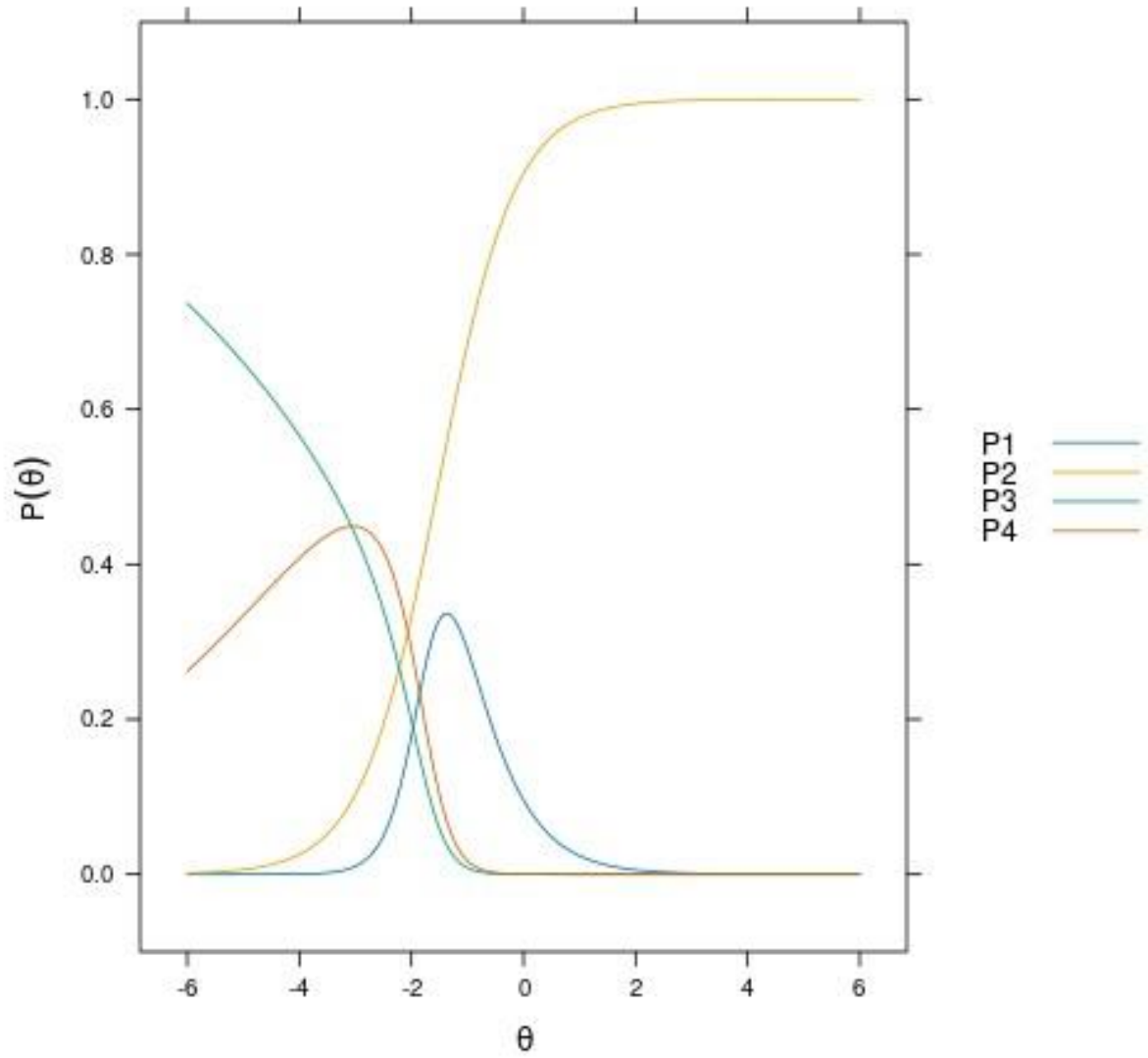
**Gráfico 80** – Função de Probabilidade – Item 42

**Fonte:** Elaborado própria

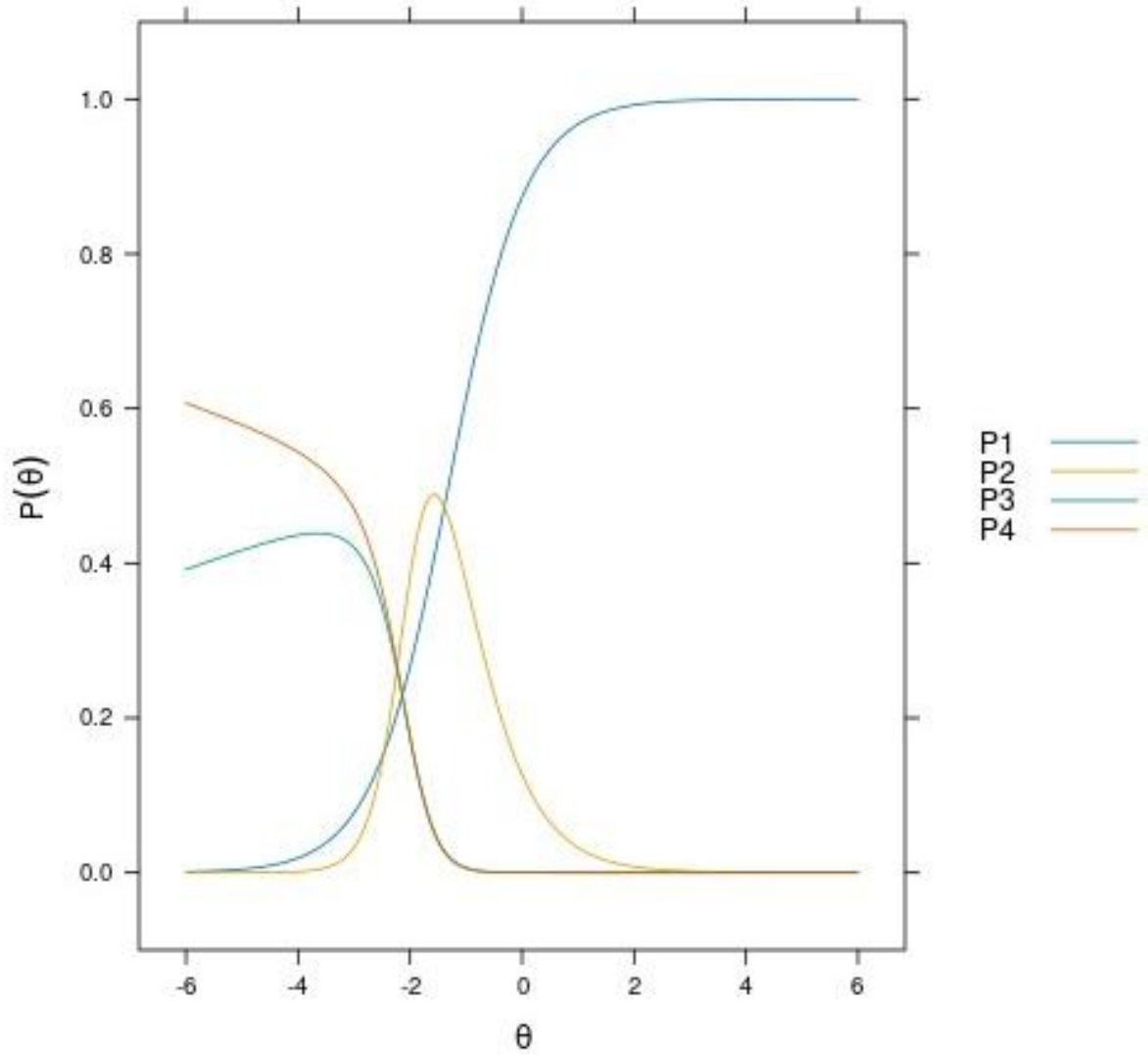
**Gráfico 81** – Função de Probabilidade – Item 43

**Fonte:** Elaborado própria

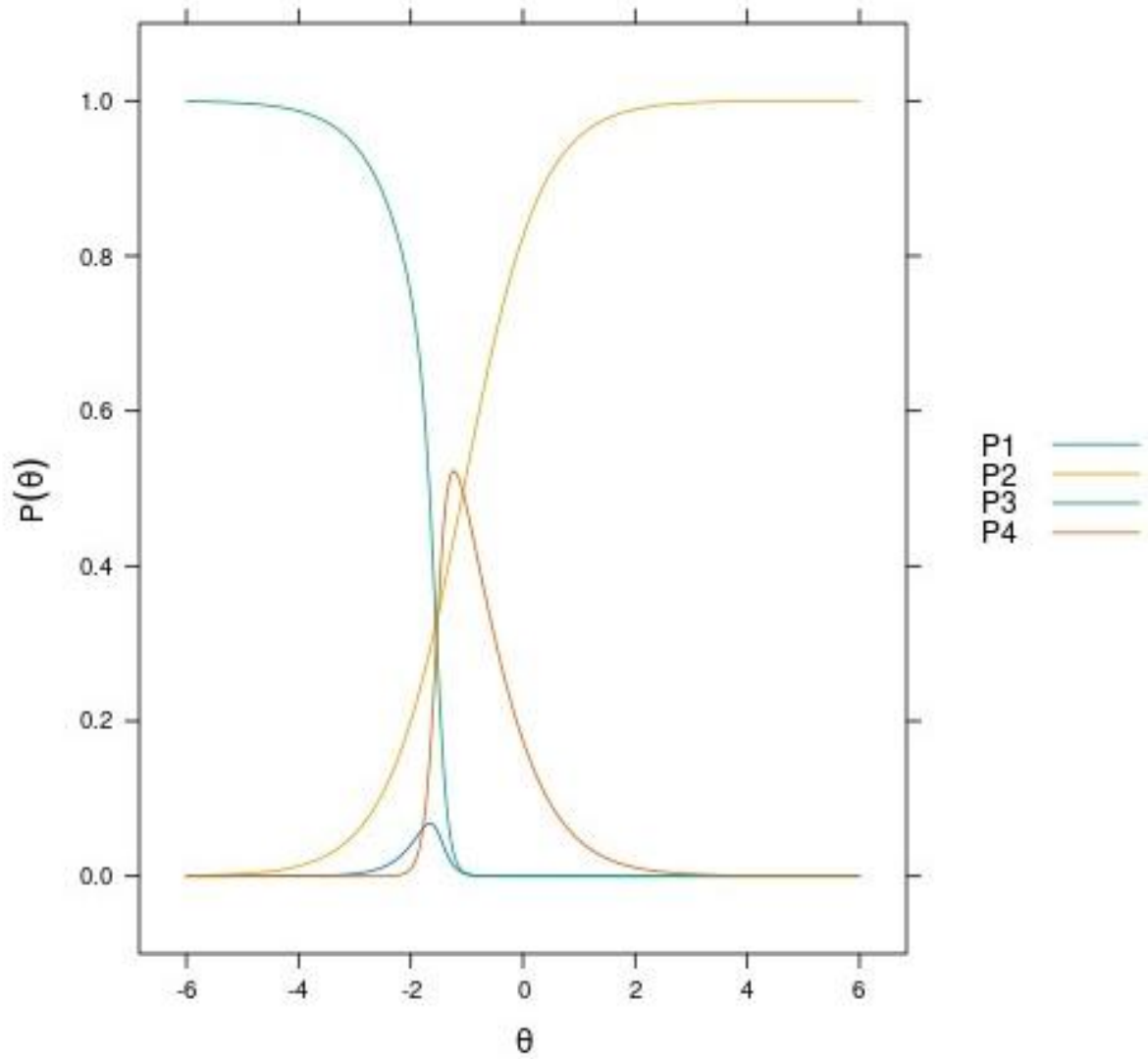


**Gráfico 82** – Função de Probabilidade – Item 44

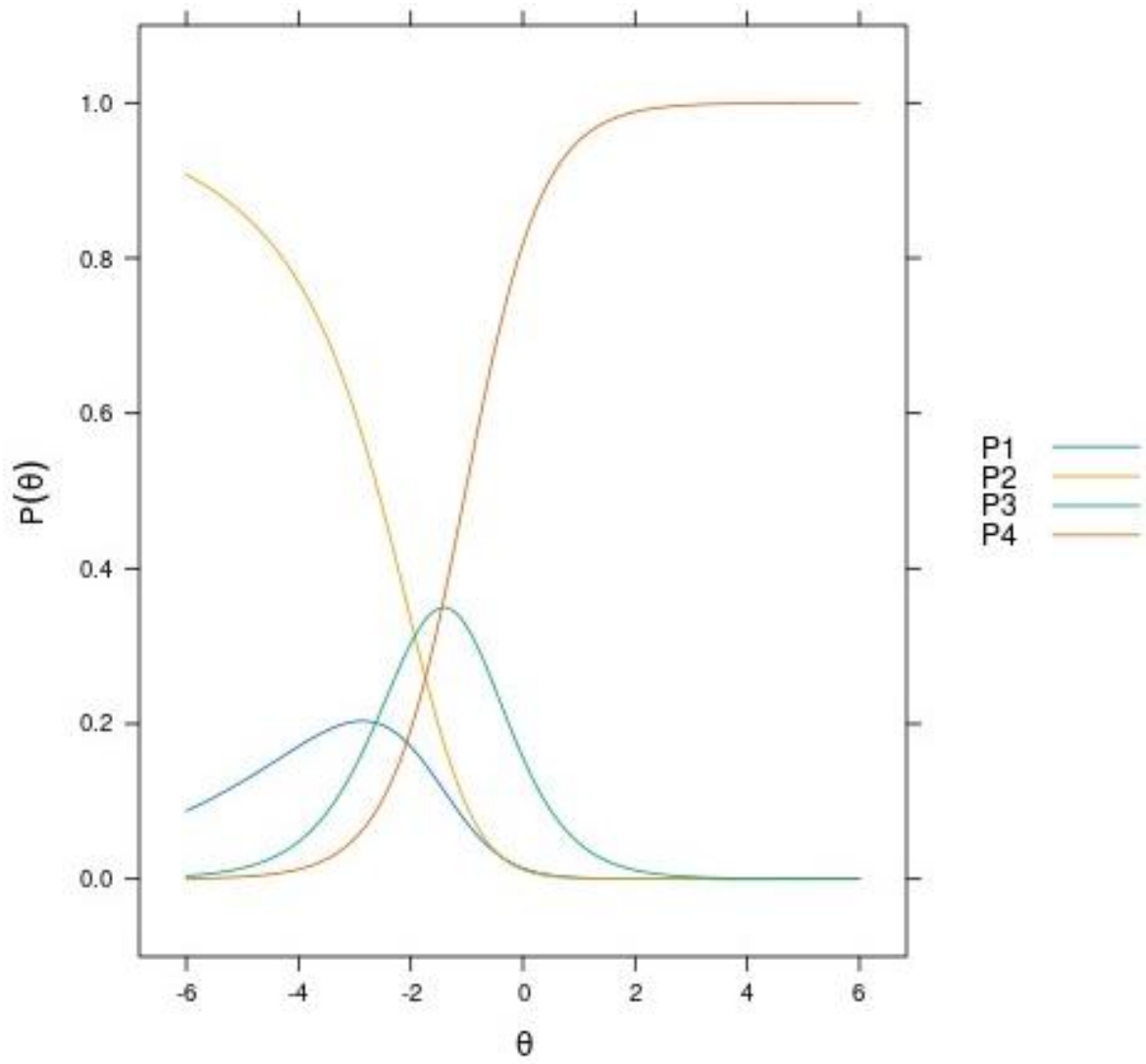
**Fonte:** Elaborado própria

**Gráfico 83** – Função de Probabilidade – Item 45

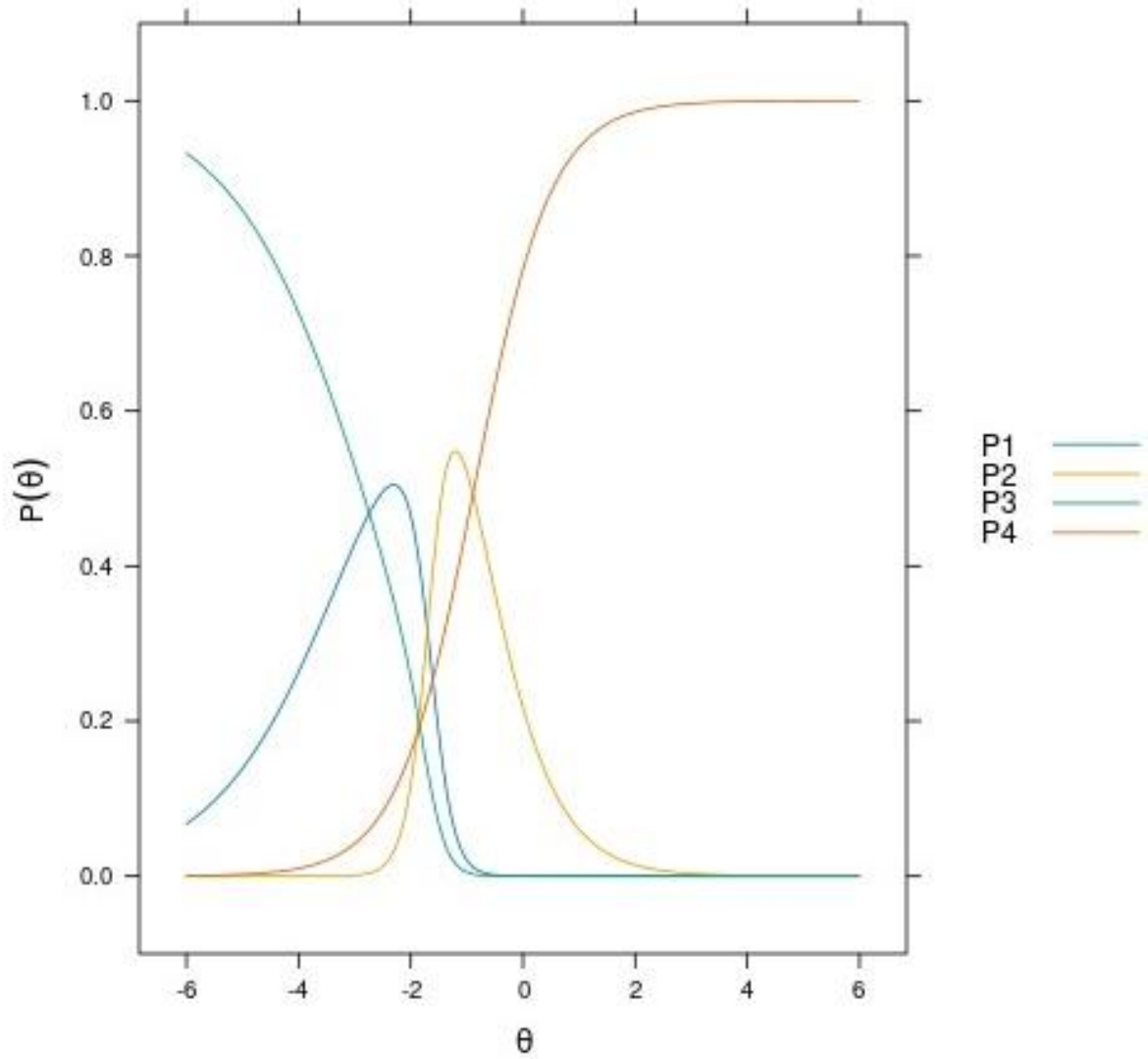
**Fonte:** Elaborado própria

**Gráfico 84** – Função de Probabilidade – Item 46

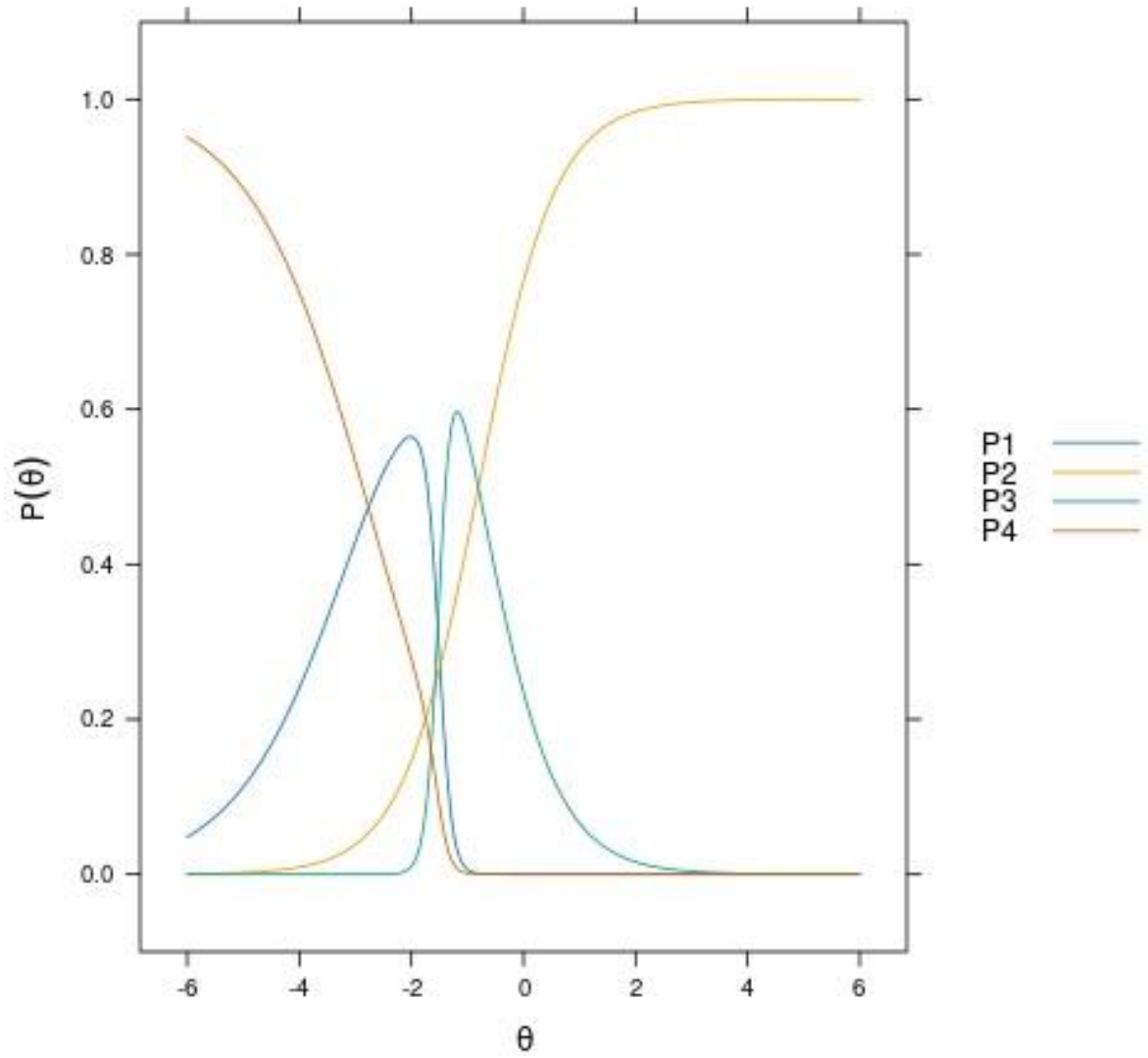
**Fonte:** Elaborado própria

**Gráfico 85** – Função de Probabilidade – Item 47

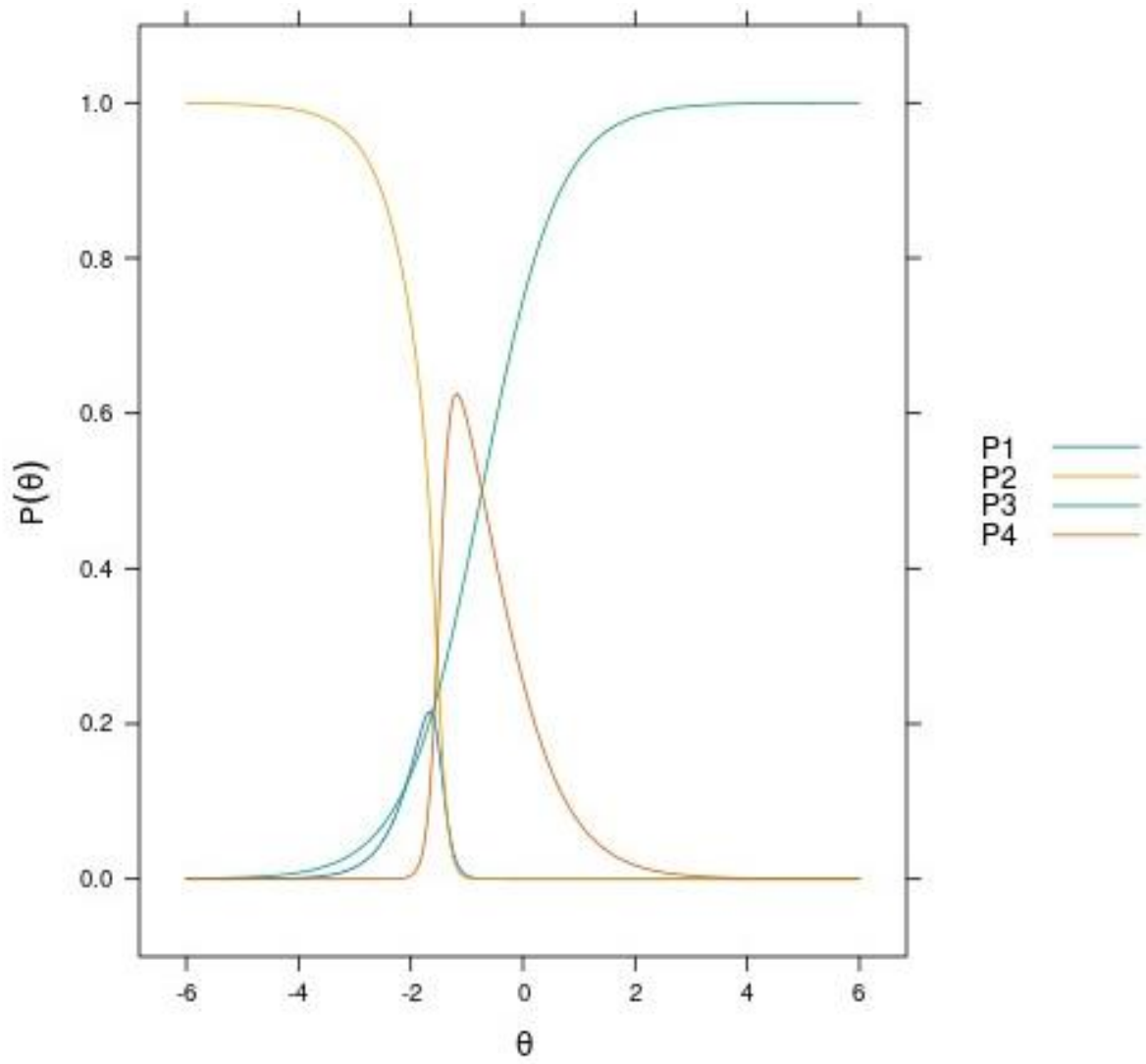
**Fonte:** Elaborado própria

**Gráfico 86** – Função de Probabilidade – Item 48

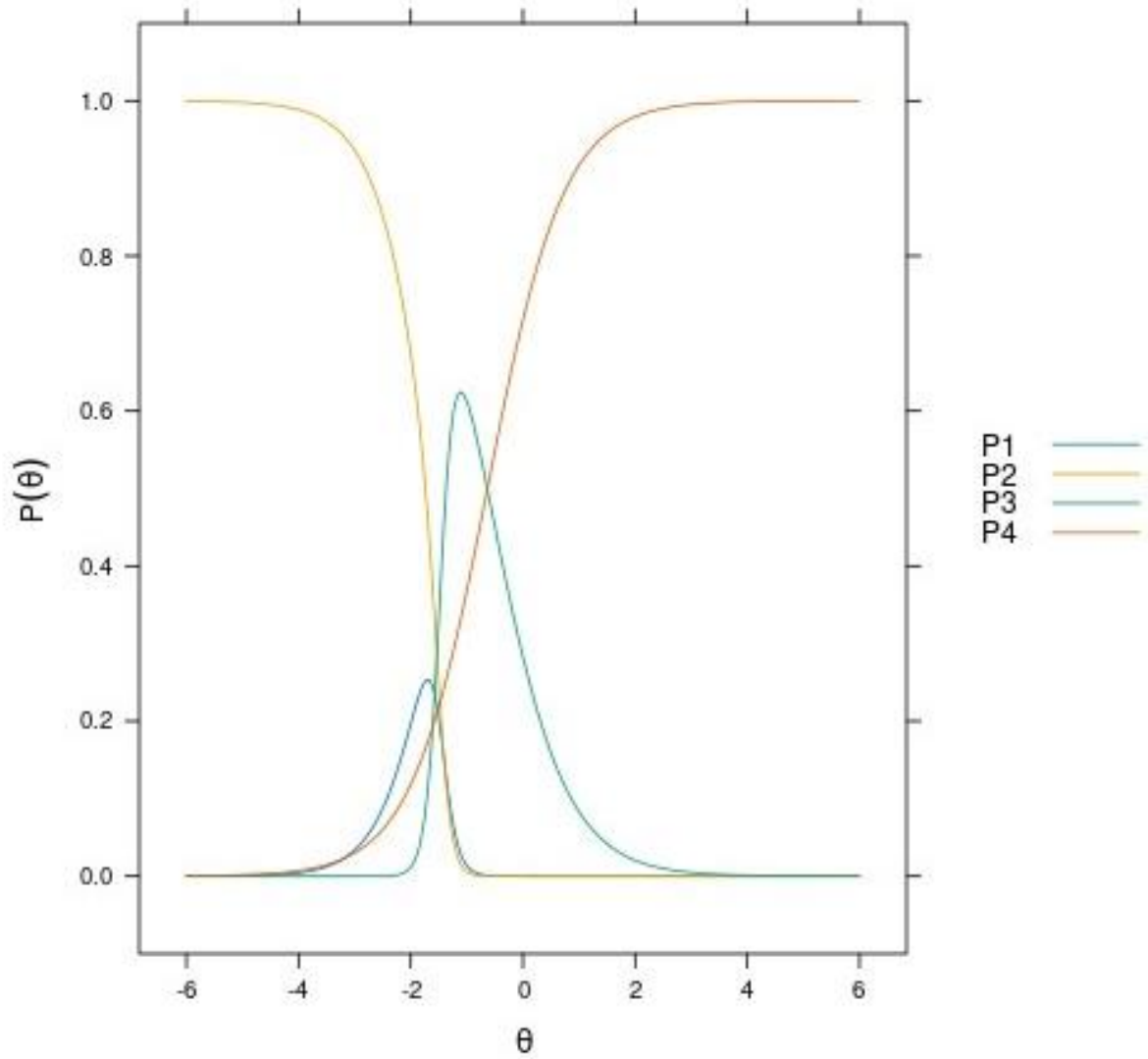
Fonte: Elaborado própria

**Gráfico 87** – Função de Probabilidade – Item 49

**Fonte:** Elaborado própria

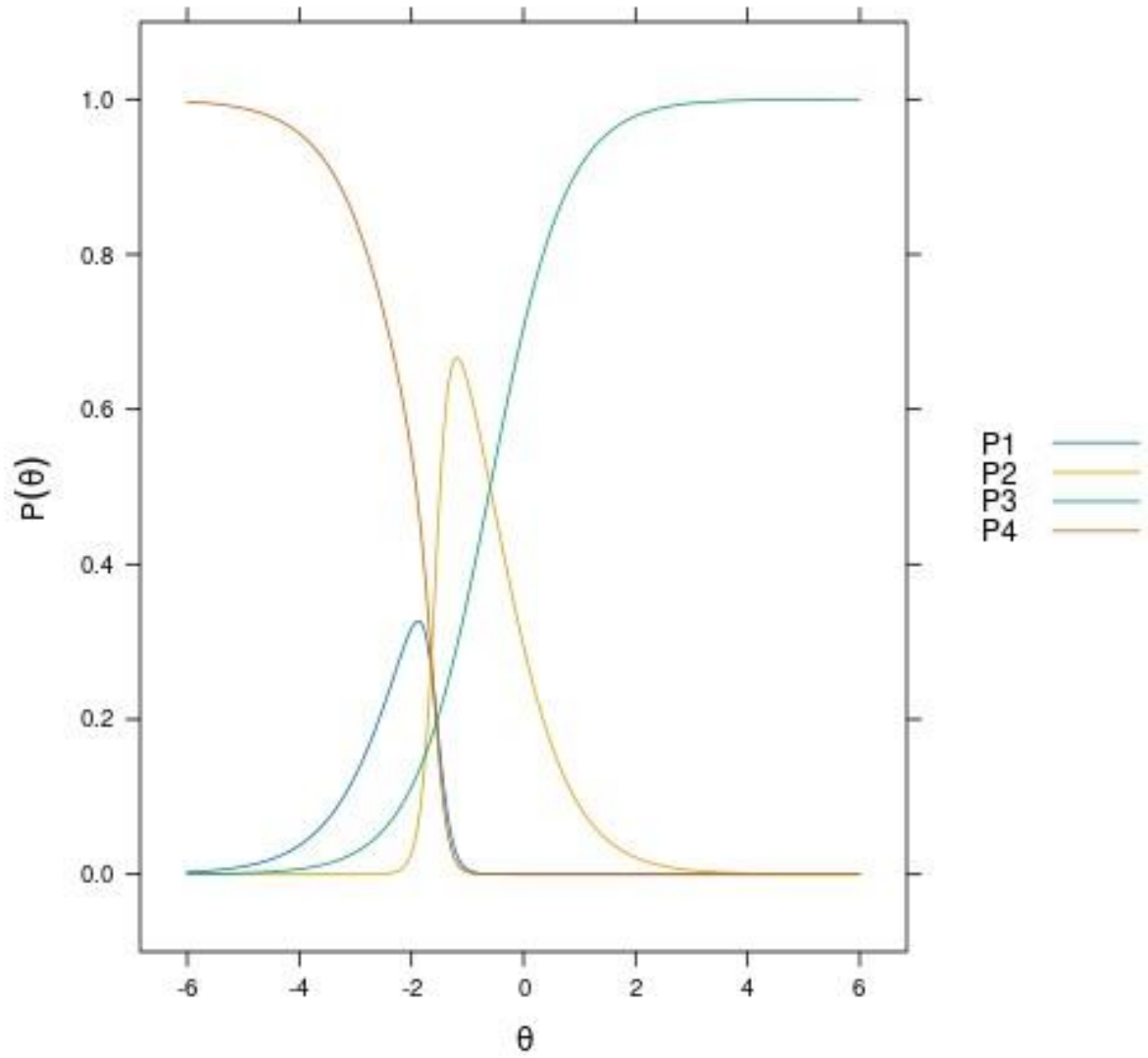
**Gráfico 88** – Função de Probabilidade – Item 50

**Fonte:** Elaborado própria

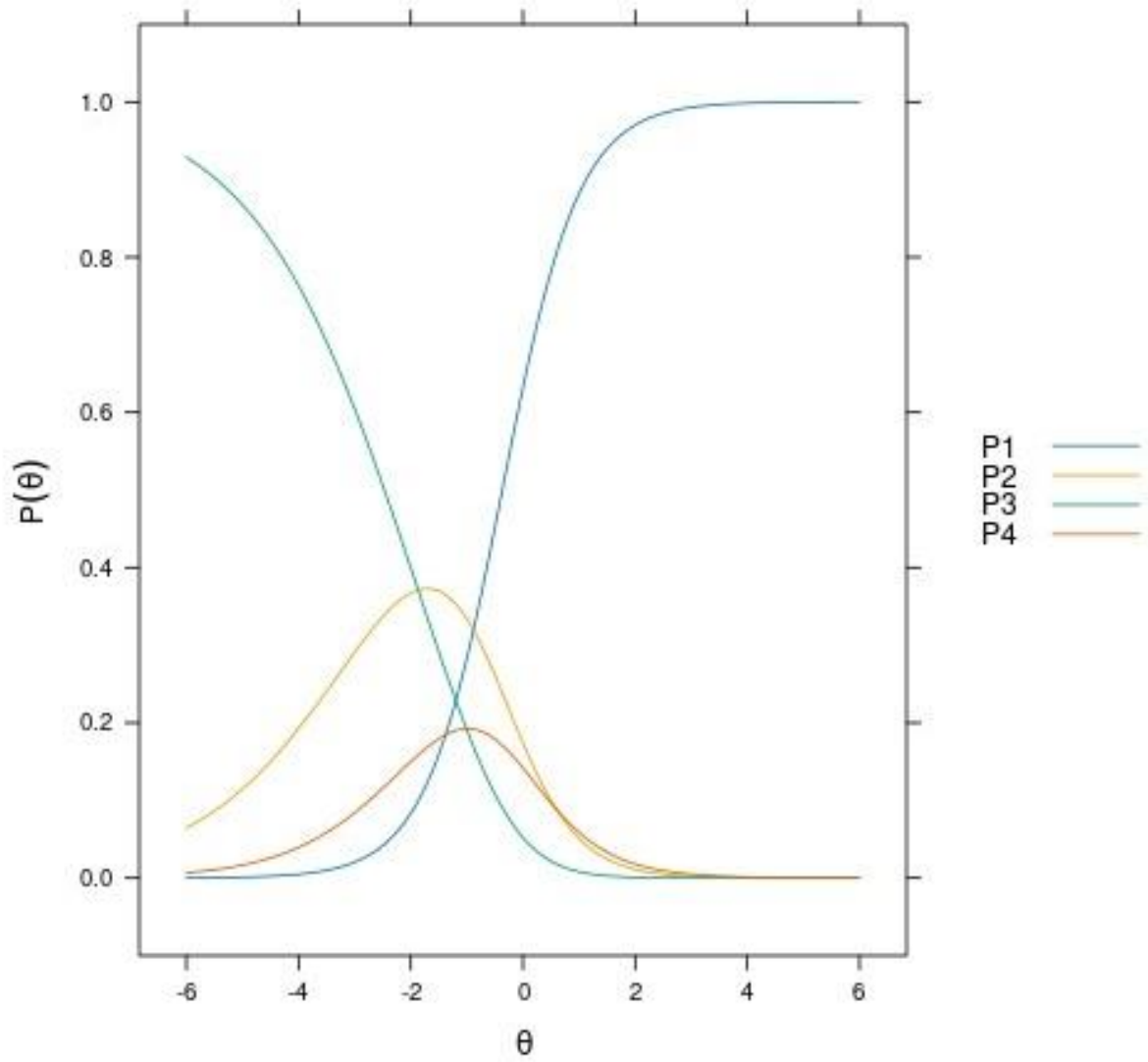
**Gráfico 89** – Função de Probabilidade – Item 51

**Fonte:** Elaborado própria

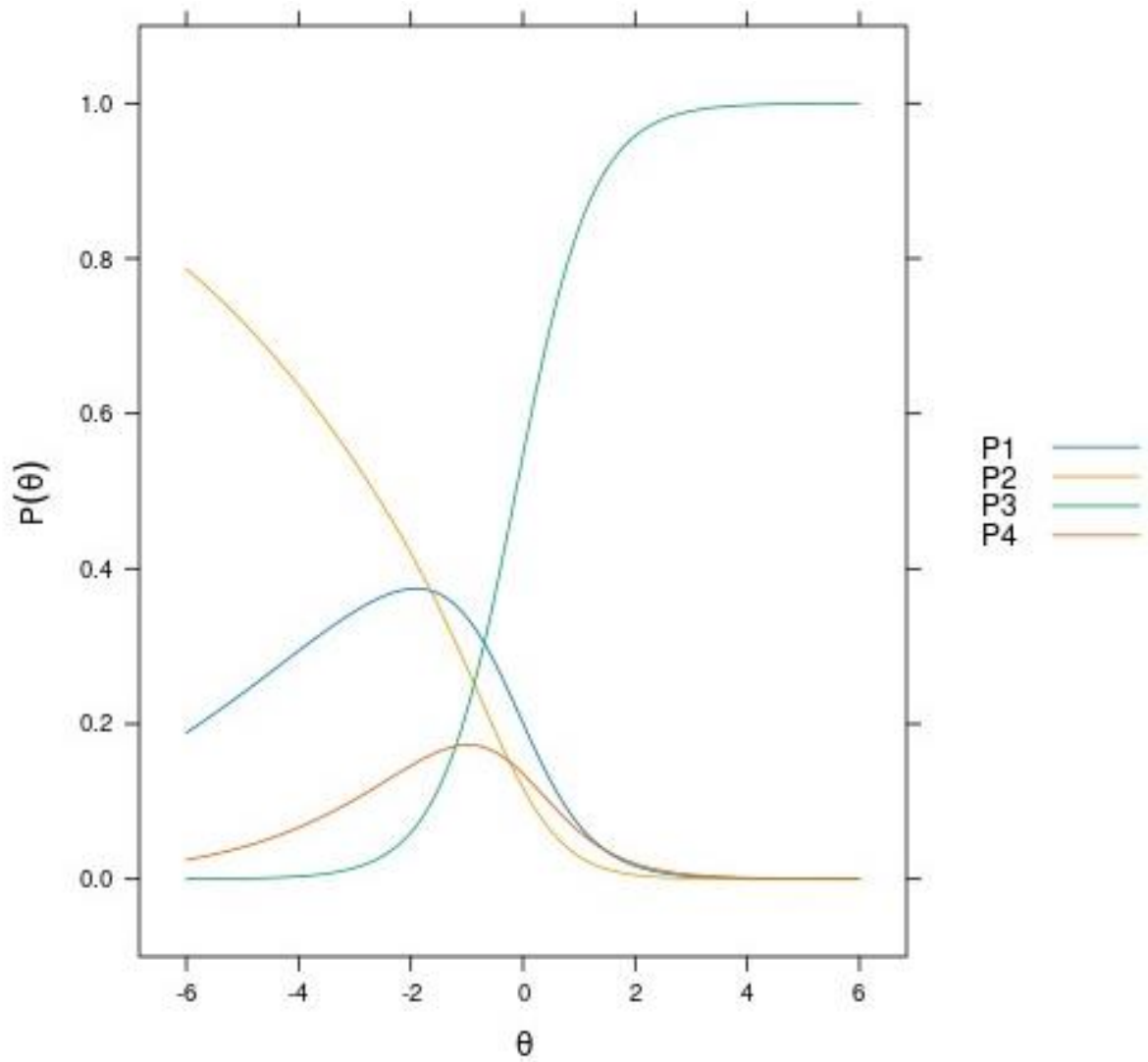


**Gráfico 90** – Função de Probabilidade – Item 52

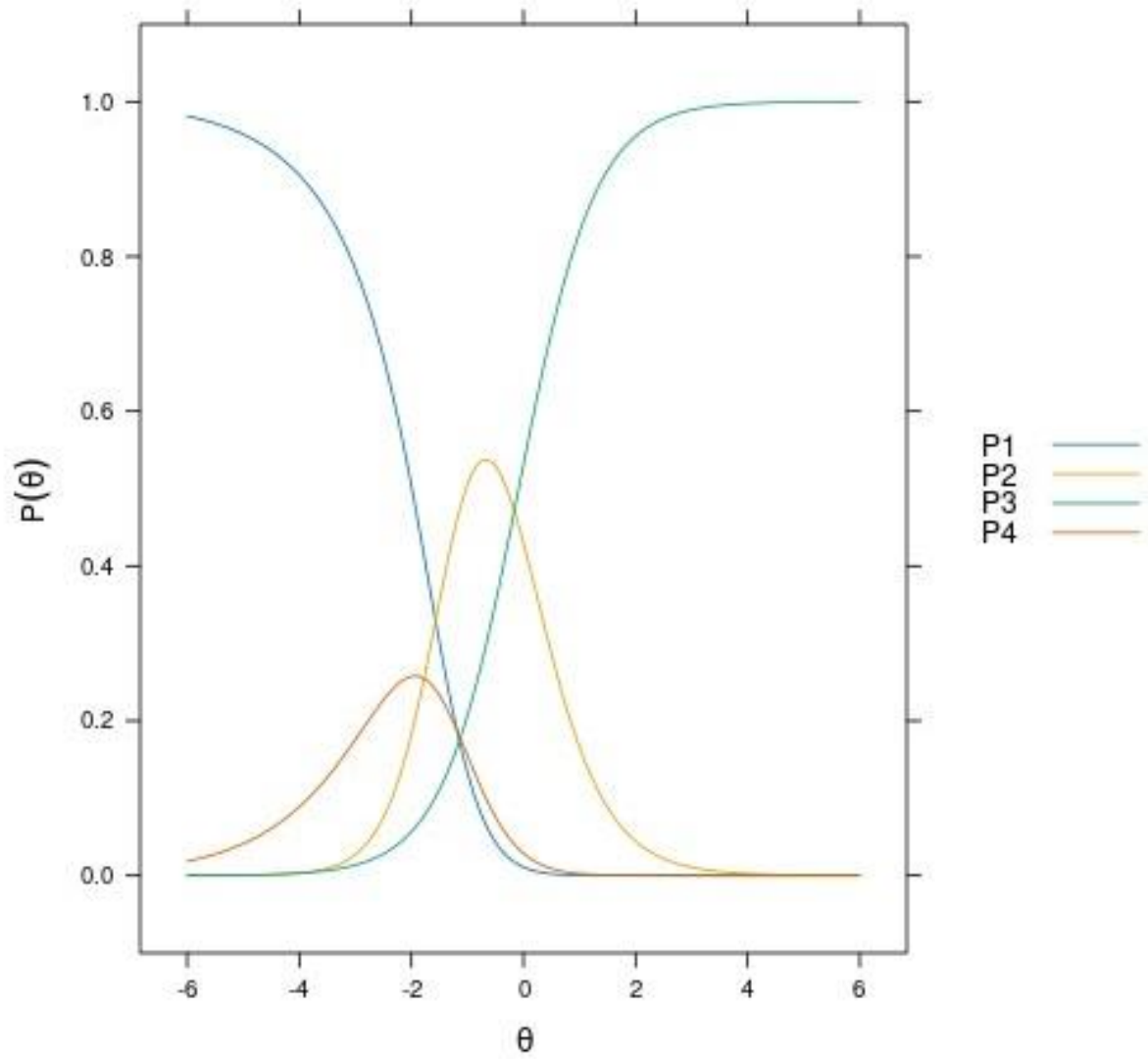
**Fonte:** Elaborado própria

**Gráfico 91** – Função de Probabilidade – Item 53

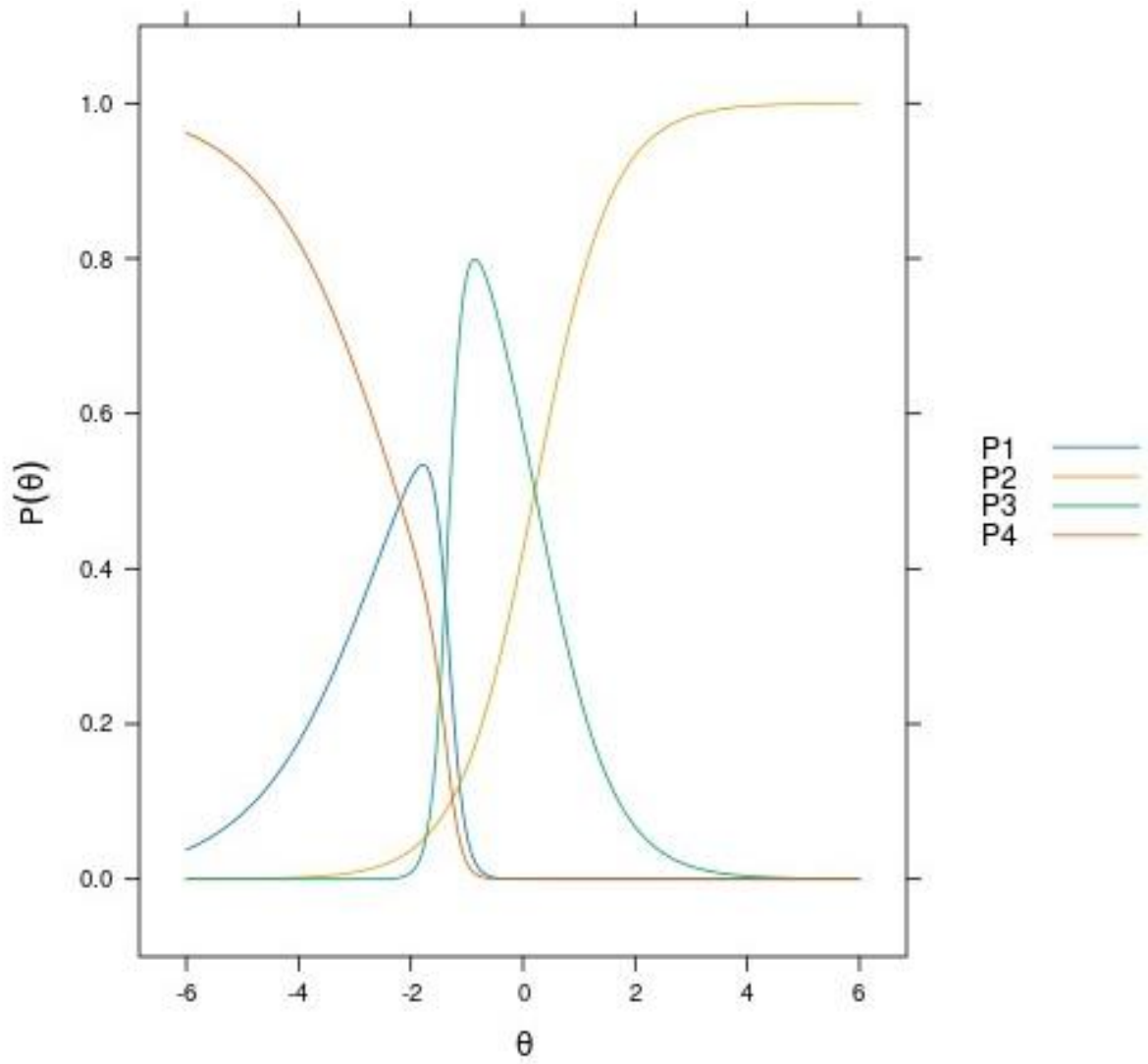
**Fonte:** Elaborado própria

**Gráfico 92** – Função de Probabilidade – Item 54

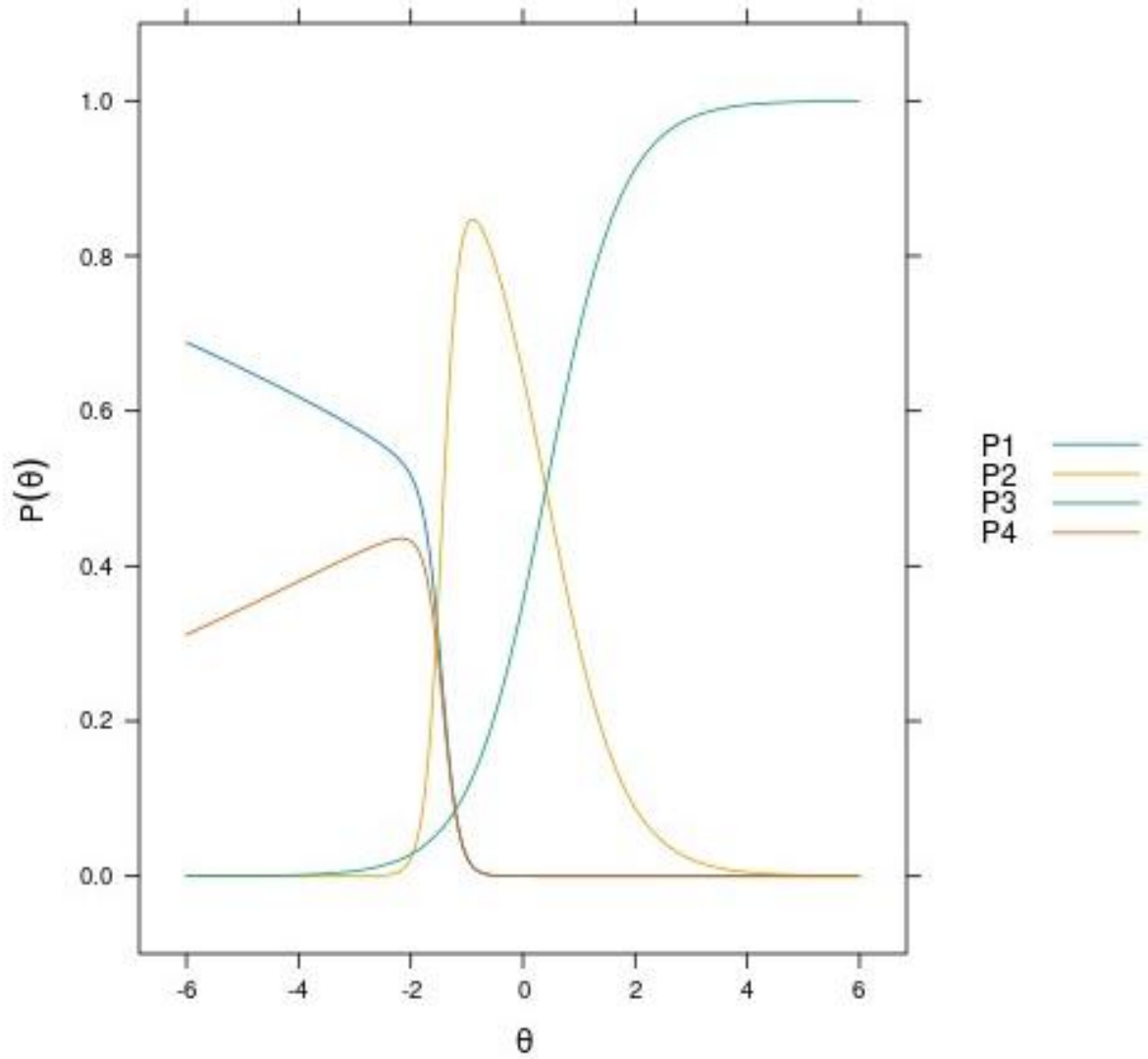
**Fonte:** Elaborado própria

**Gráfico 93** – Função de Probabilidade – Item 55

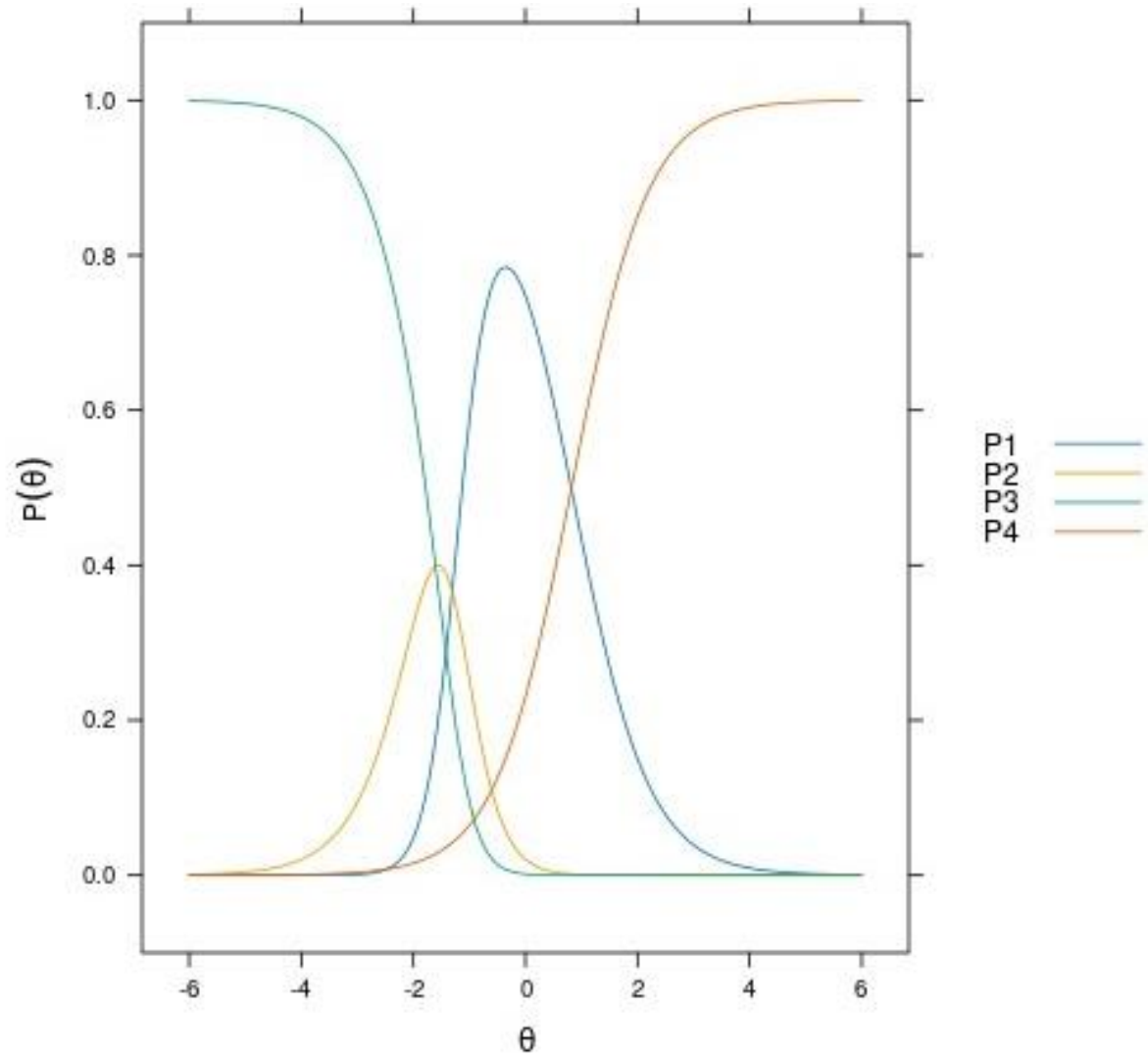
**Fonte:** Elaborado própria

**Gráfico 94** – Função de Probabilidade – Item 56

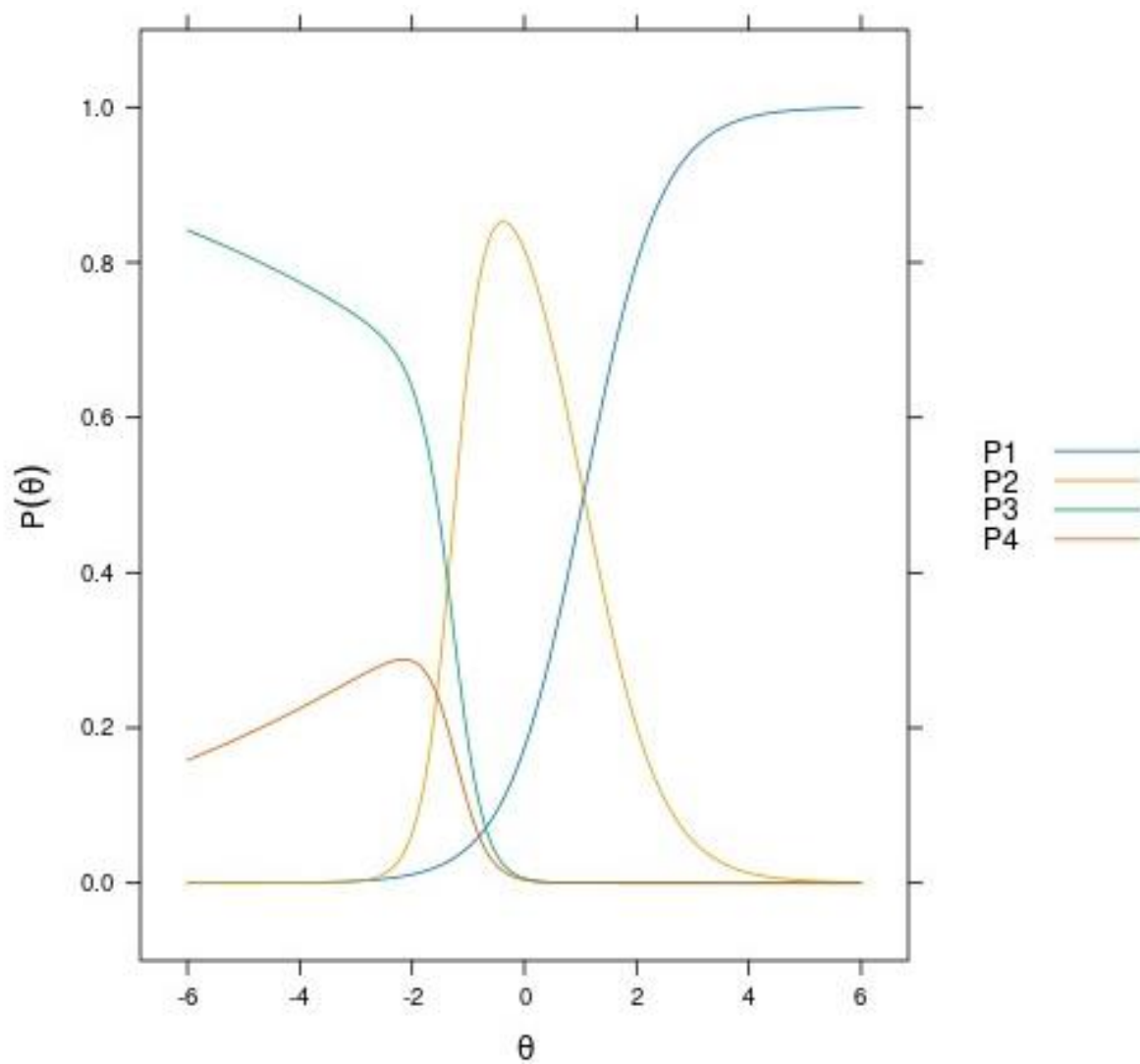
**Fonte:** Elaborado própria

**Gráfico 95** – Função de Probabilidade – Item 57

**Fonte:** Elaborado própria

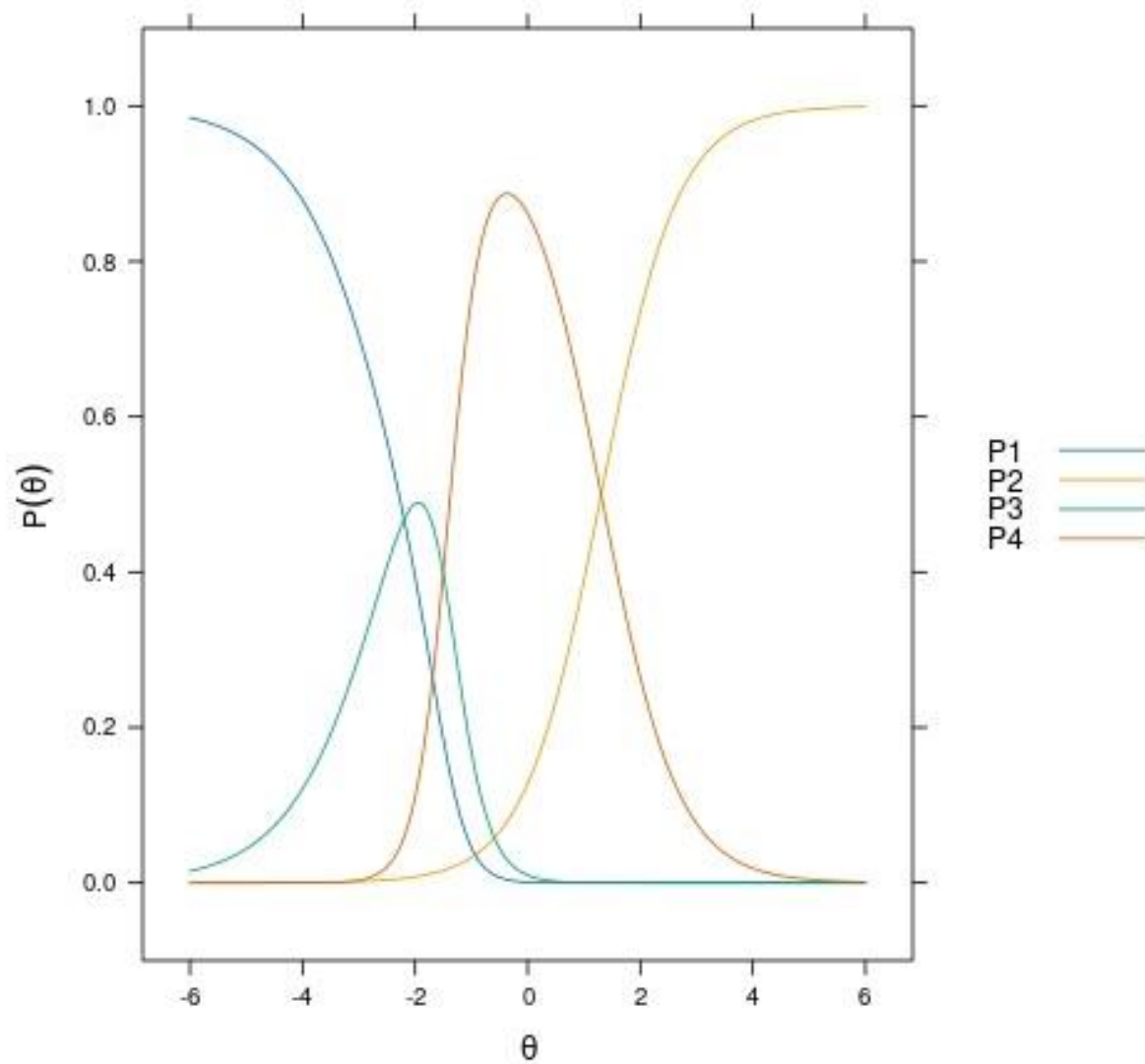
**Gráfico 96** – Função de Probabilidade – Item 58

**Fonte:** Elaborado própria

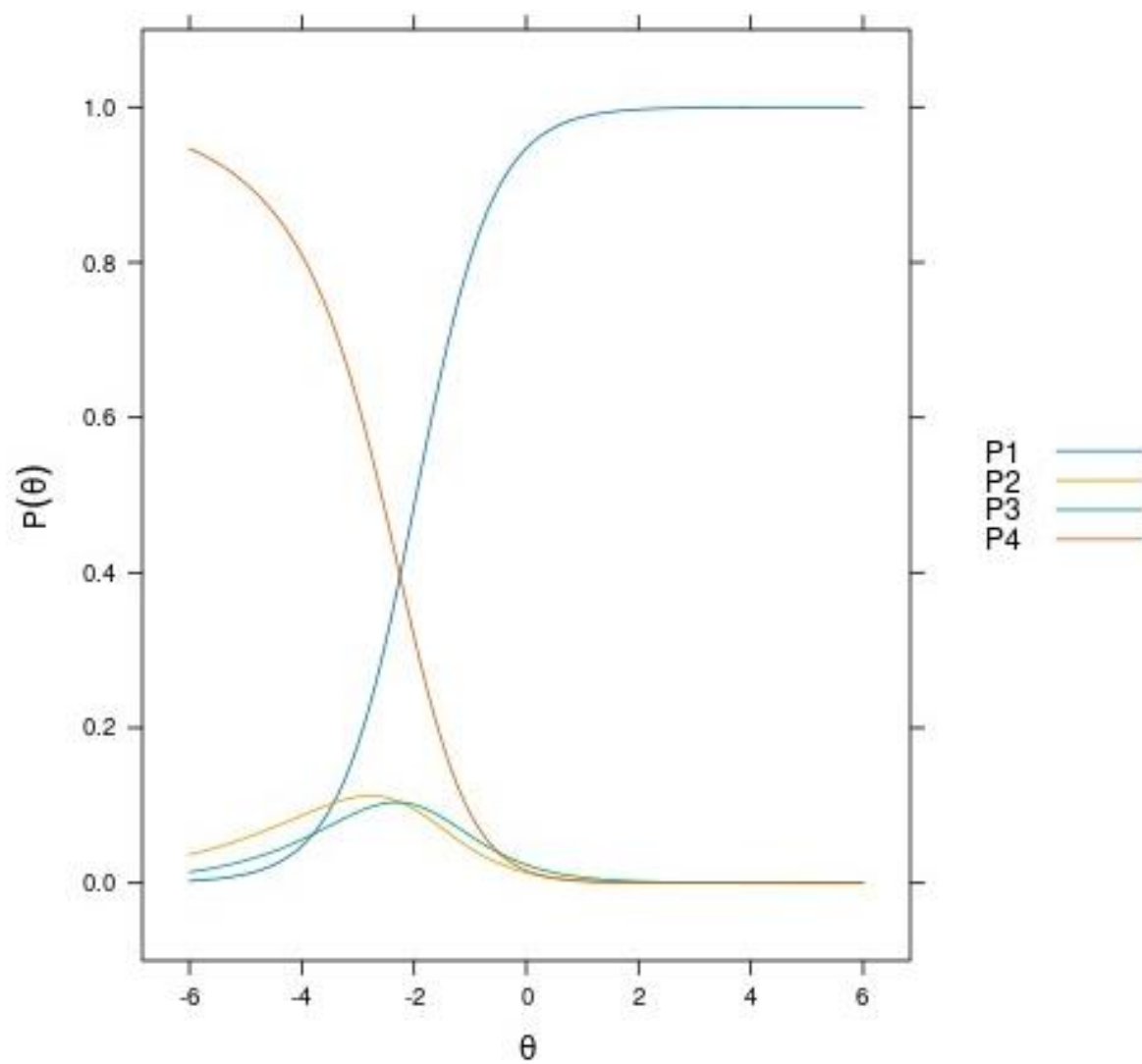
**Gráfico 97** – Função de Probabilidade – Item 59

**Fonte:** Elaborado própria

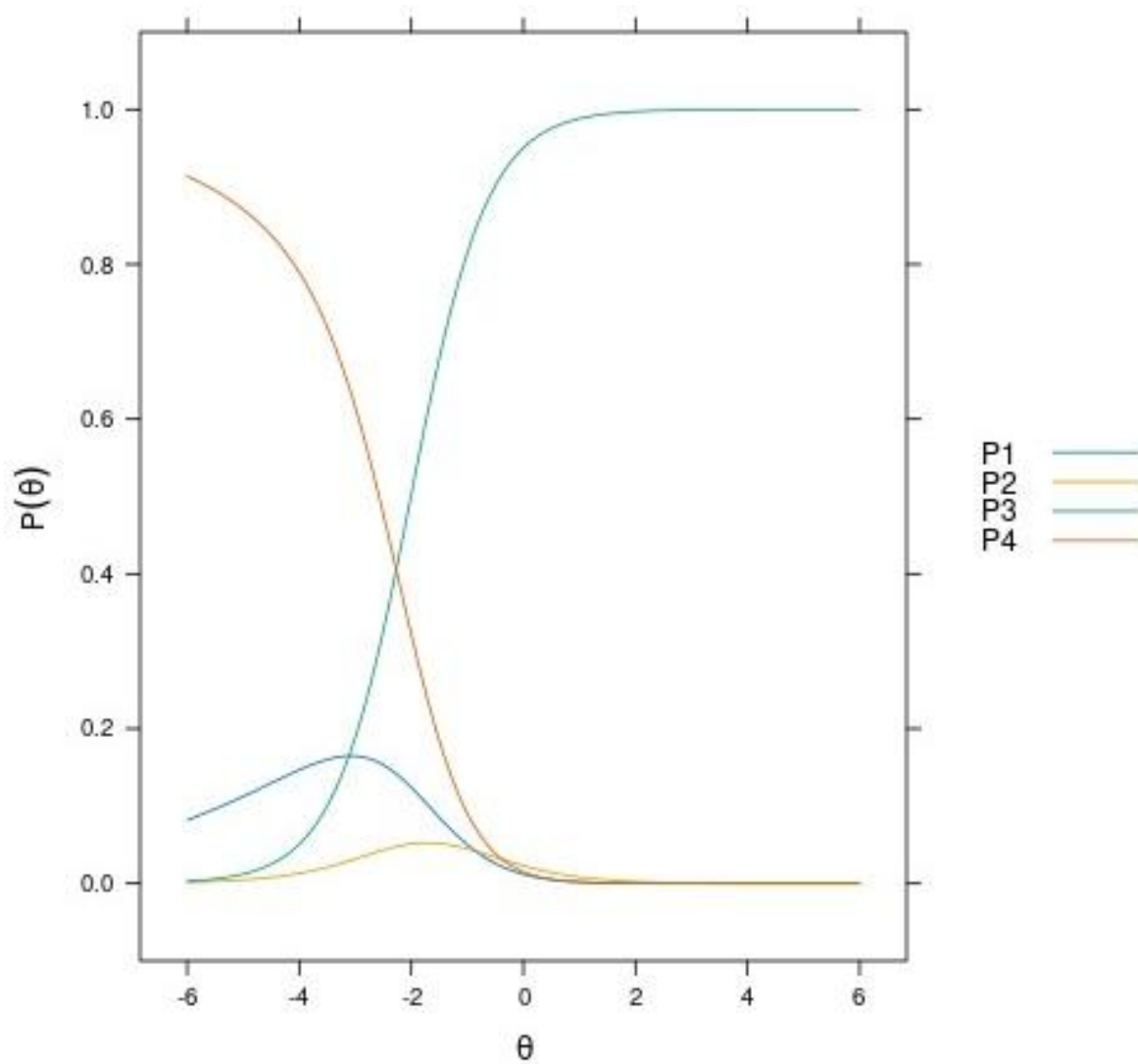


**Gráfico 98** – Função de Probabilidade – Item 60

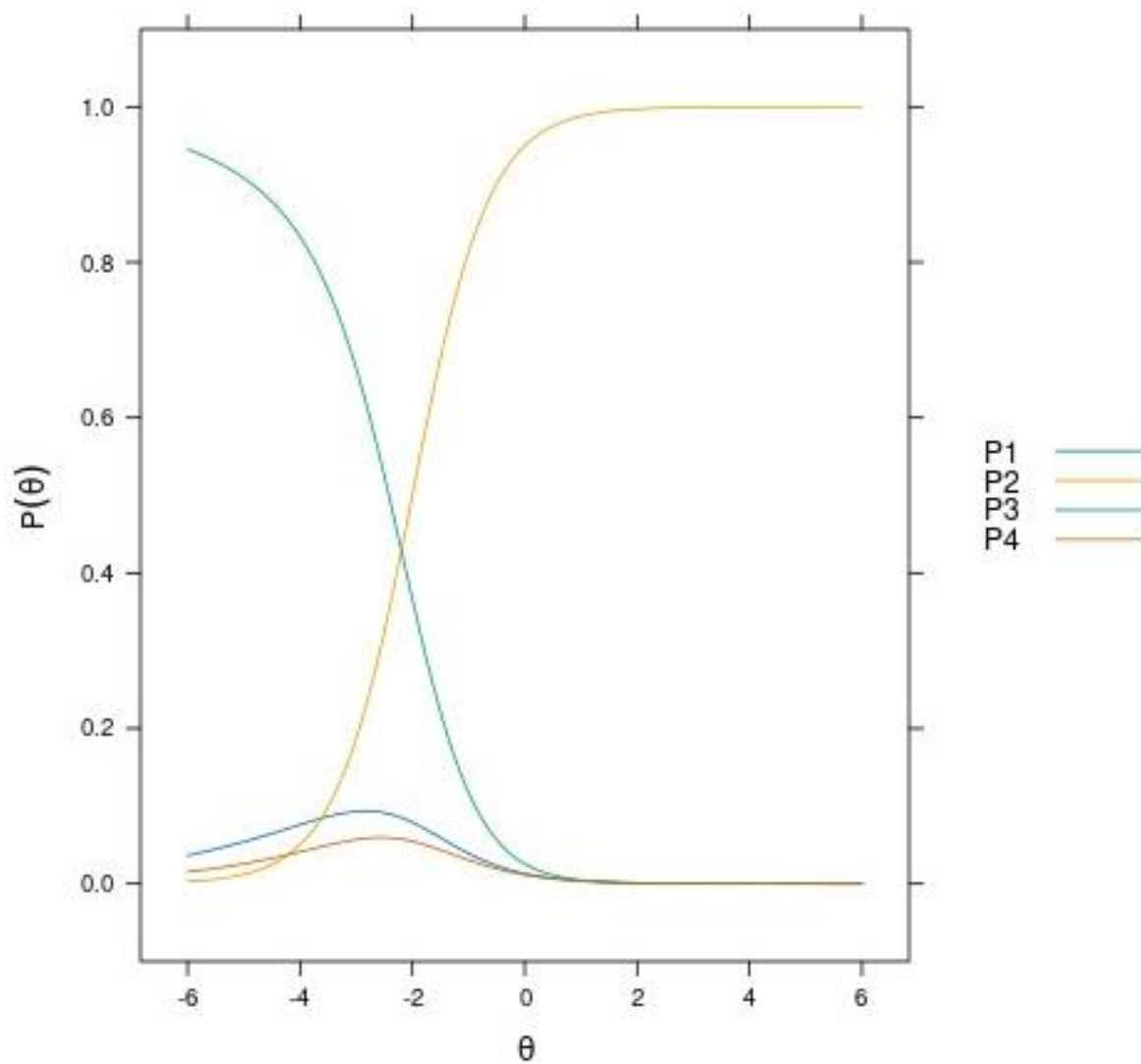
**Fonte:** Elaborado própria

**Gráfico 99** – Função de Probabilidade – Item 61

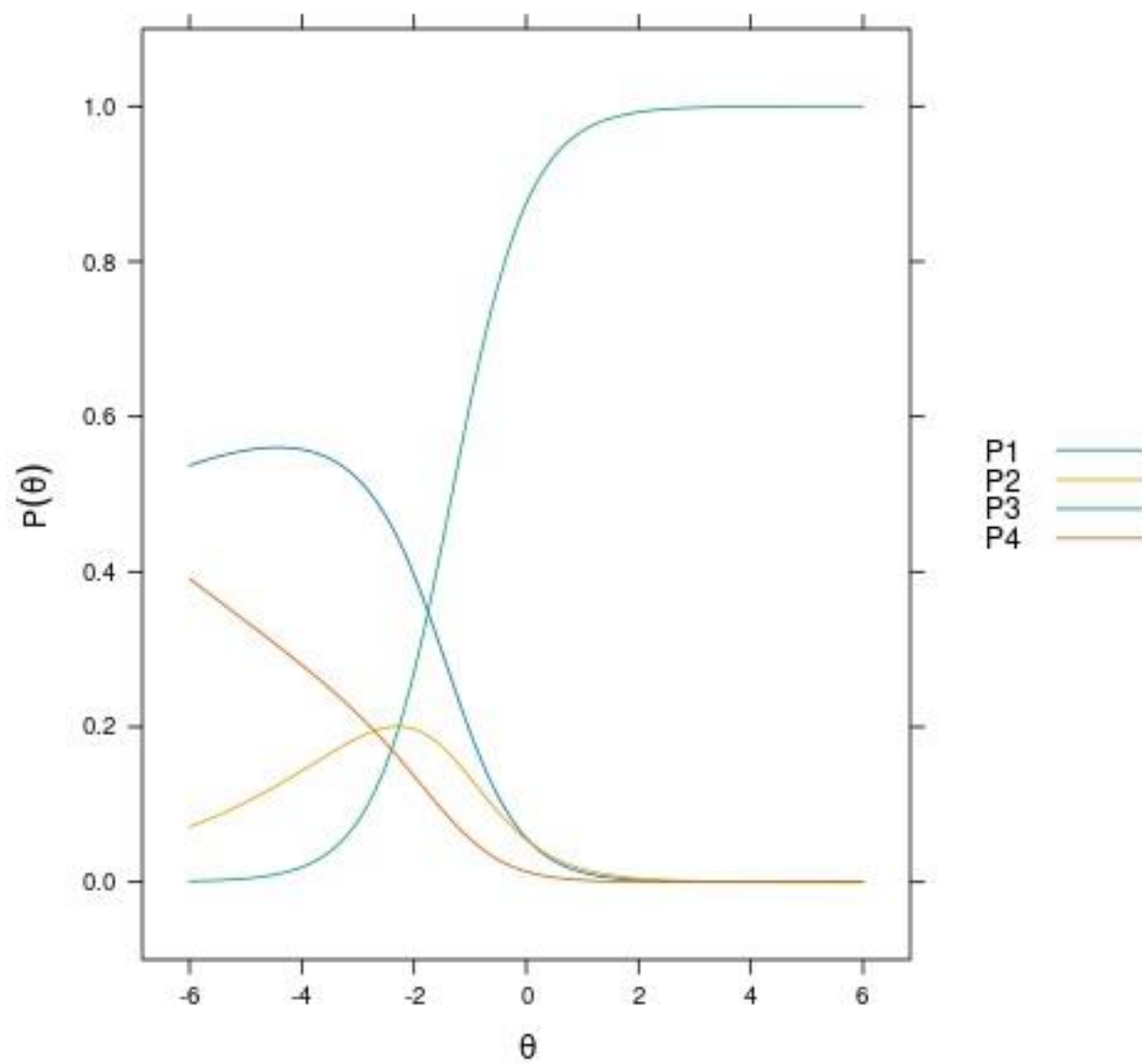
**Fonte:** Elaborado própria

**Gráfico 100** – Função de Probabilidade – Item 62

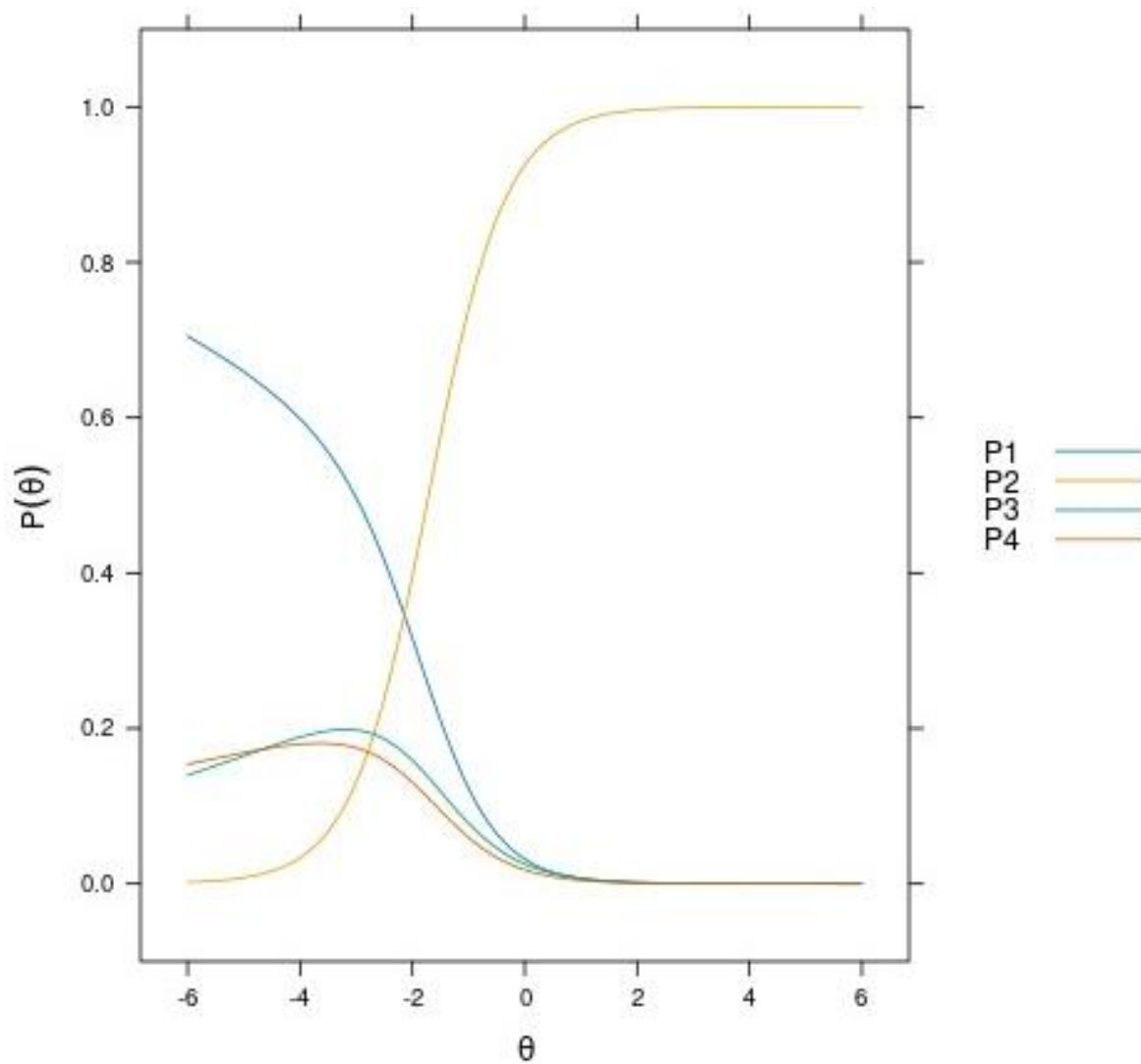
Fonte: Elaborado própria

**Gráfico 101** – Função de Probabilidade – Item 63

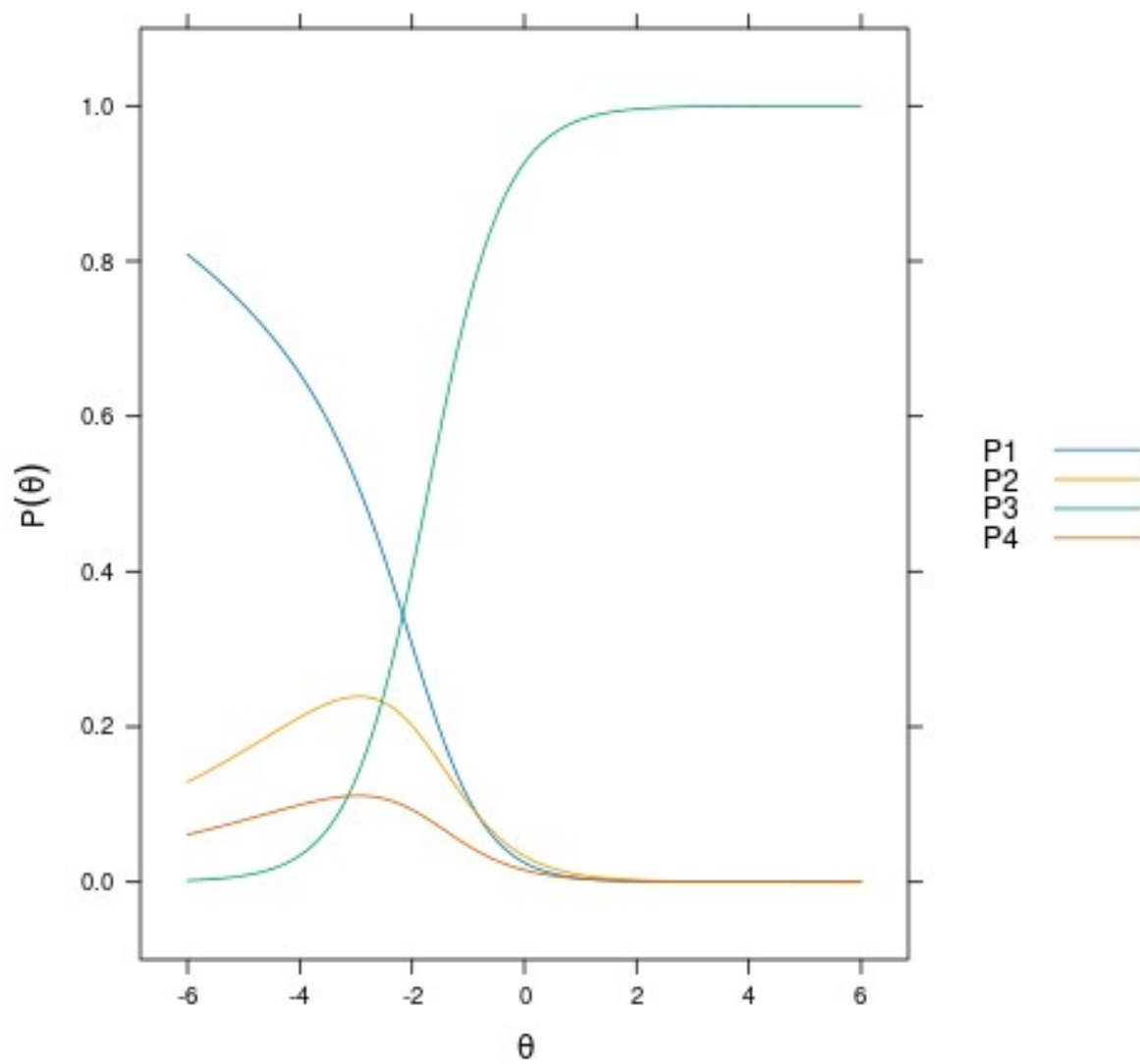
**Fonte:** Elaborado própria

**Gráfico 102** – Função de Probabilidade – Item 64

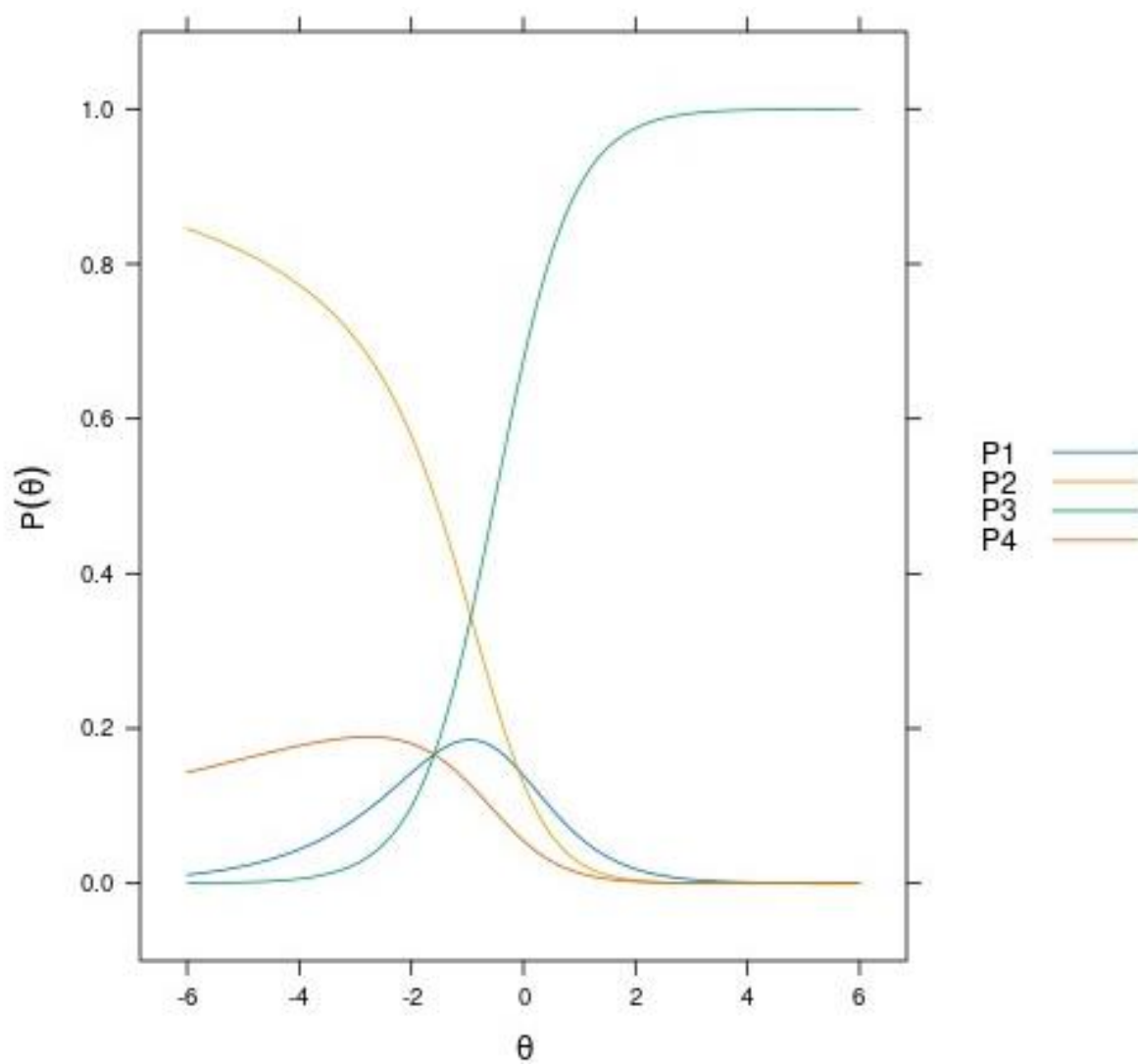
**Fonte:** Elaborado própria

**Gráfico 103** – Função de Probabilidade – Item 65

**Fonte:** Elaborado própria

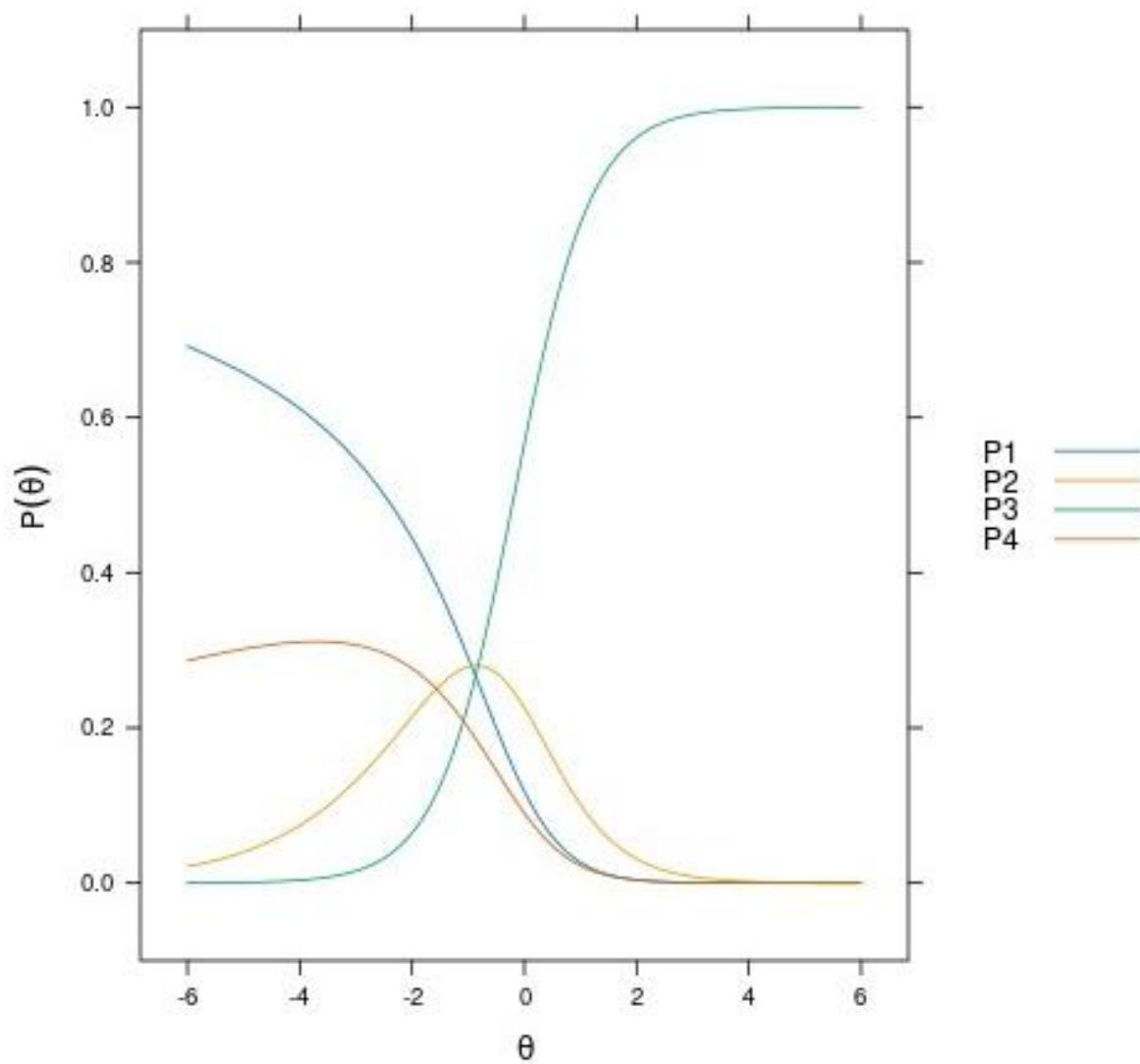
**Gráfico 104** – Função de Probabilidade – Item 66

**Fonte:** Elaborado própria

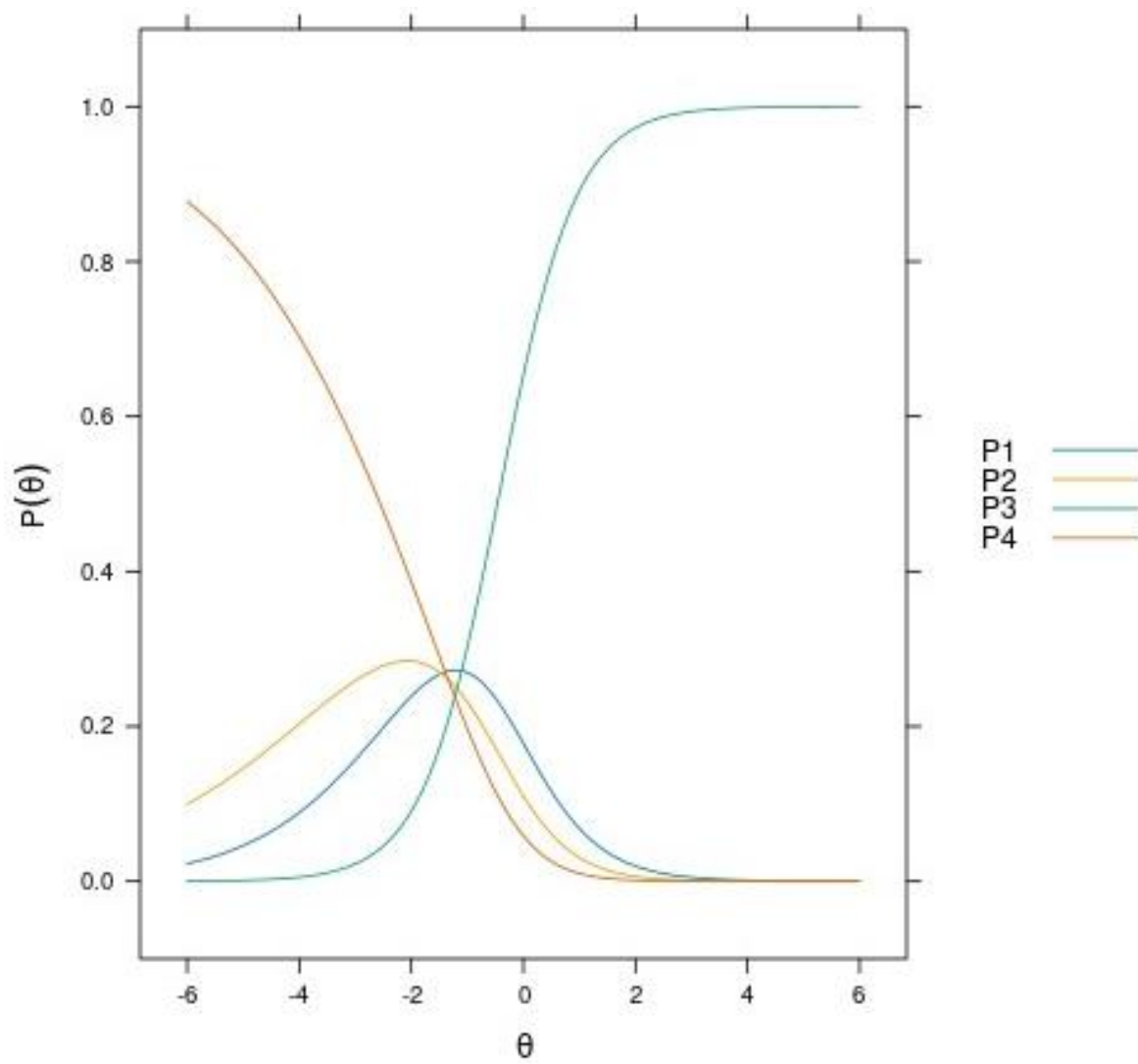
**Gráfico 105** – Função de Probabilidade – Item 67

**Fonte:** Elaborado própria

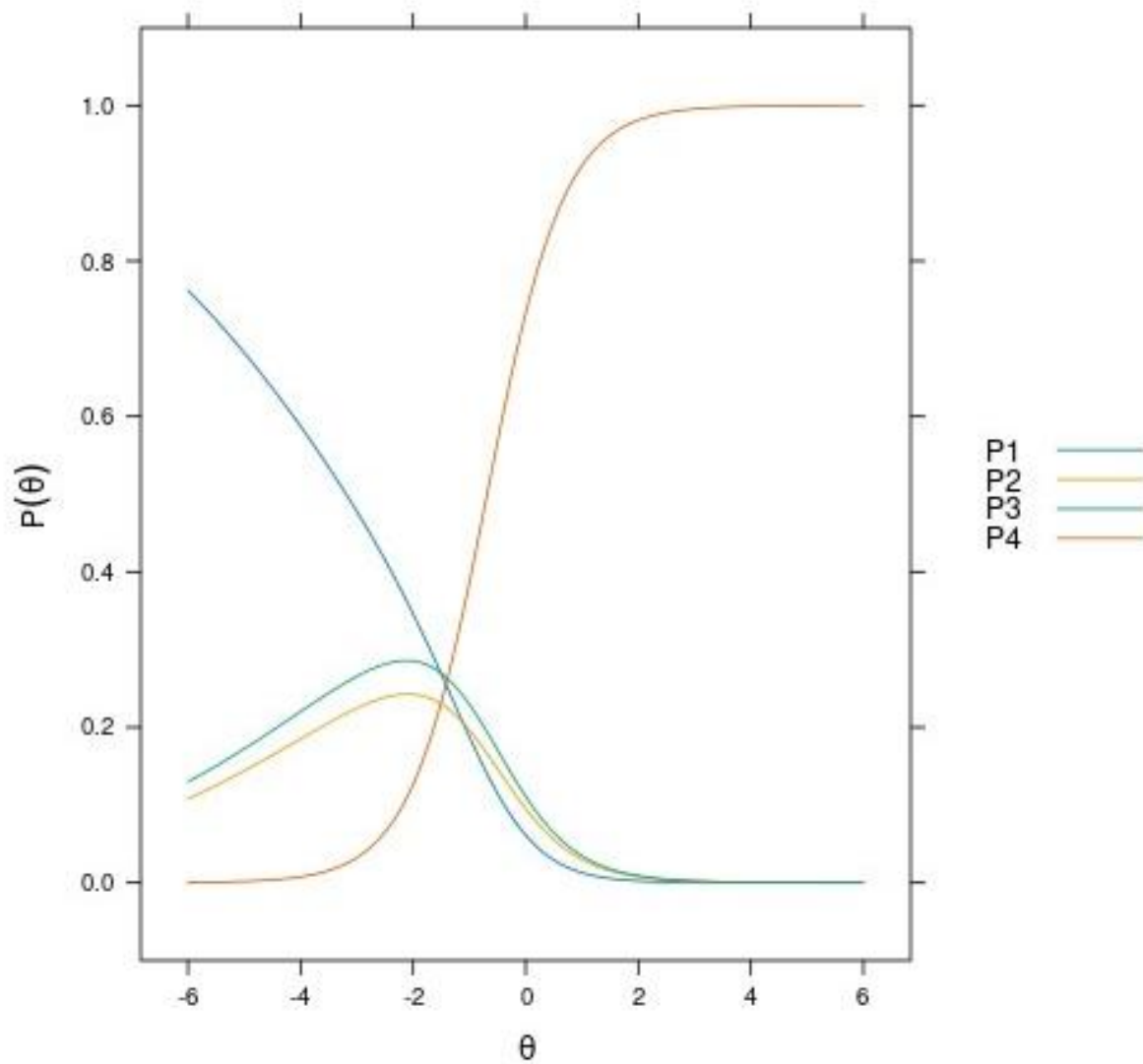


**Gráfico 106** – Função de Probabilidade – Item 68

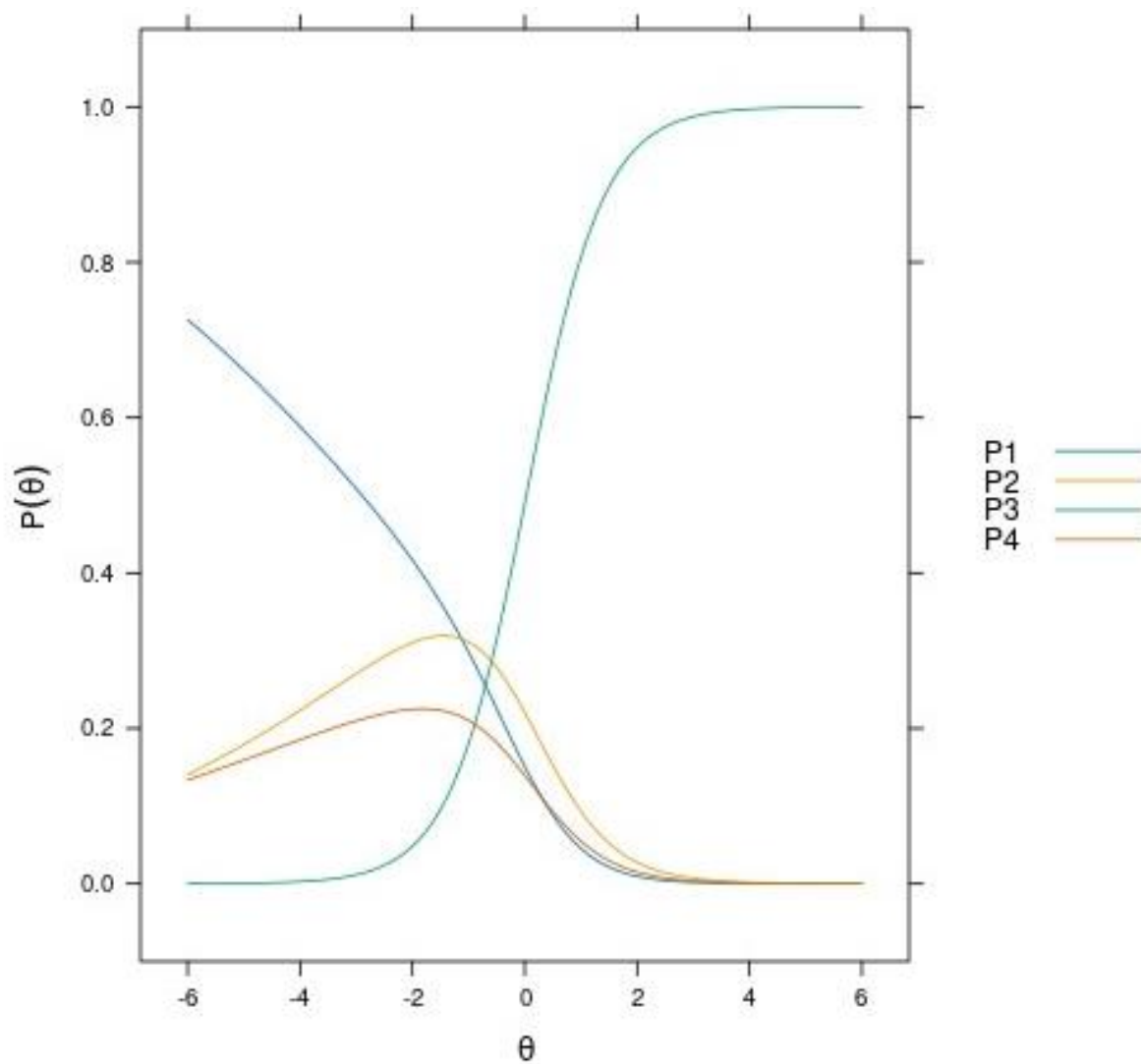
**Fonte:** Elaborado própria

**Gráfico 107** – Função de Probabilidade – Item 69

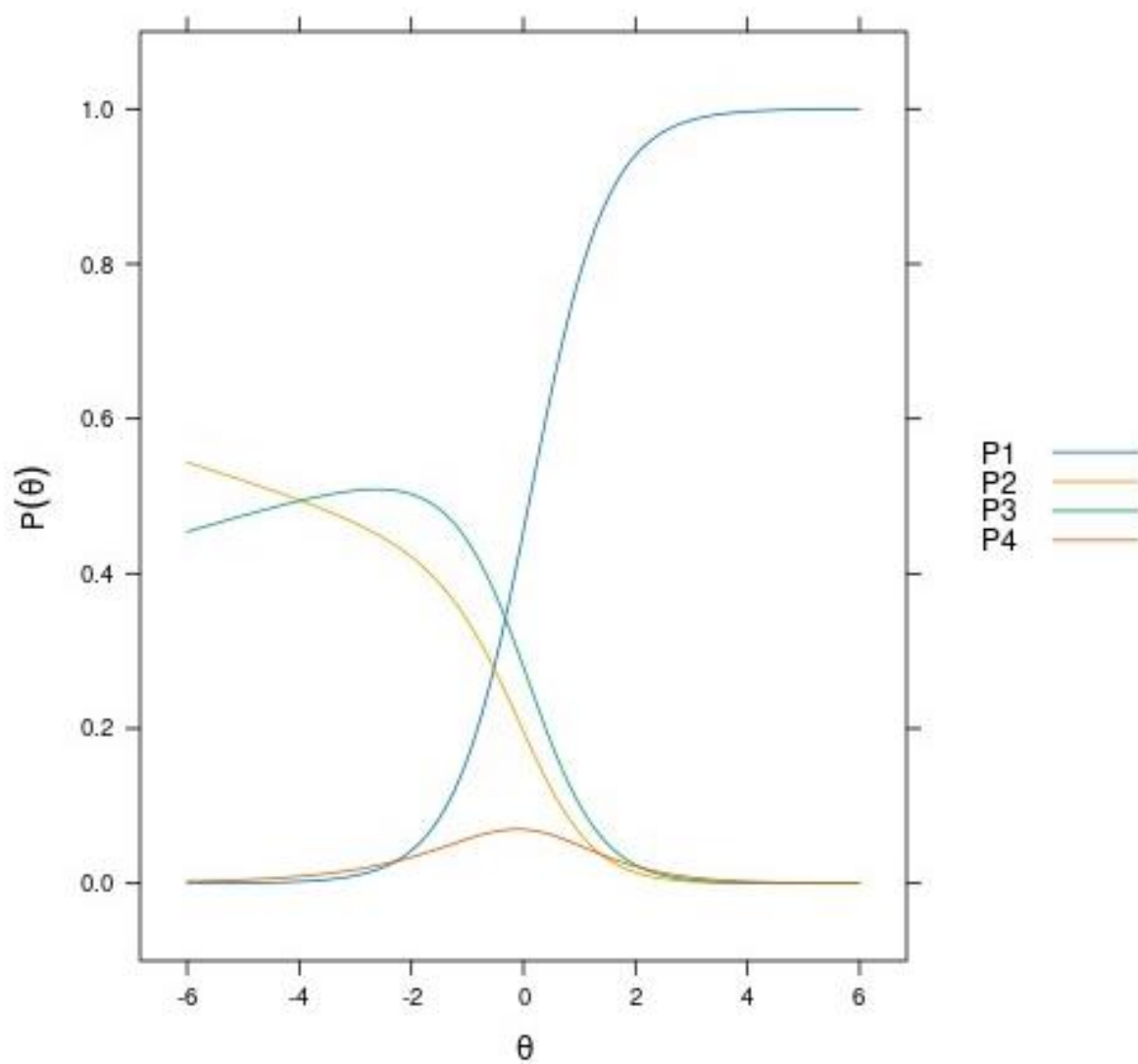
**Fonte:** Elaborado própria

**Gráfico 108** – Função de Probabilidade – Item 70

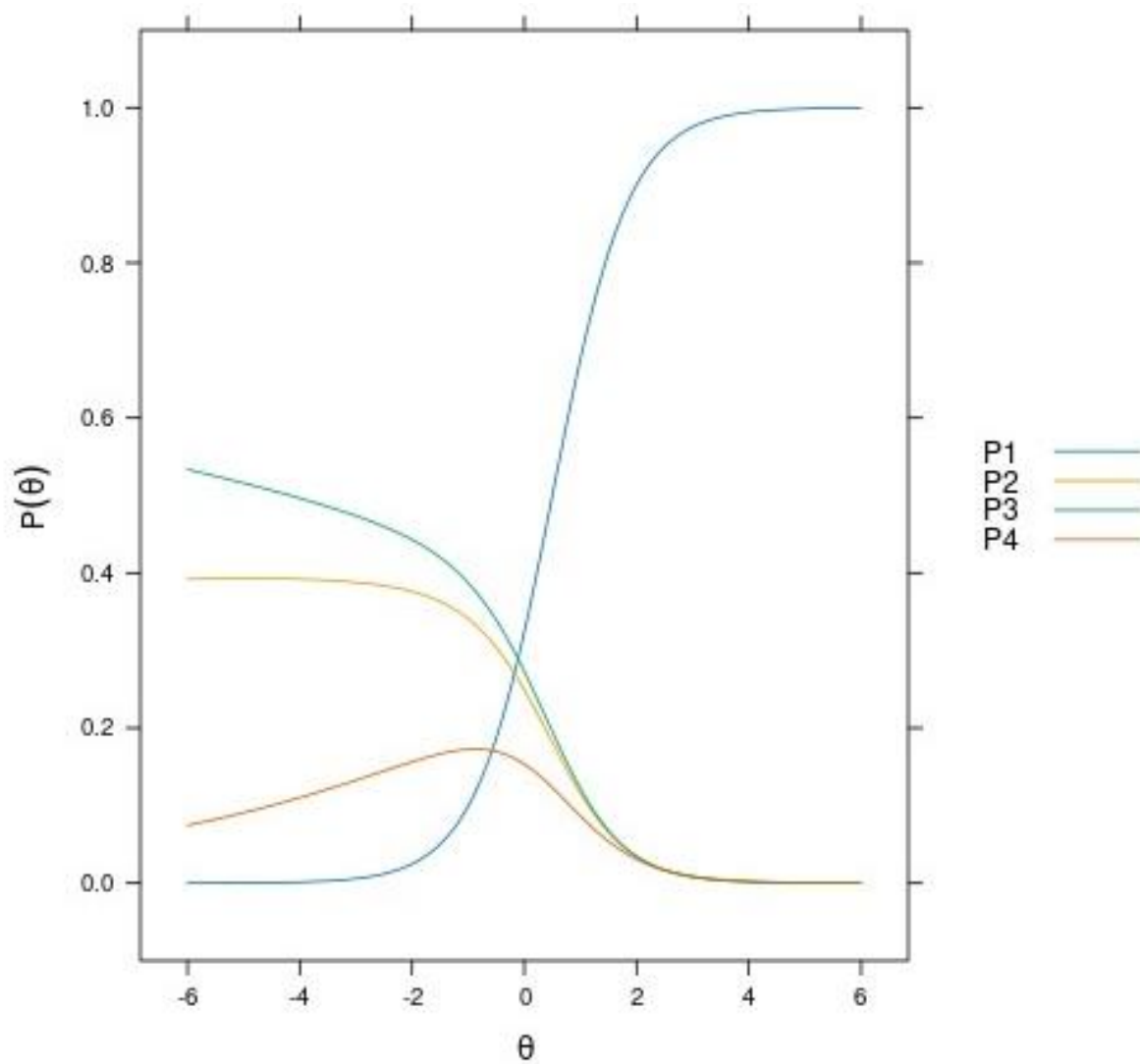
**Fonte:** Elaborado própria

**Gráfico 109** – Função de Probabilidade – Item 71

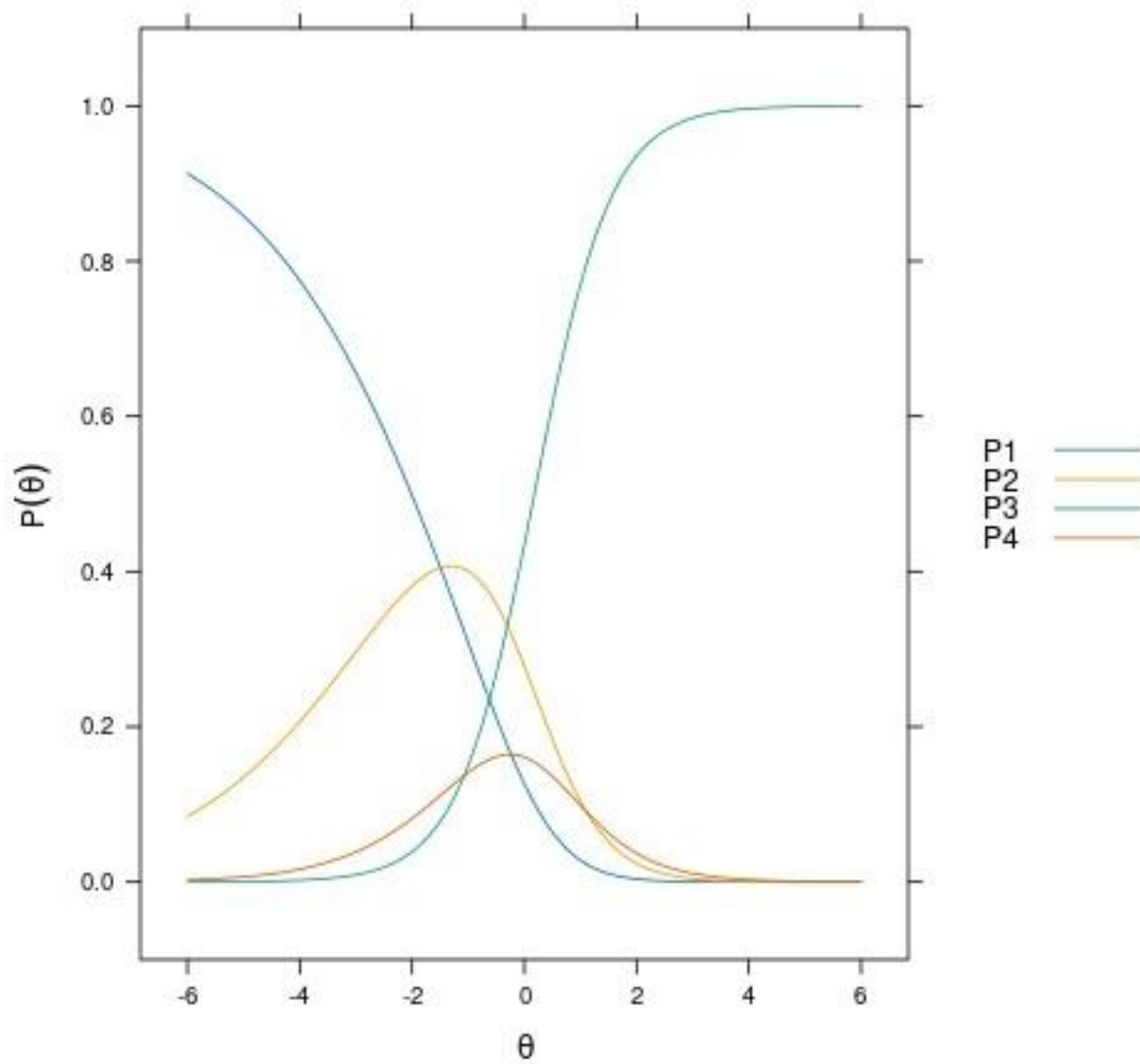
**Fonte:** Elaborado própria

**Gráfico 110** – Função de Probabilidade – Item 72

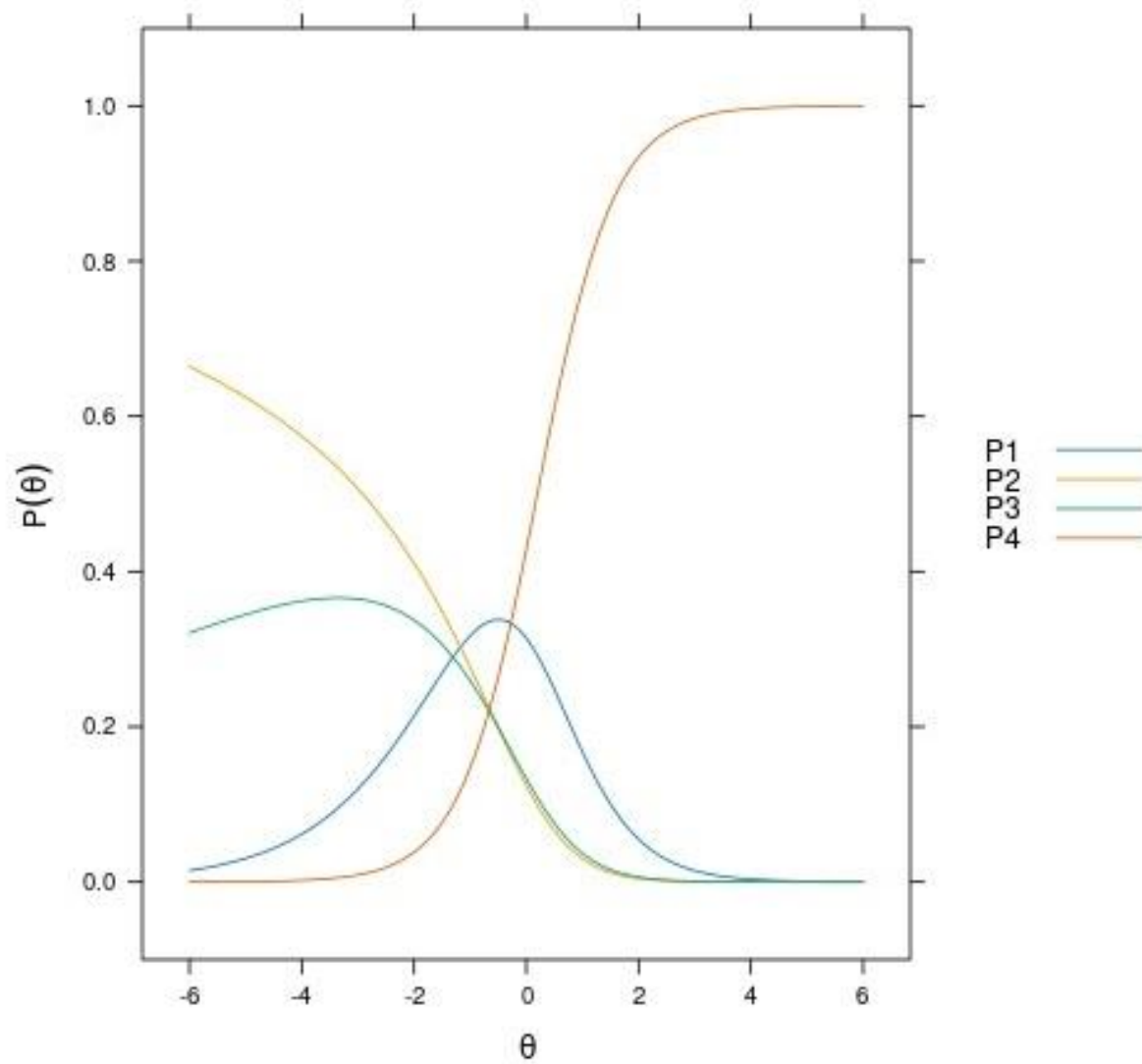
**Fonte:** Elaborado própria

**Gráfico 111** – Função de Probabilidade – Item 73

**Fonte:** Elaborado própria

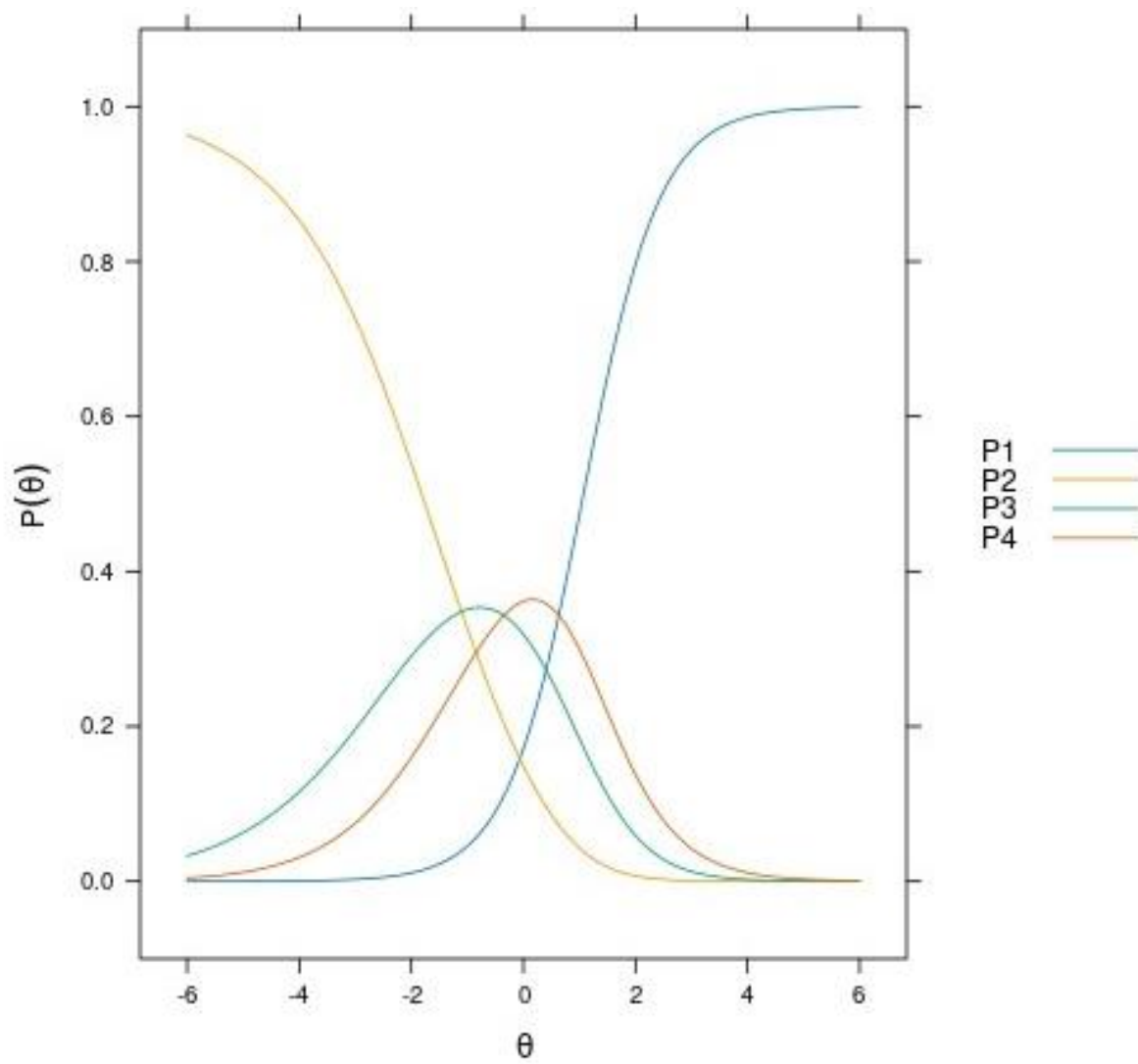
**Gráfico 112** – Função de Probabilidade – Item 74

**Fonte:** Elaborado própria

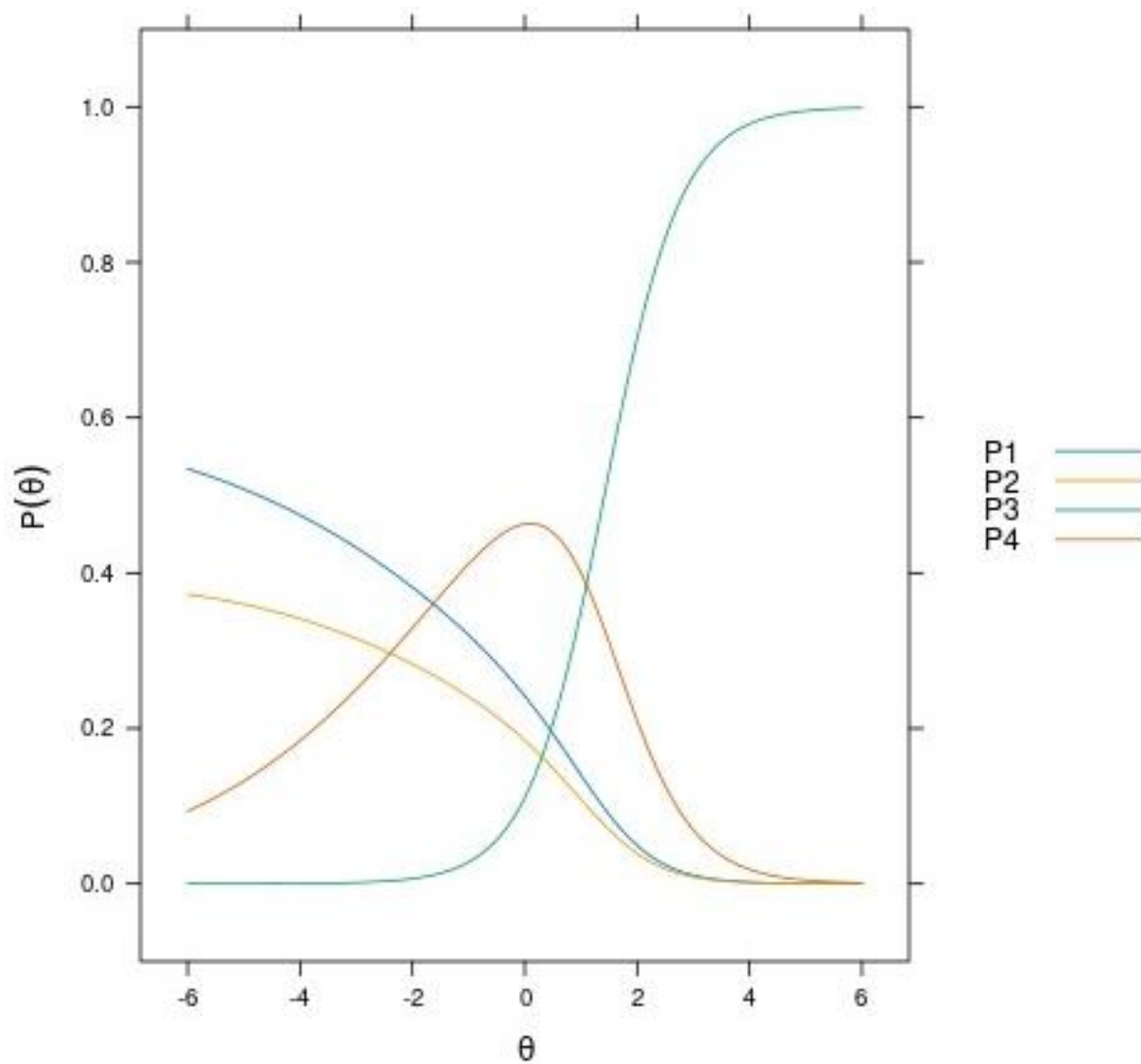
**Gráfico 113** – Função de Probabilidade – Item 75

**Fonte:** Elaborado própria

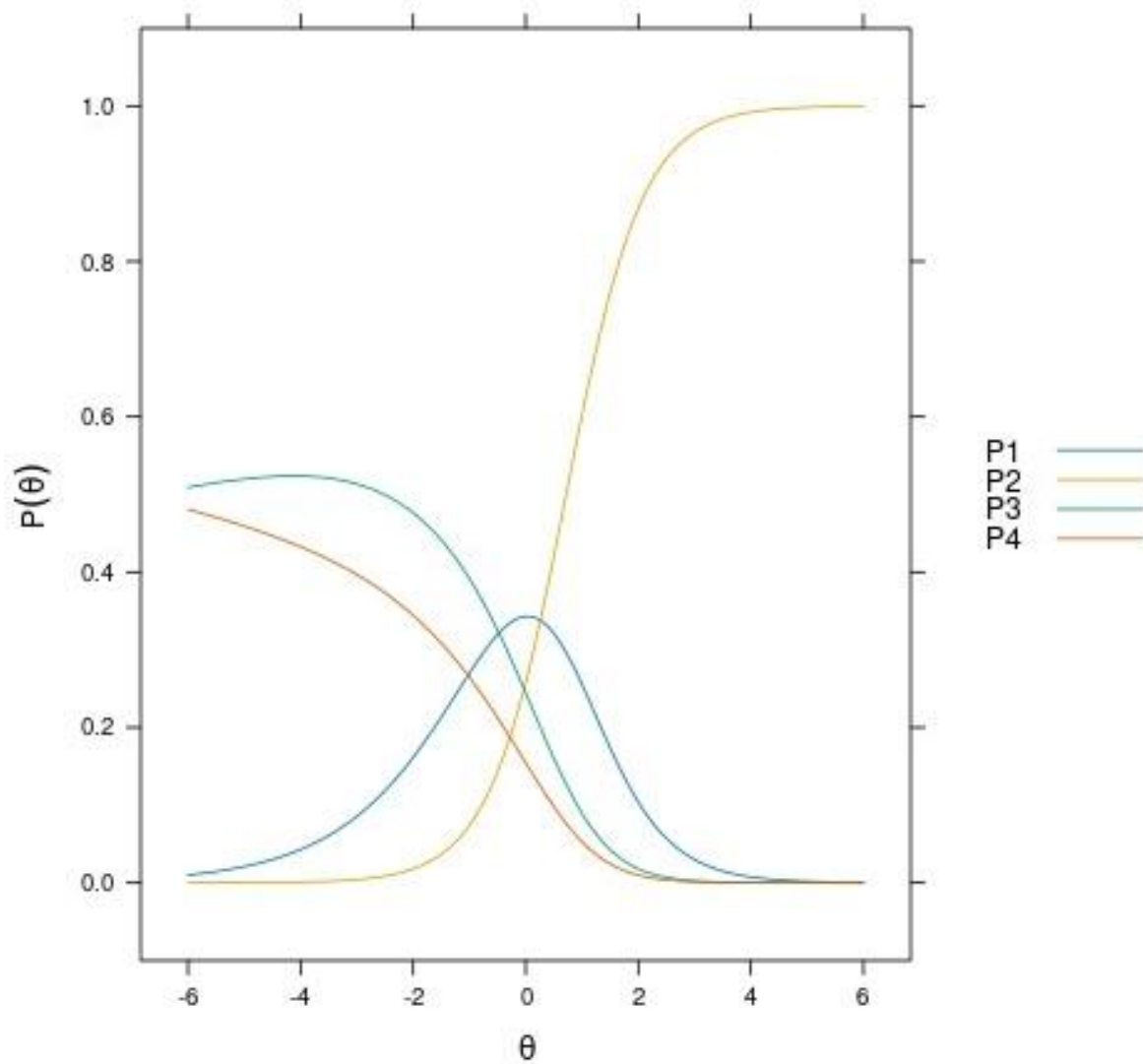


**Gráfico 114** – Função de Probabilidade – Item 76

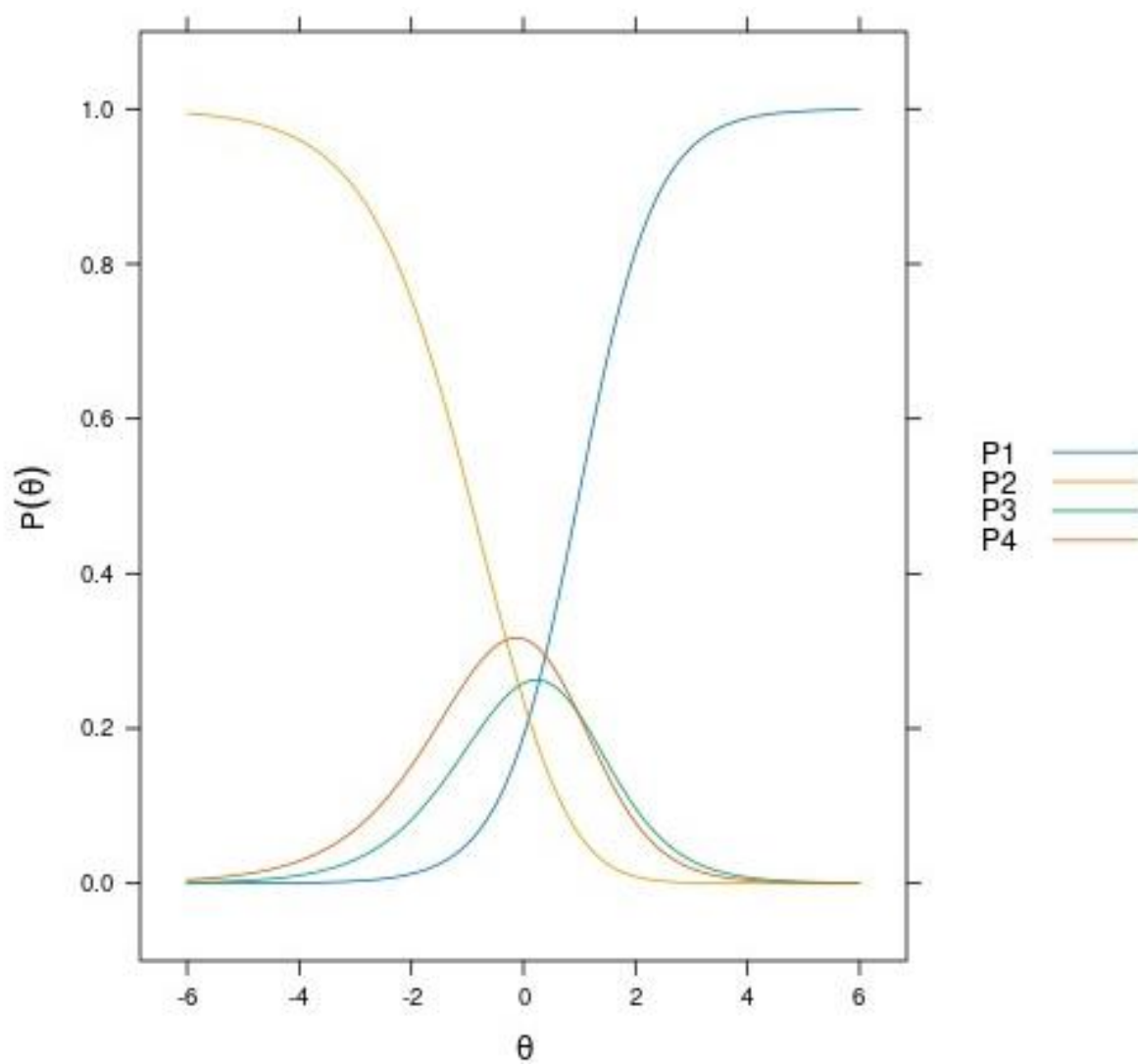
**Fonte:** Elaborado própria

**Gráfico 115** – Função de Probabilidade – Item 77

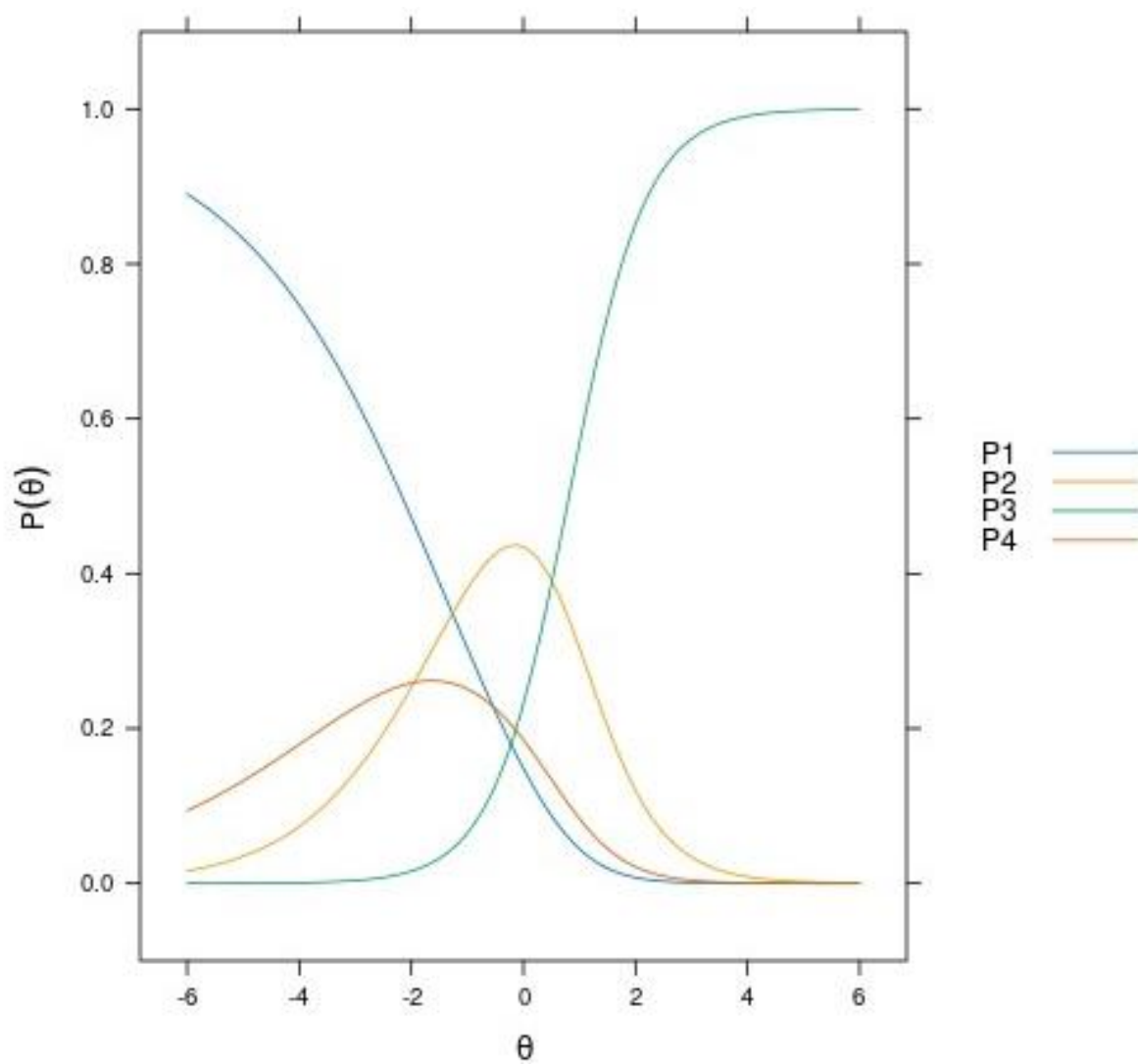
**Fonte:** Elaborado própria

**Gráfico 116** – Função de Probabilidade – Item 78

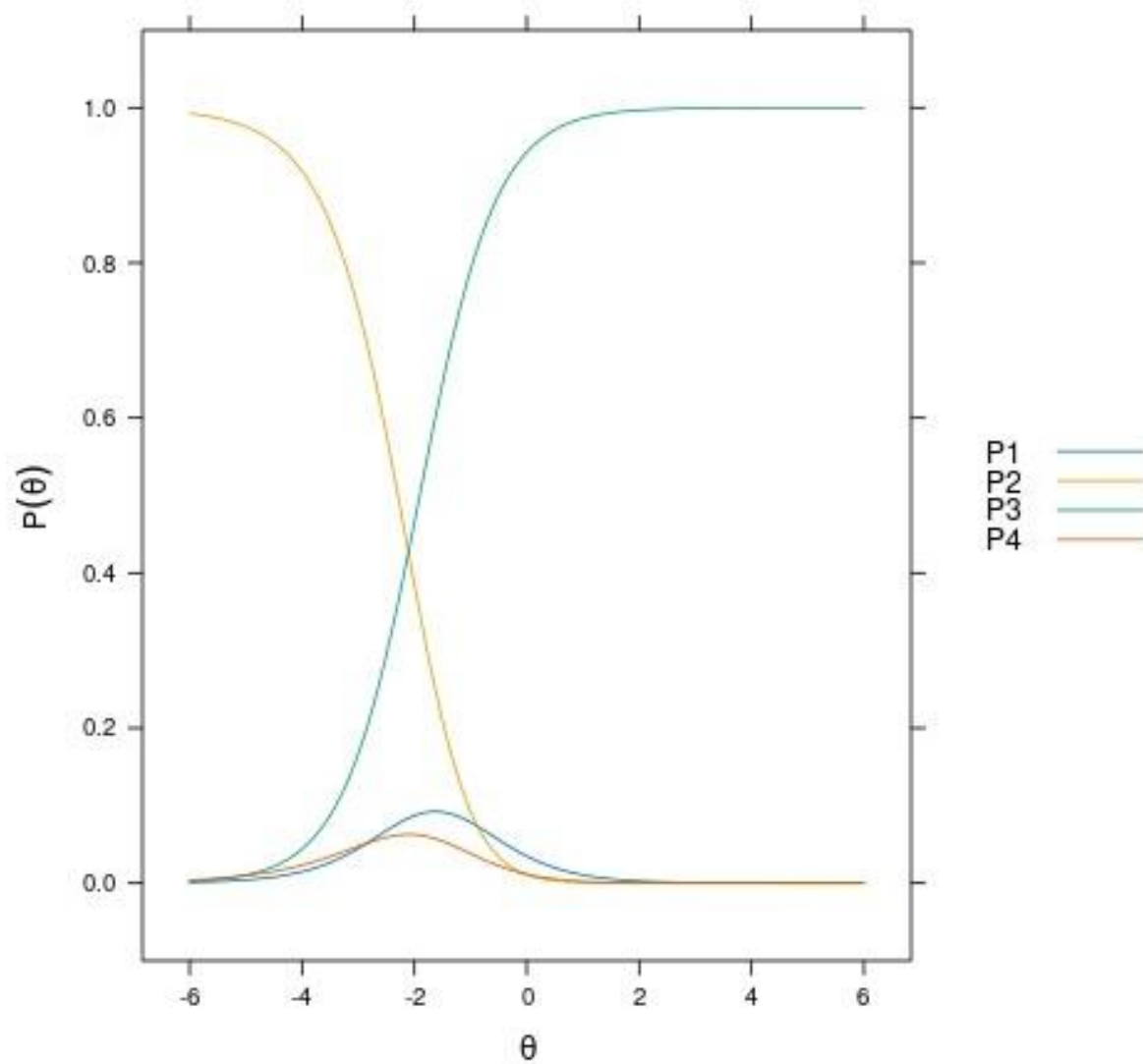
**Fonte:** Elaborado própria

**Gráfico 117** – Função de Probabilidade – Item 79

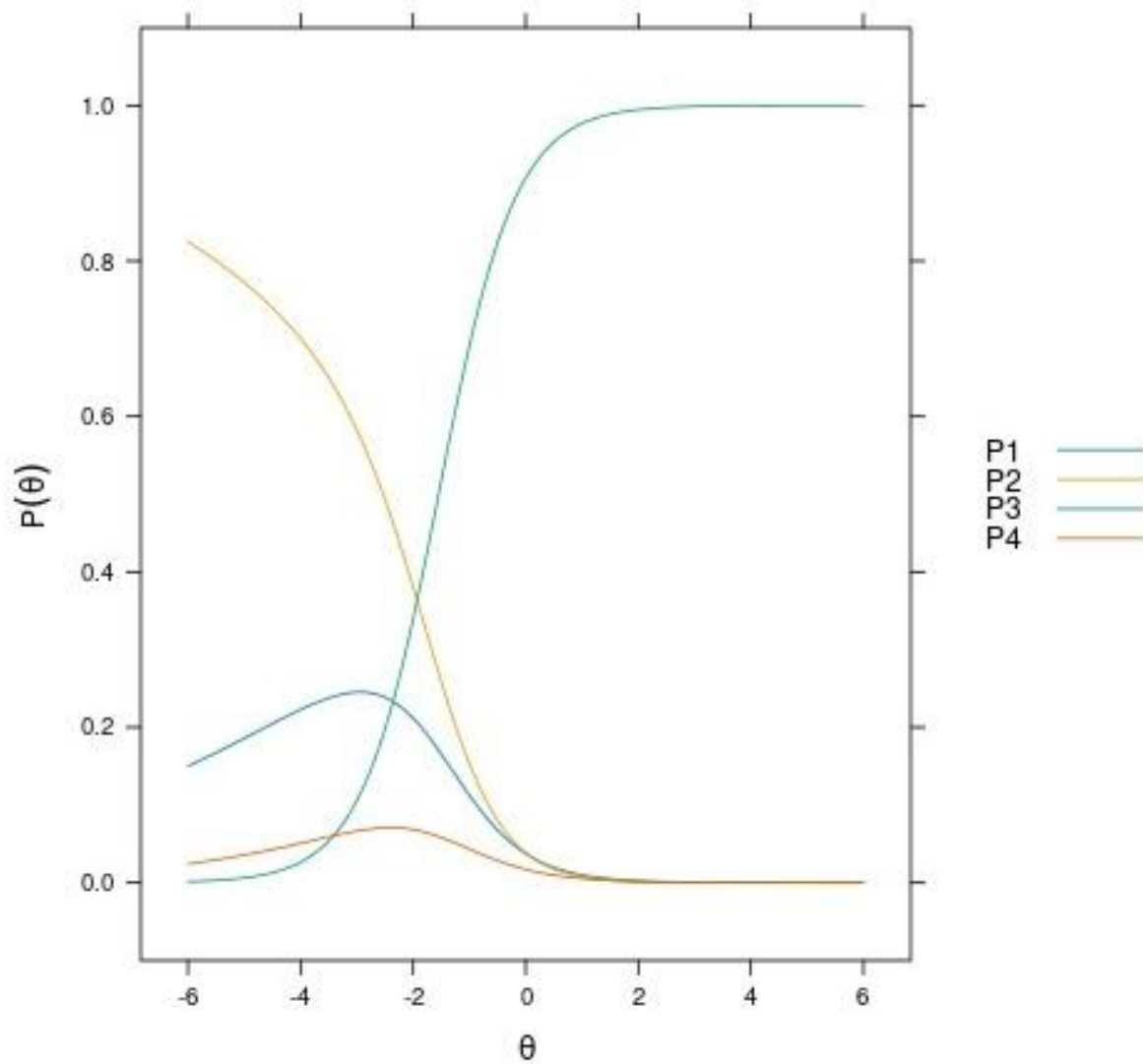
**Fonte:** Elaborado própria

**Gráfico 118** – Função de Probabilidade – Item 80

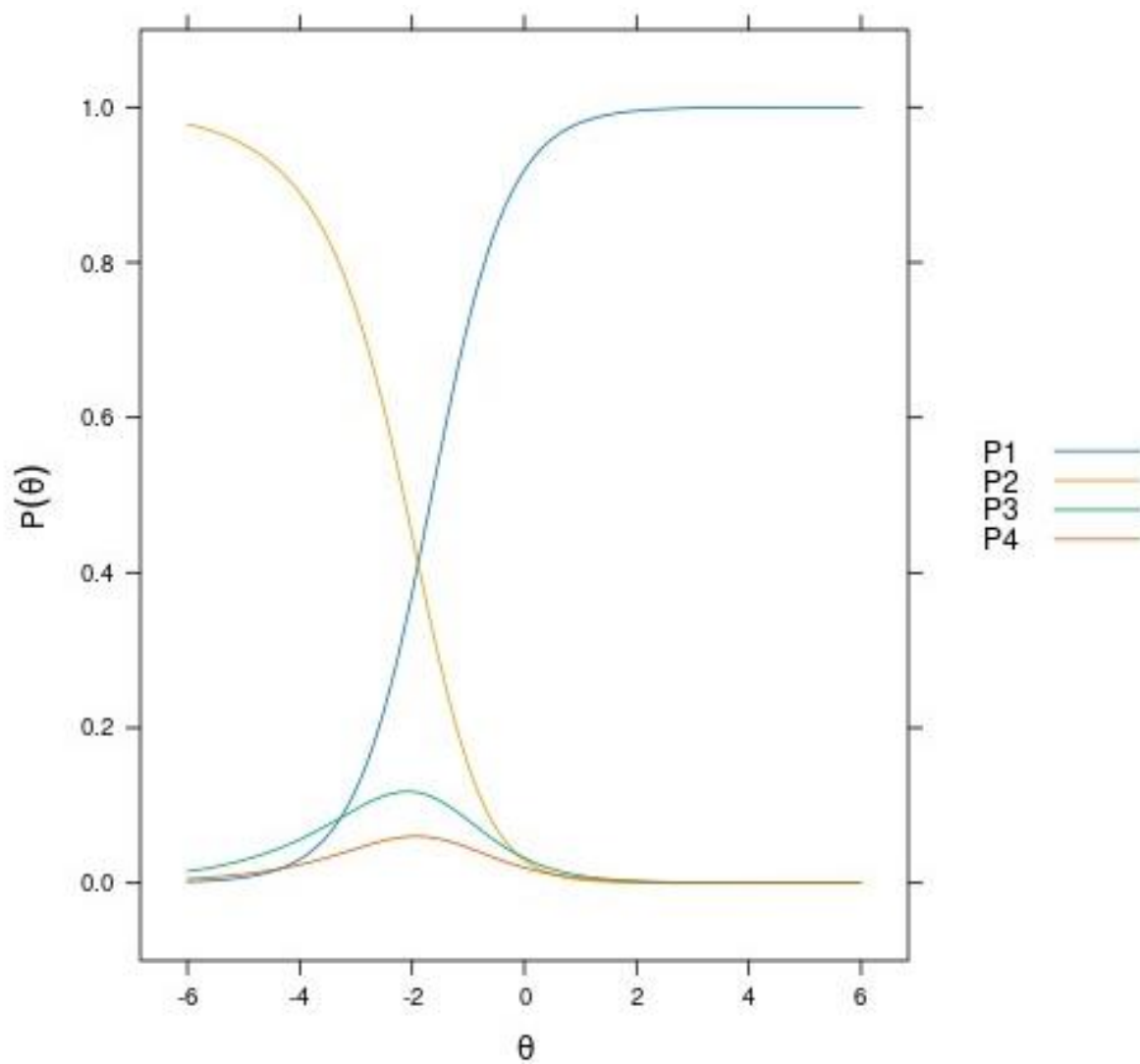
**Fonte:** Elaborado própria

**Gráfico 119** – Função de Probabilidade – Item 81

**Fonte:** Elaborado própria

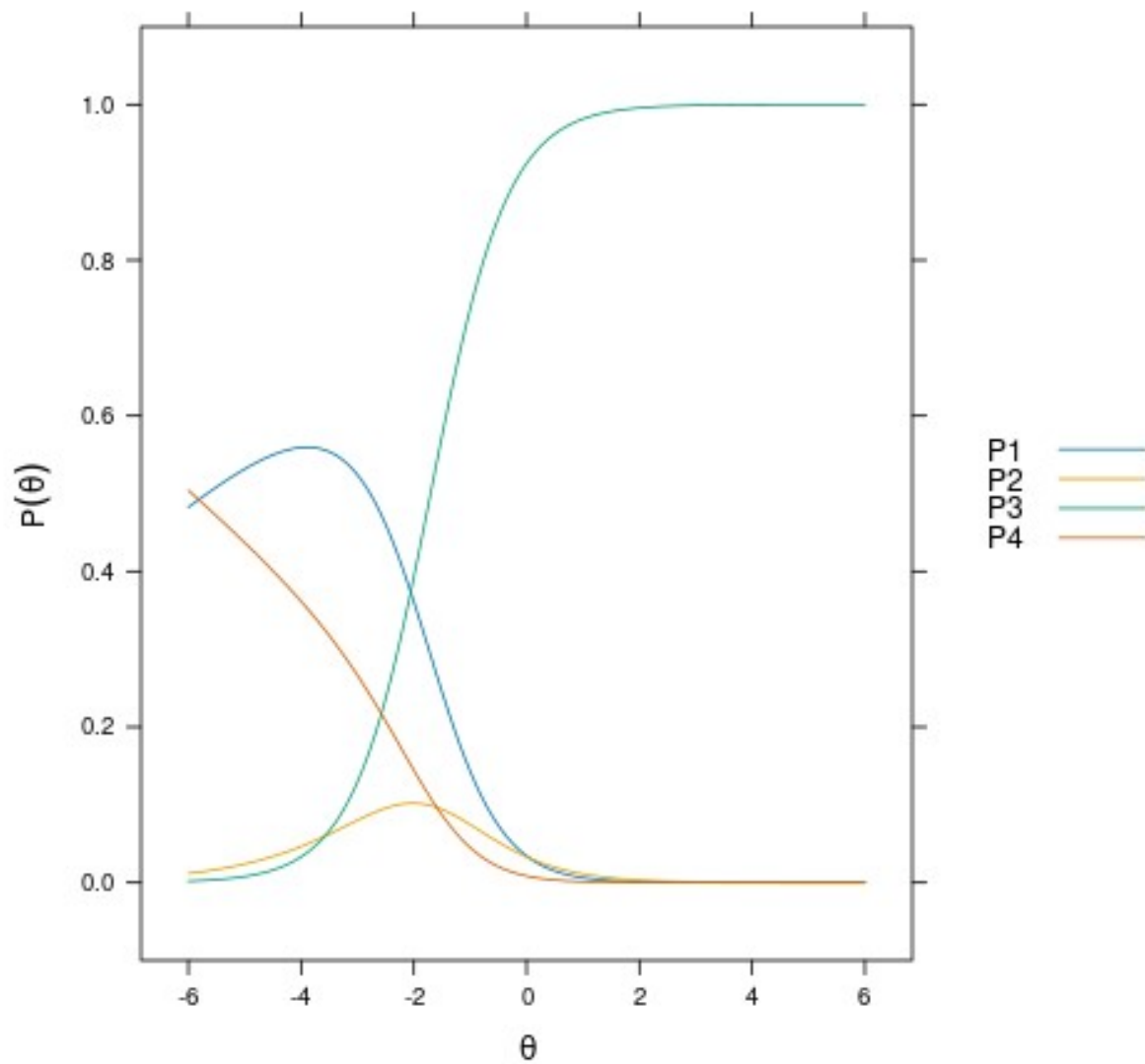
**Gráfico 120** – Função de Probabilidade – Item 82

**Fonte:** Elaborado própria

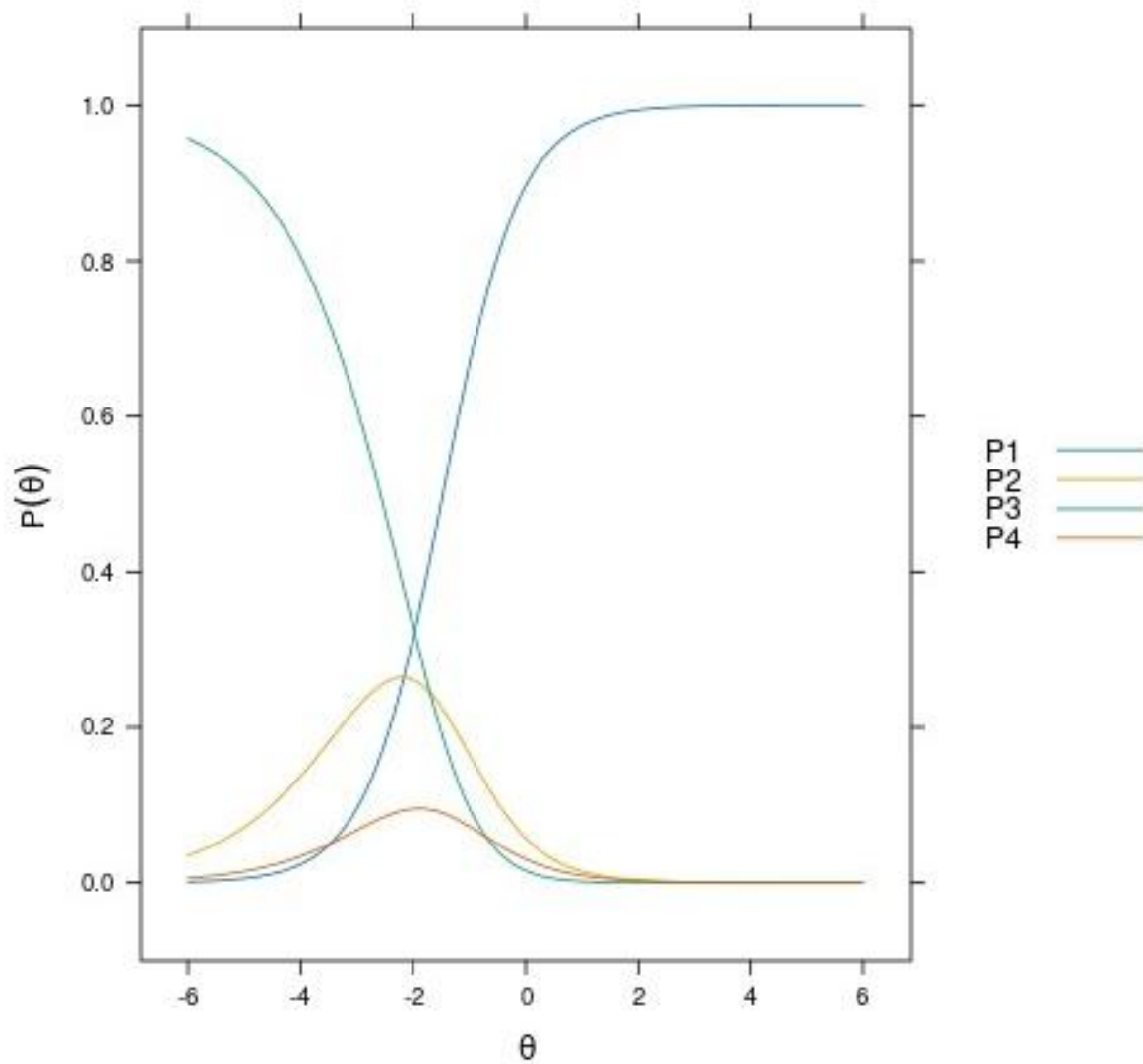
**Gráfico 121** – Função de Probabilidade – Item 83

**Fonte:** Elaborado própria

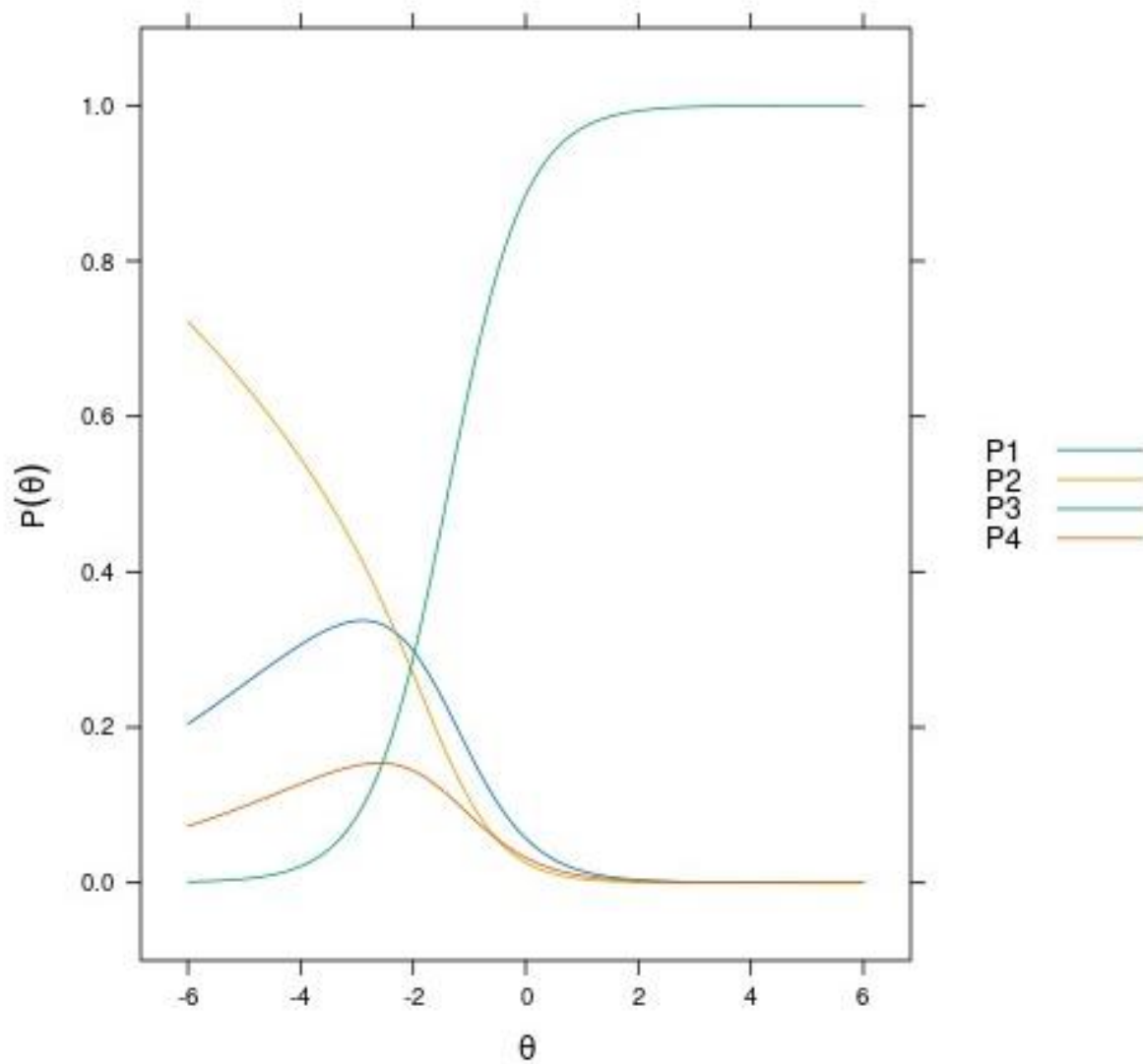


**Gráfico 122** – Função de Probabilidade – Item 84

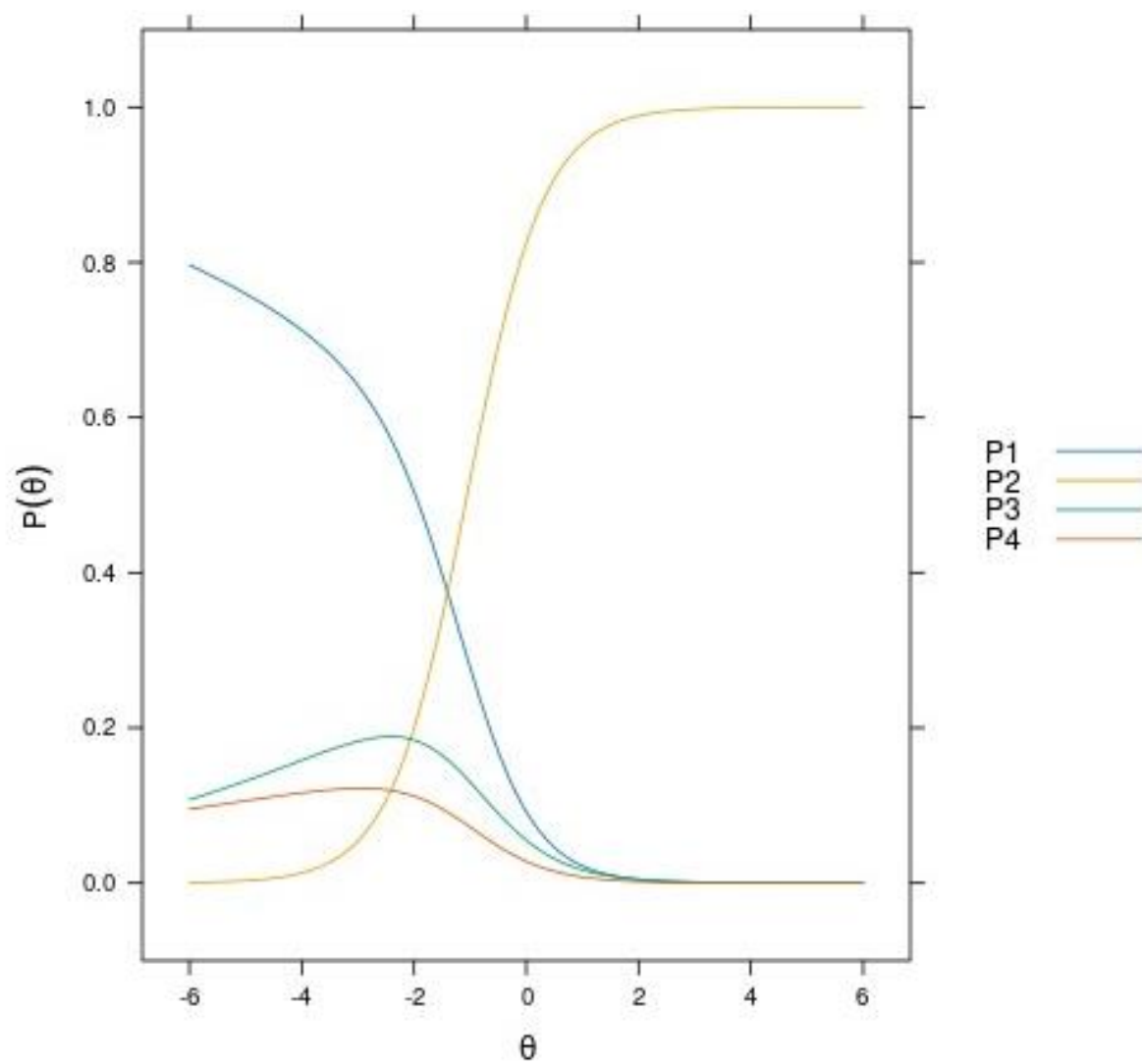
**Fonte:** Elaborado própria

**Gráfico 123** – Função de Probabilidade – Item 85

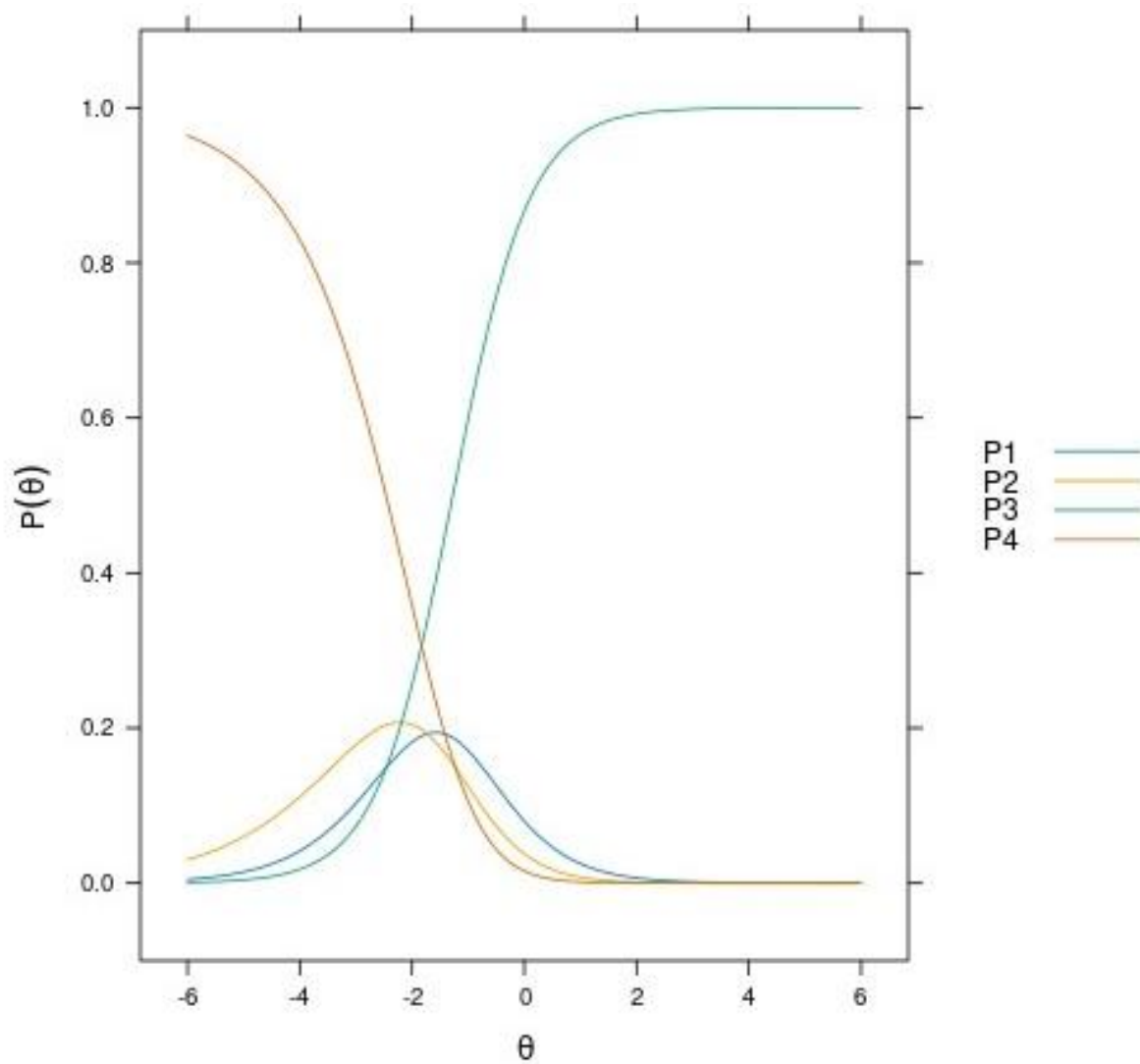
**Fonte:** Elaborado própria

**Gráfico 124** – Função de Probabilidade – Item 86

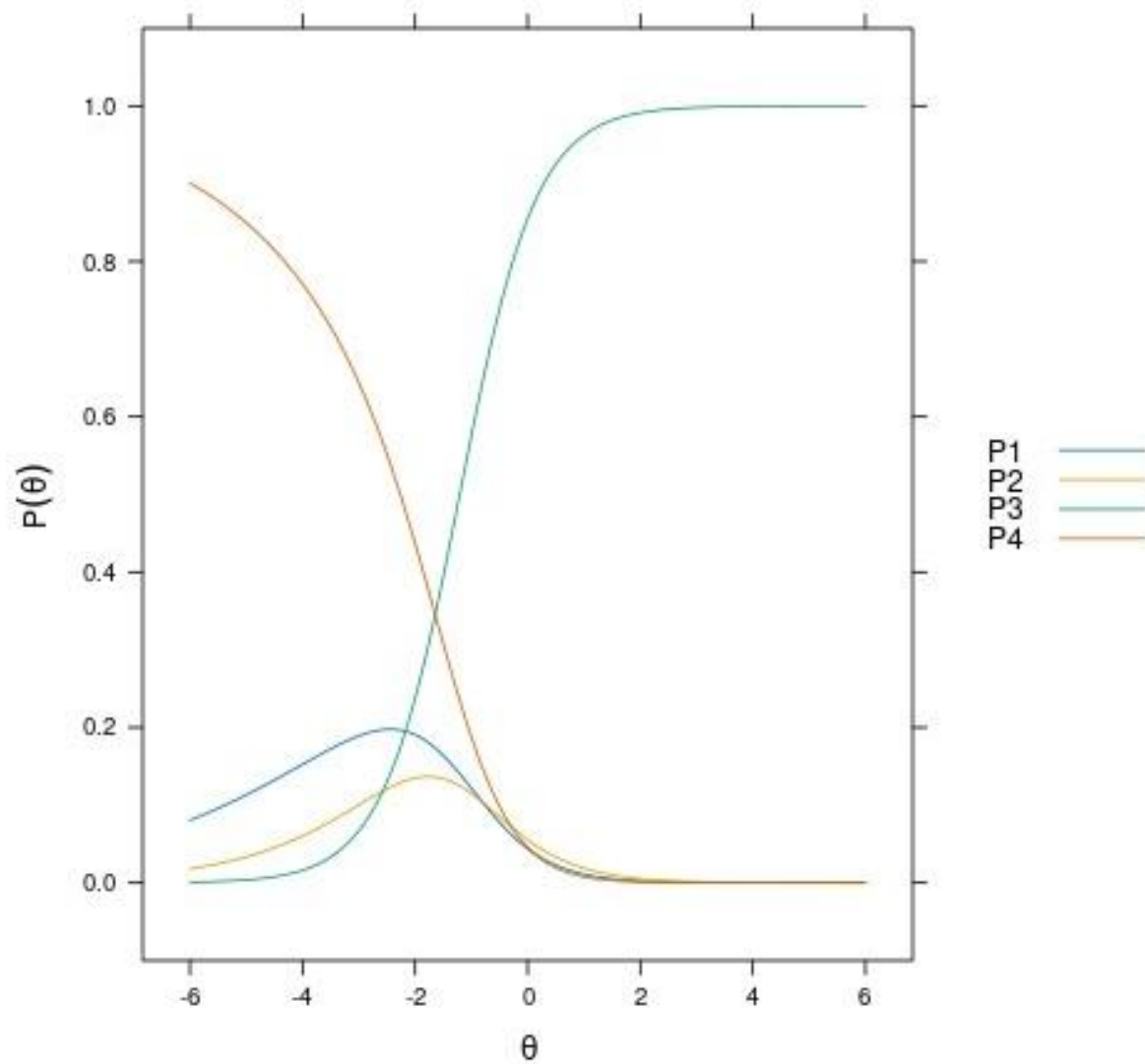
**Fonte:** Elaborado própria

**Gráfico 125** – Função de Probabilidade – Item 87

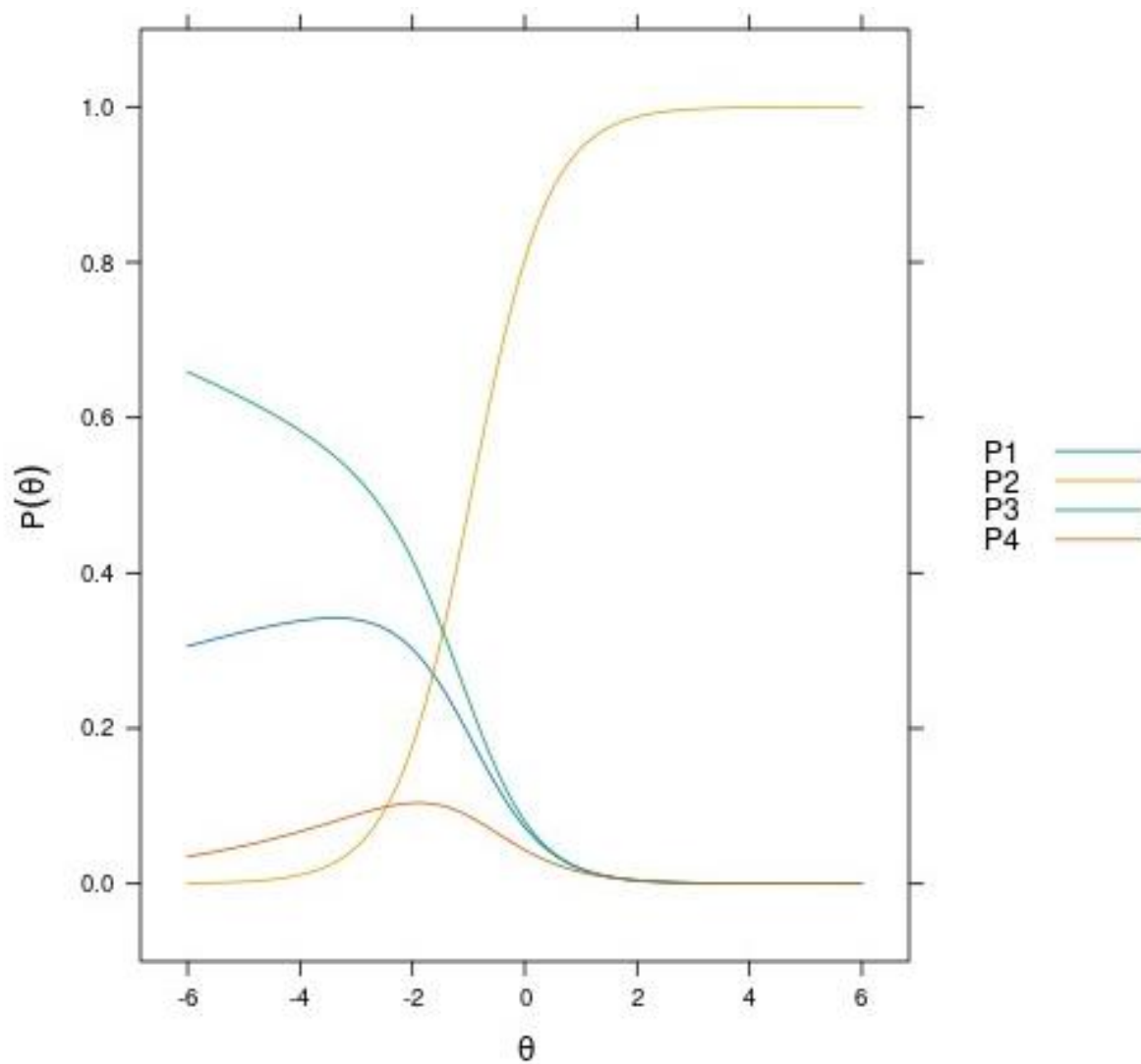
**Fonte:** Elaborado própria

**Gráfico 126** – Função de Probabilidade – Item 88

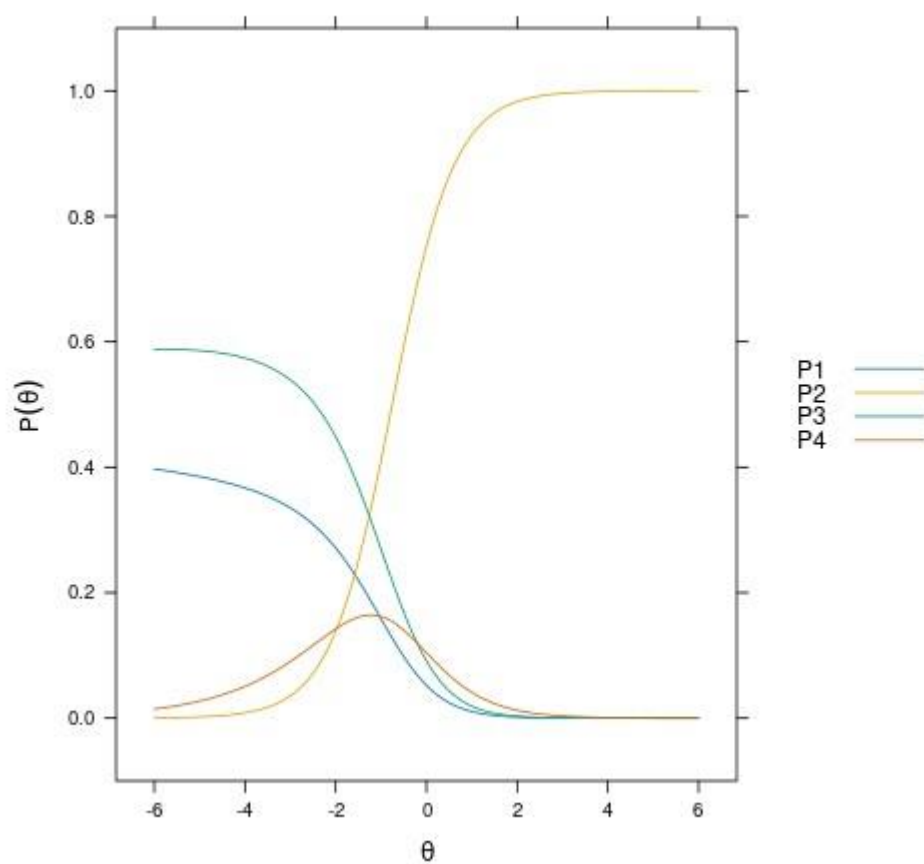
Fonte: Elaborado própria

**Gráfico 127** – Função de Probabilidade – Item 89

**Fonte:** Elaborado própria

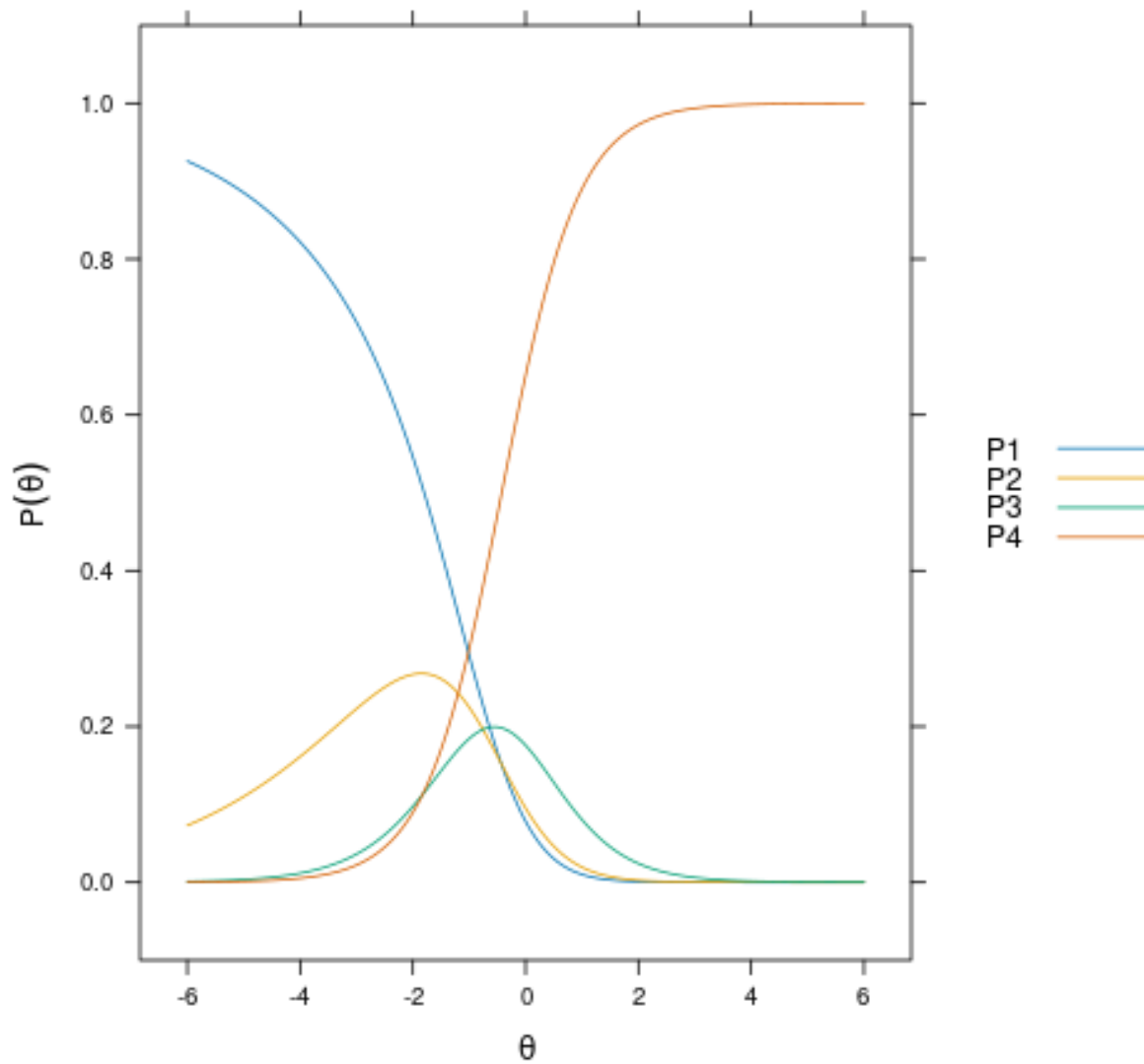
**Gráfico 128** – Função de Probabilidade – Item 90

**Fonte:** Elaborado própria

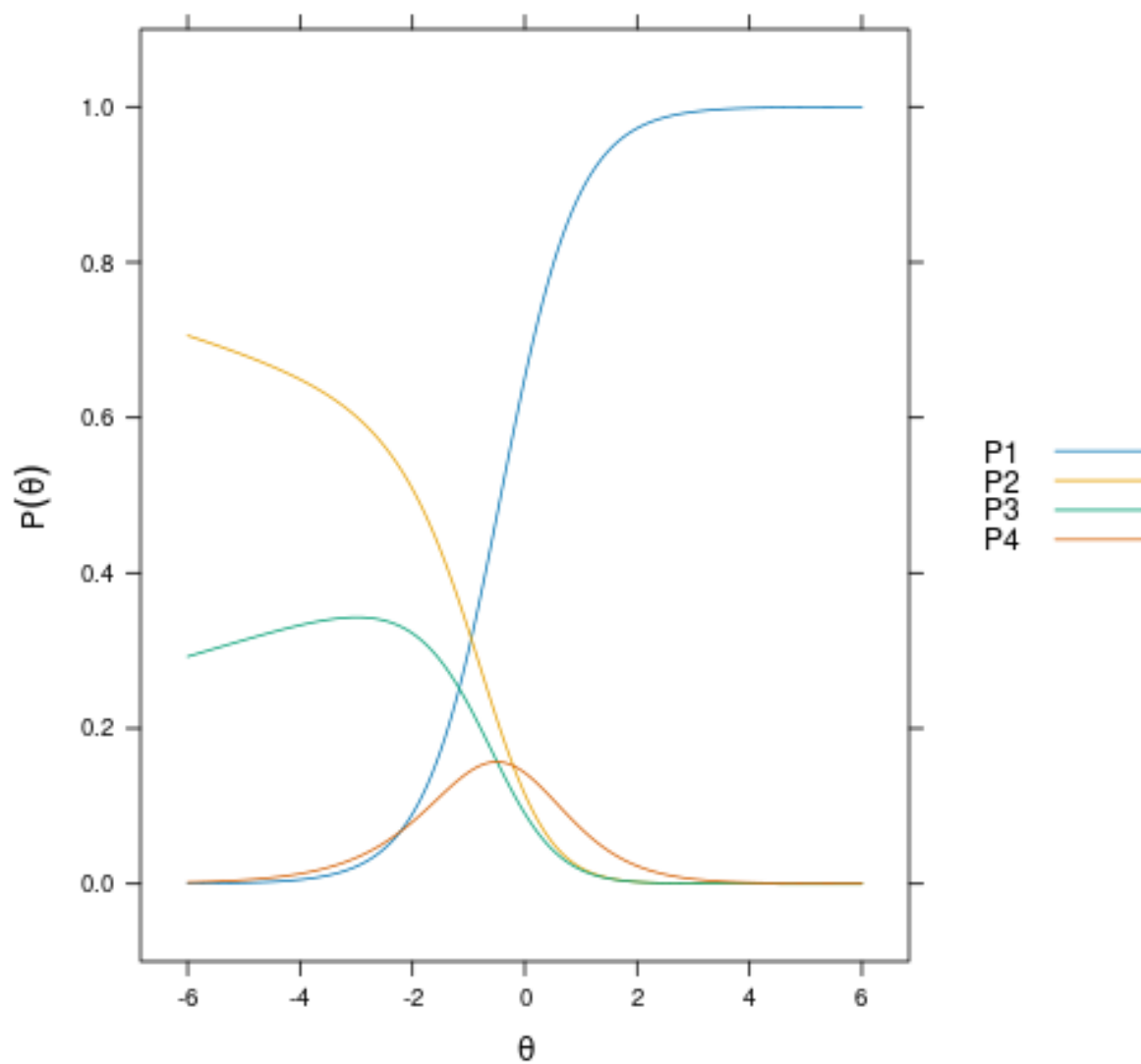
**Gráfico 129** – Função de Probabilidade – Item 91**Probability Function for Item 91**

**Fonte:** Elaborado própria

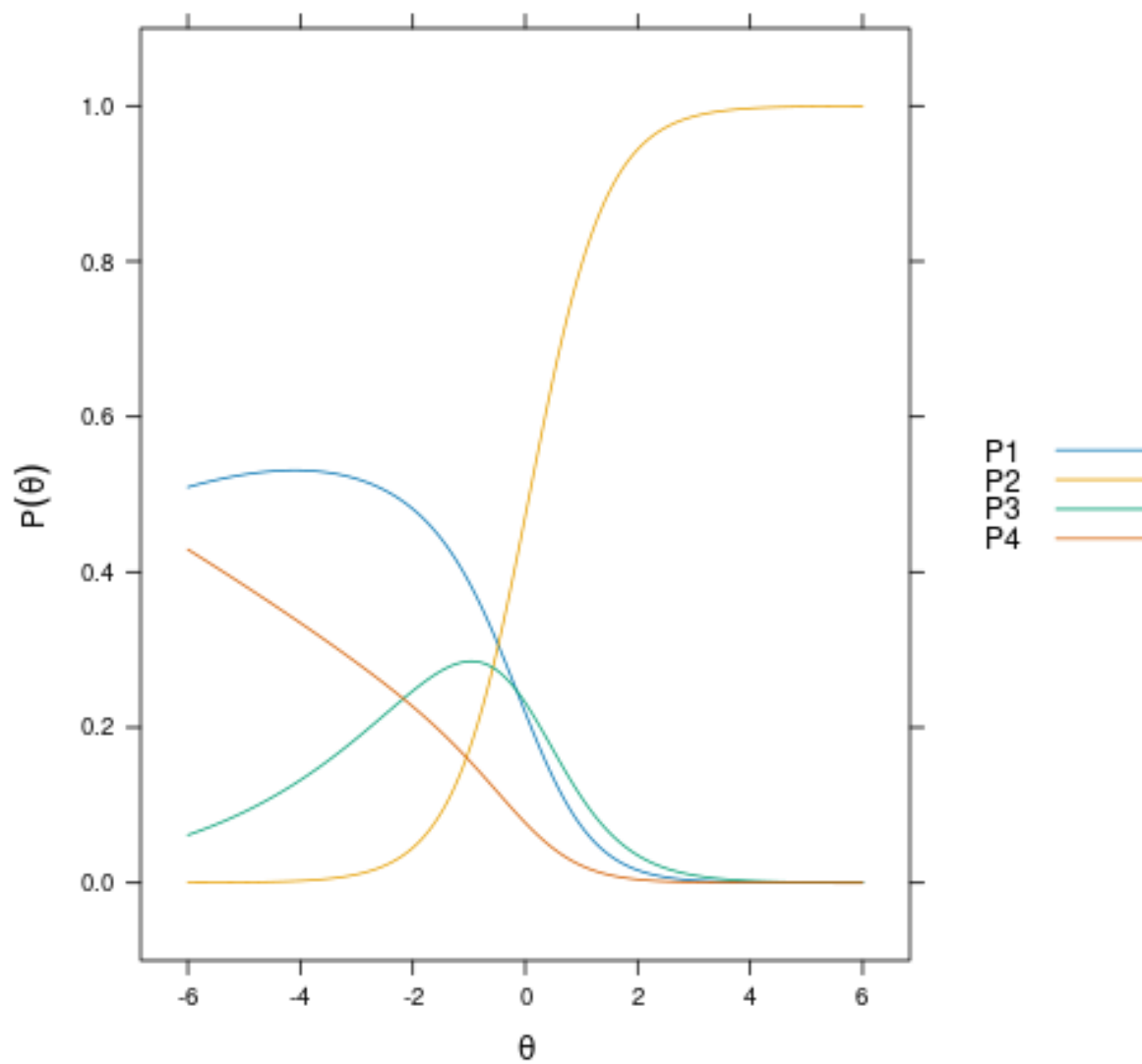


**Gráfico 130** – Função de Probabilidade – Item 92

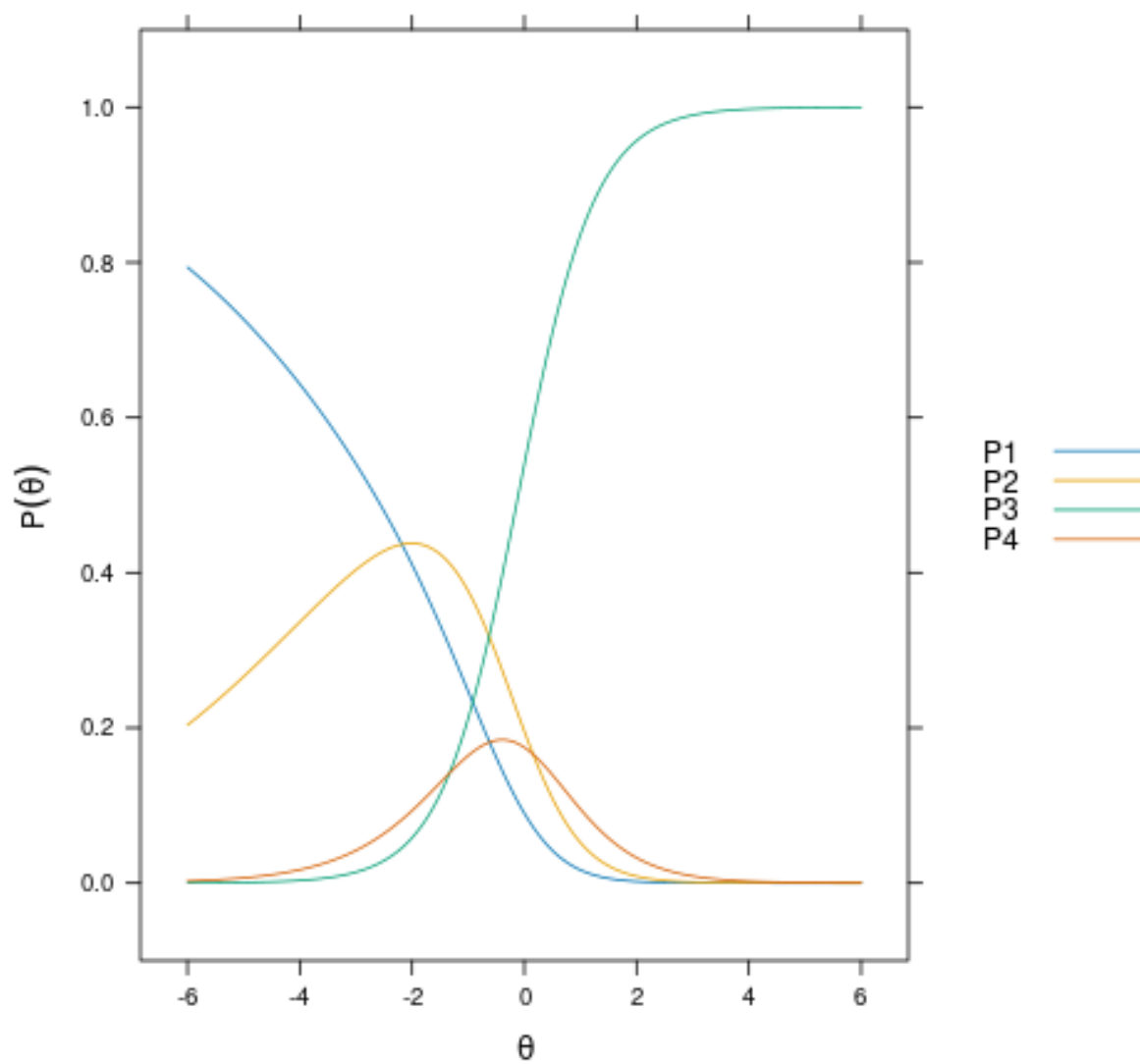
Fonte: Elaborado própria

**Gráfico 131** – Função de Probabilidade – Item 93

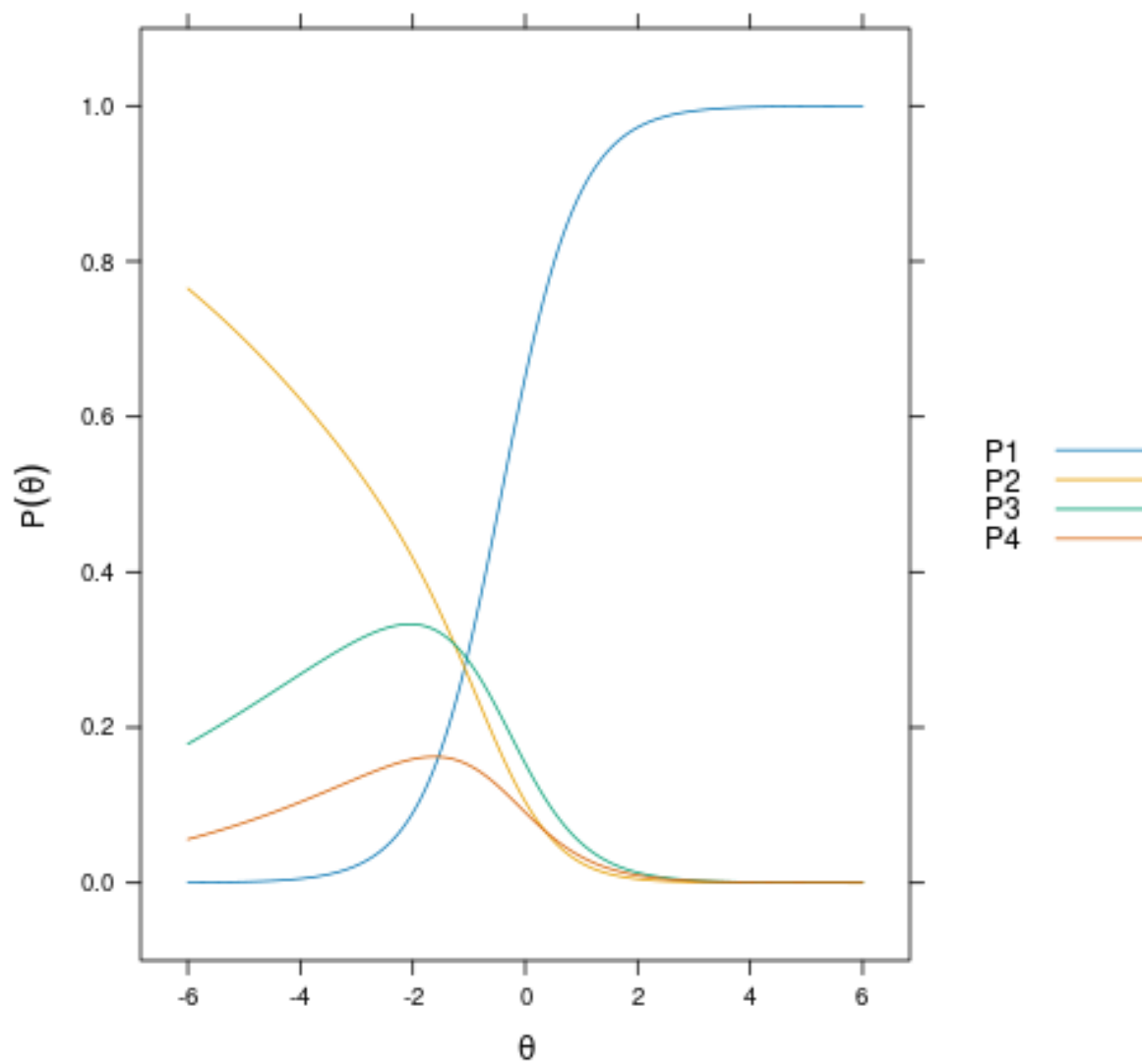
**Fonte:** Elaborado própria

**Gráfico 132** – Função de Probabilidade – Item 94

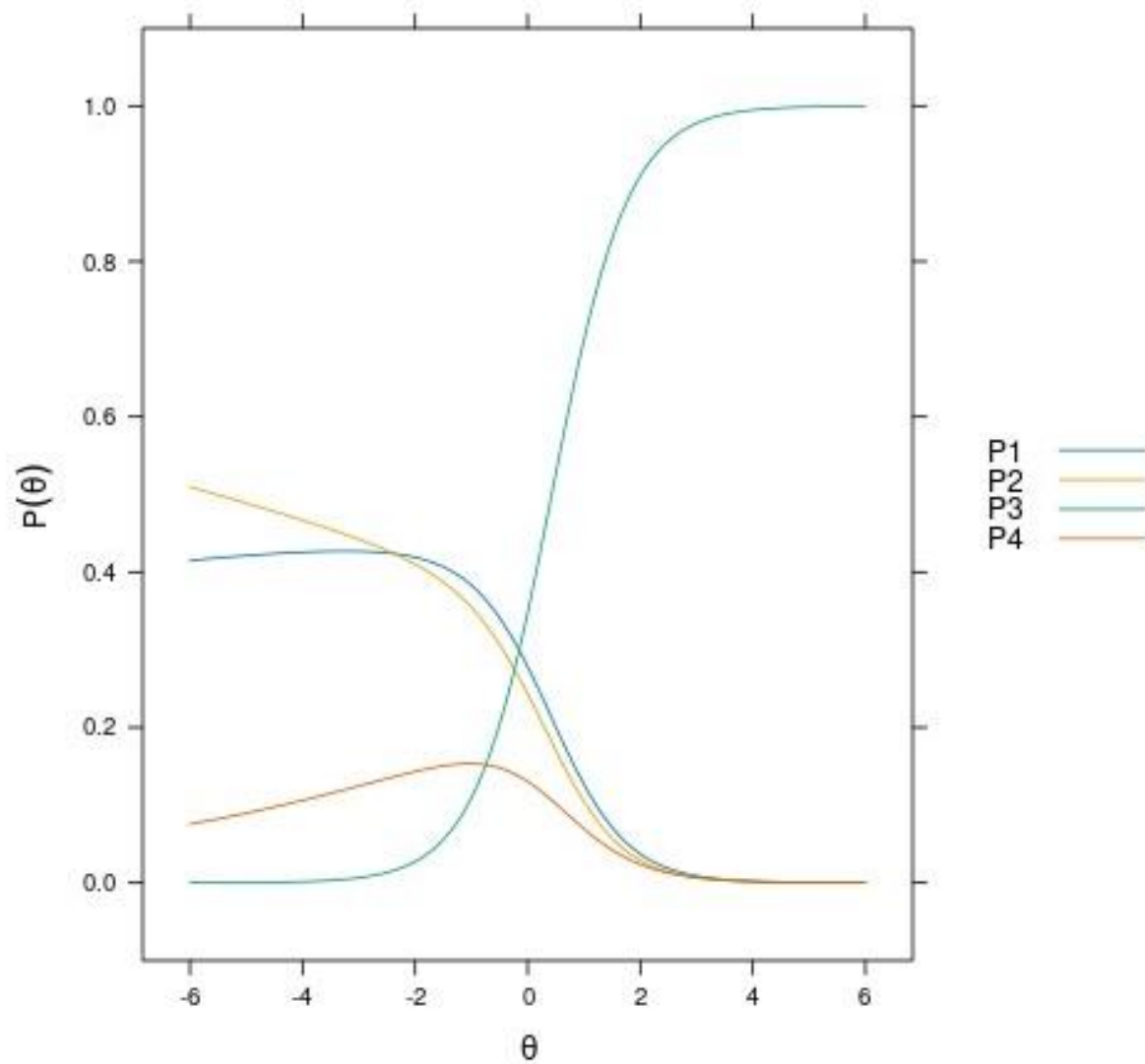
Fonte: Elaborado própria

**Gráfico 133** – Função de Probabilidade – Item 95

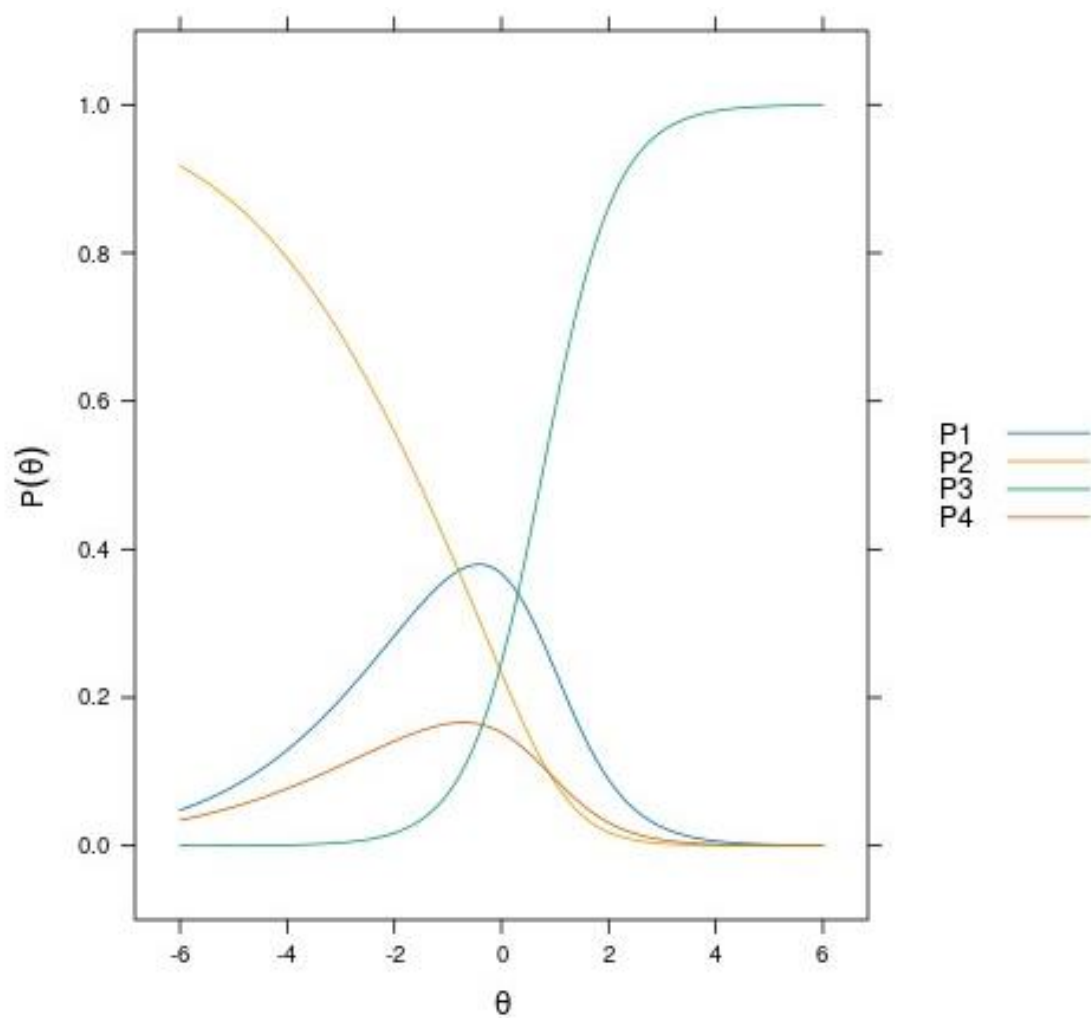
**Fonte:** Elaborado própria

**Gráfico 134** – Função de Probabilidade – Item 96

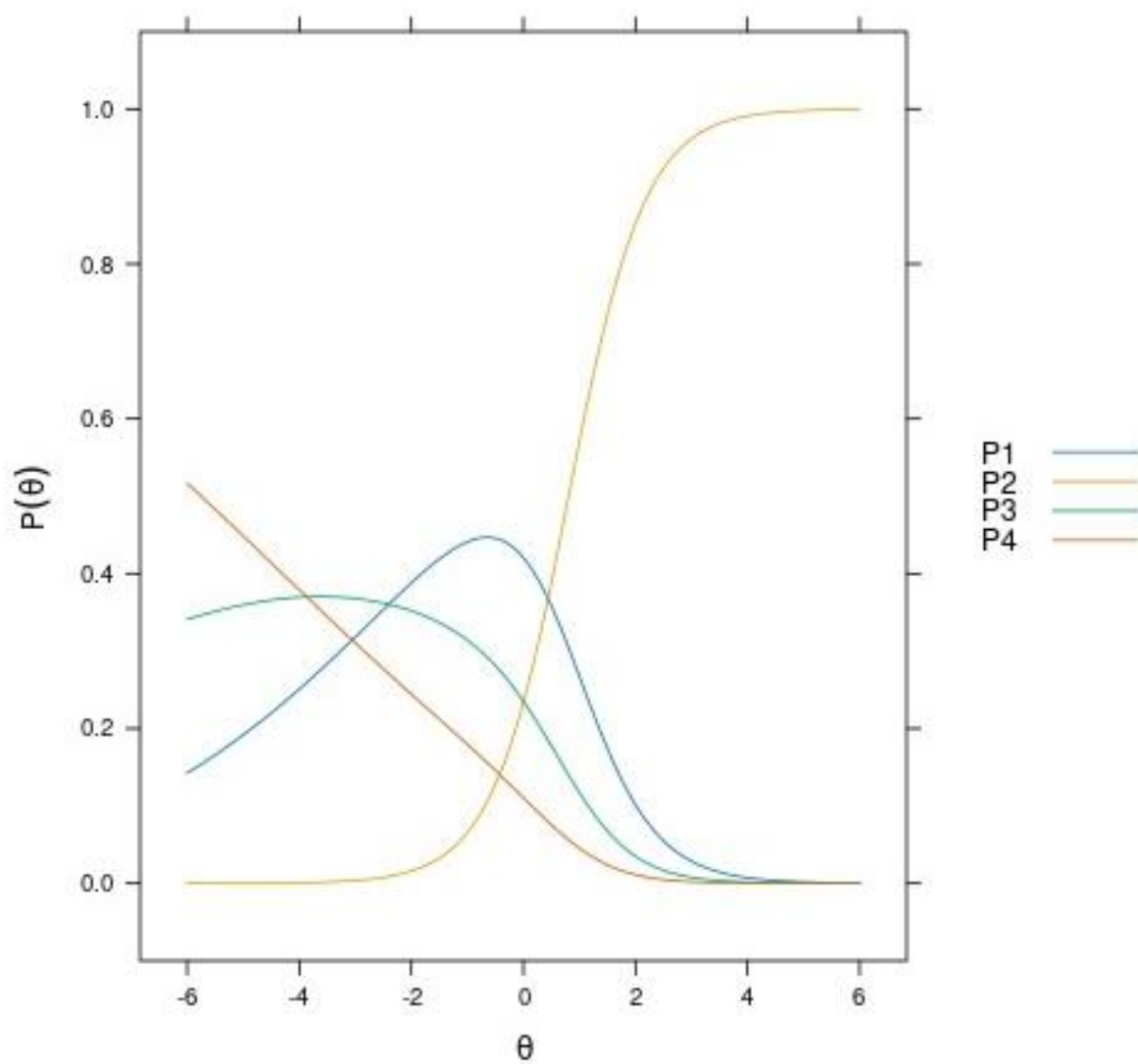
Fonte: Elaborado própria

**Gráfico 135** – Função de Probabilidade – Item 97

**Fonte:** Elaborado própria

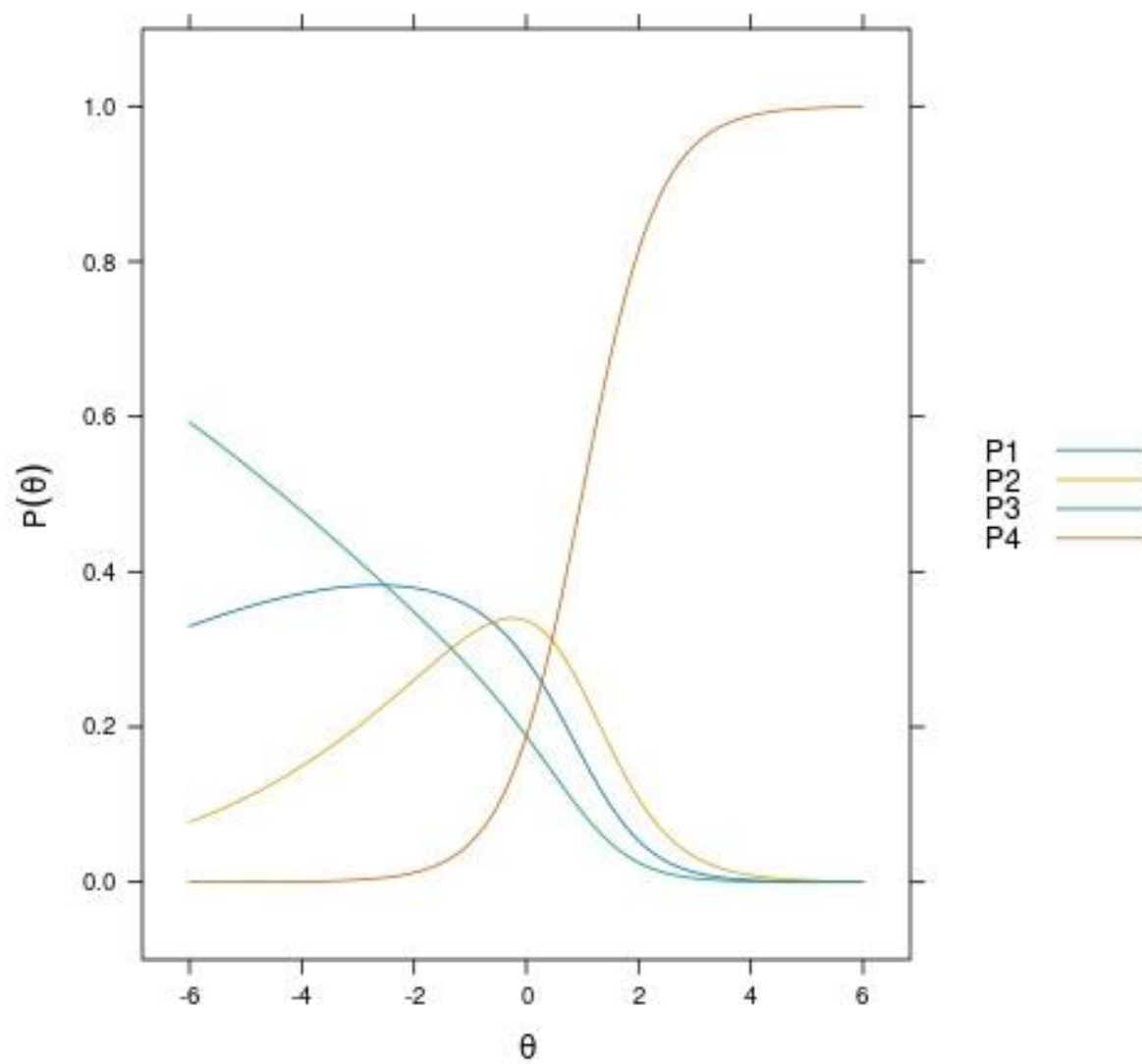
**Gráfico 136** – Função de Probabilidade – Item 98

**Fonte:** Elaborado própria

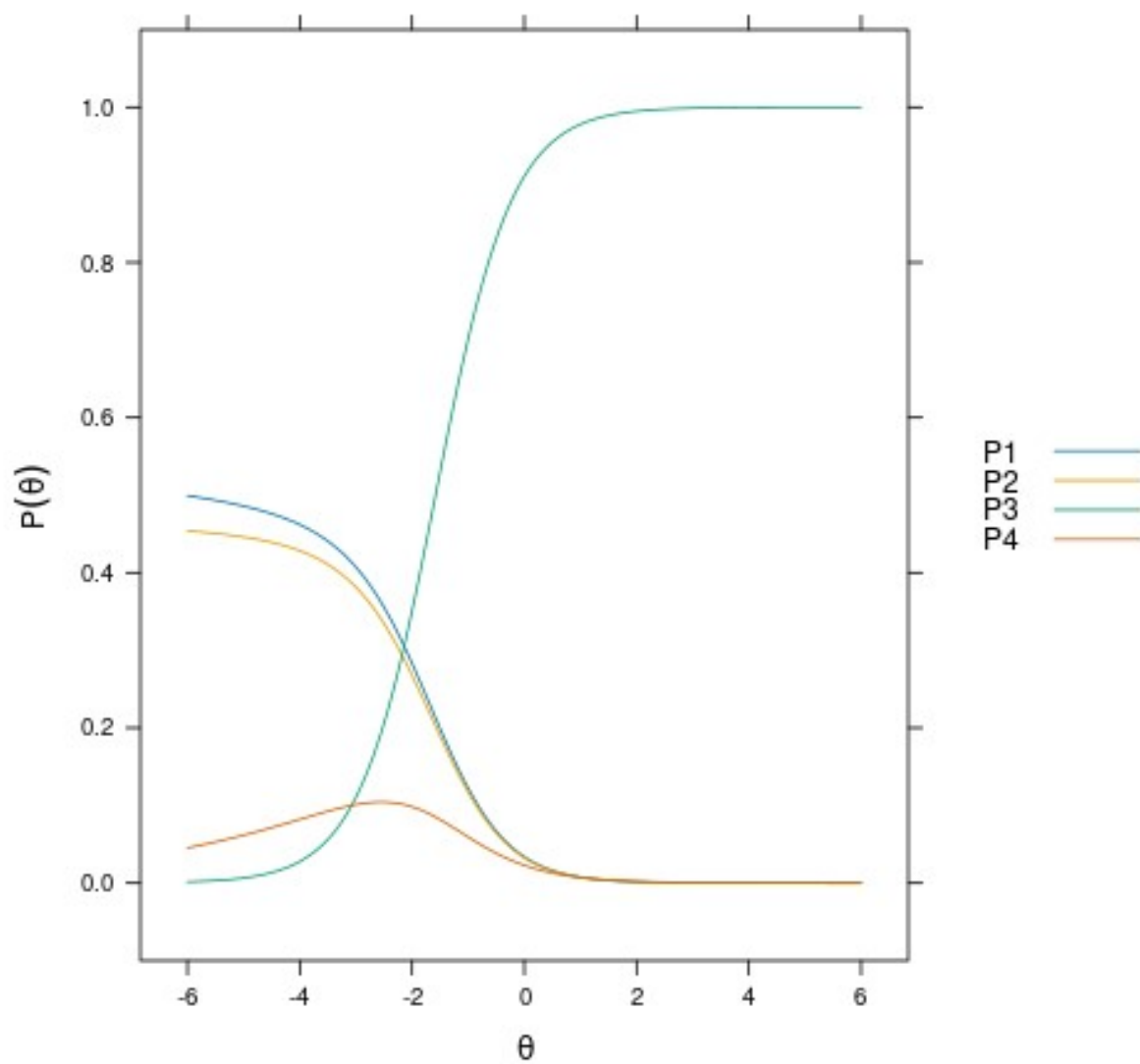
**Gráfico 137** – Função de Probabilidade – Item 99

**Fonte:** Elaborado própria

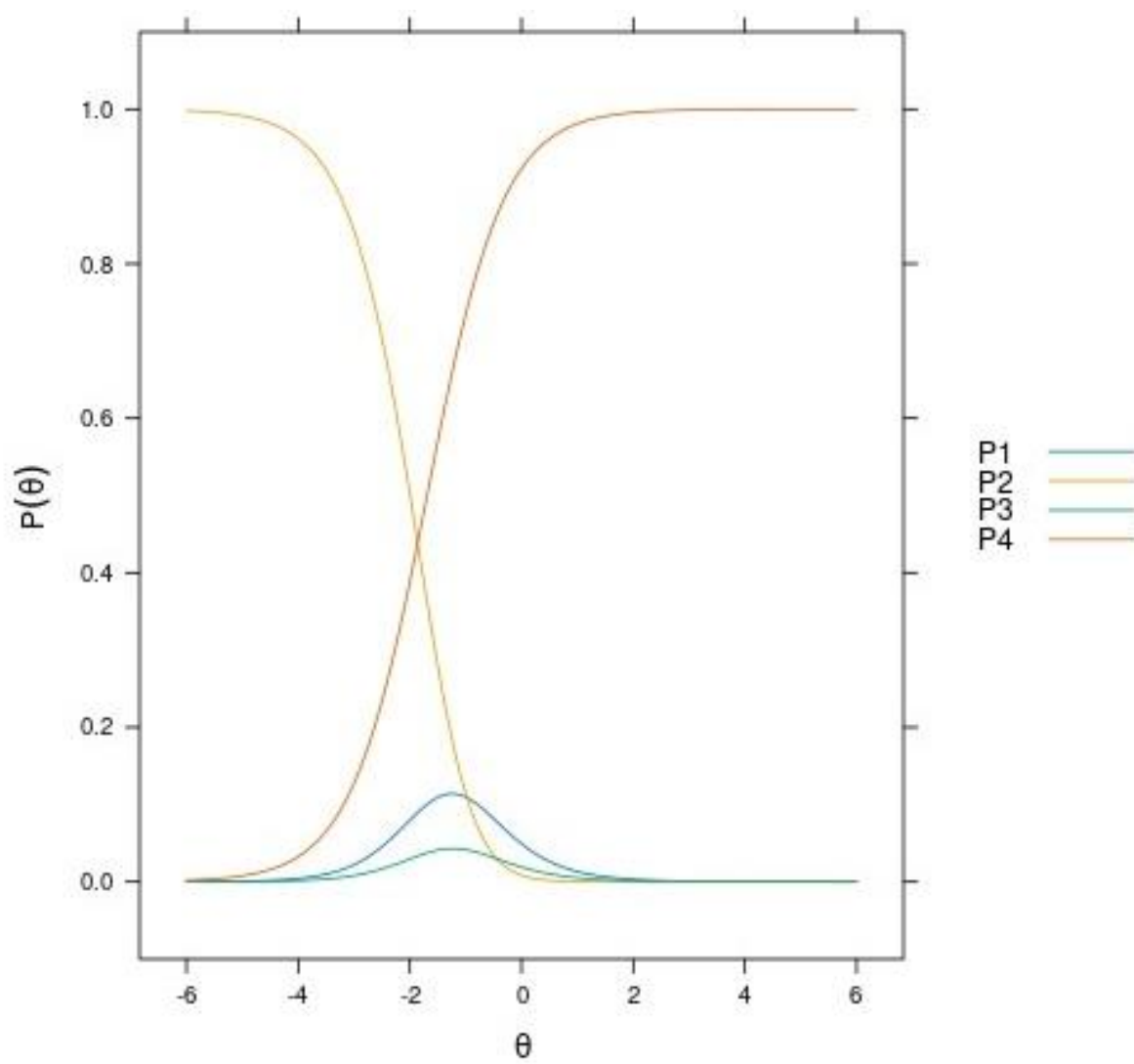


**Gráfico 138** – Função de Probabilidade – Item 100

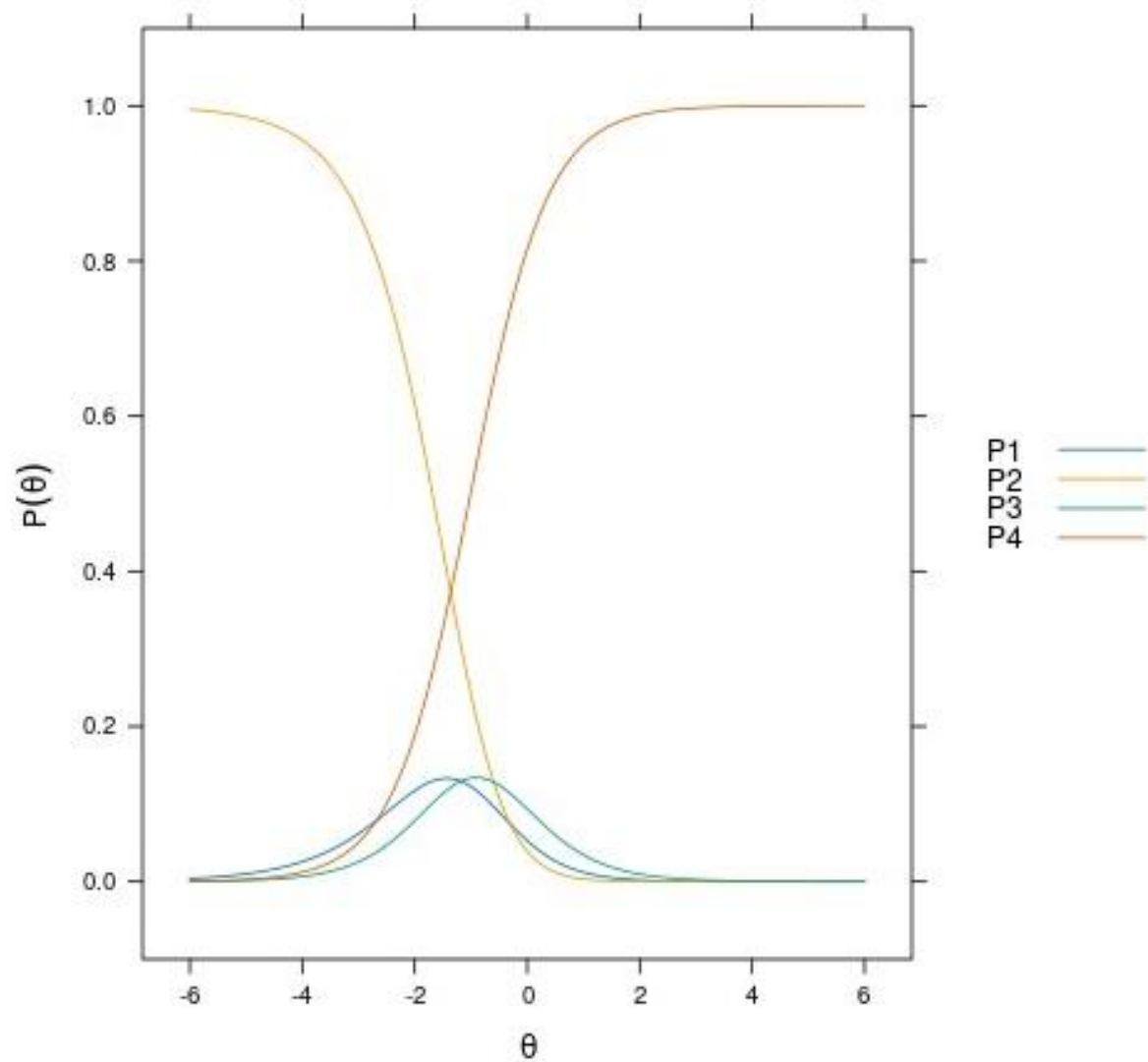
**Fonte:** Elaborado própria

**Gráfico 149** – Função de Probabilidade – Item 101

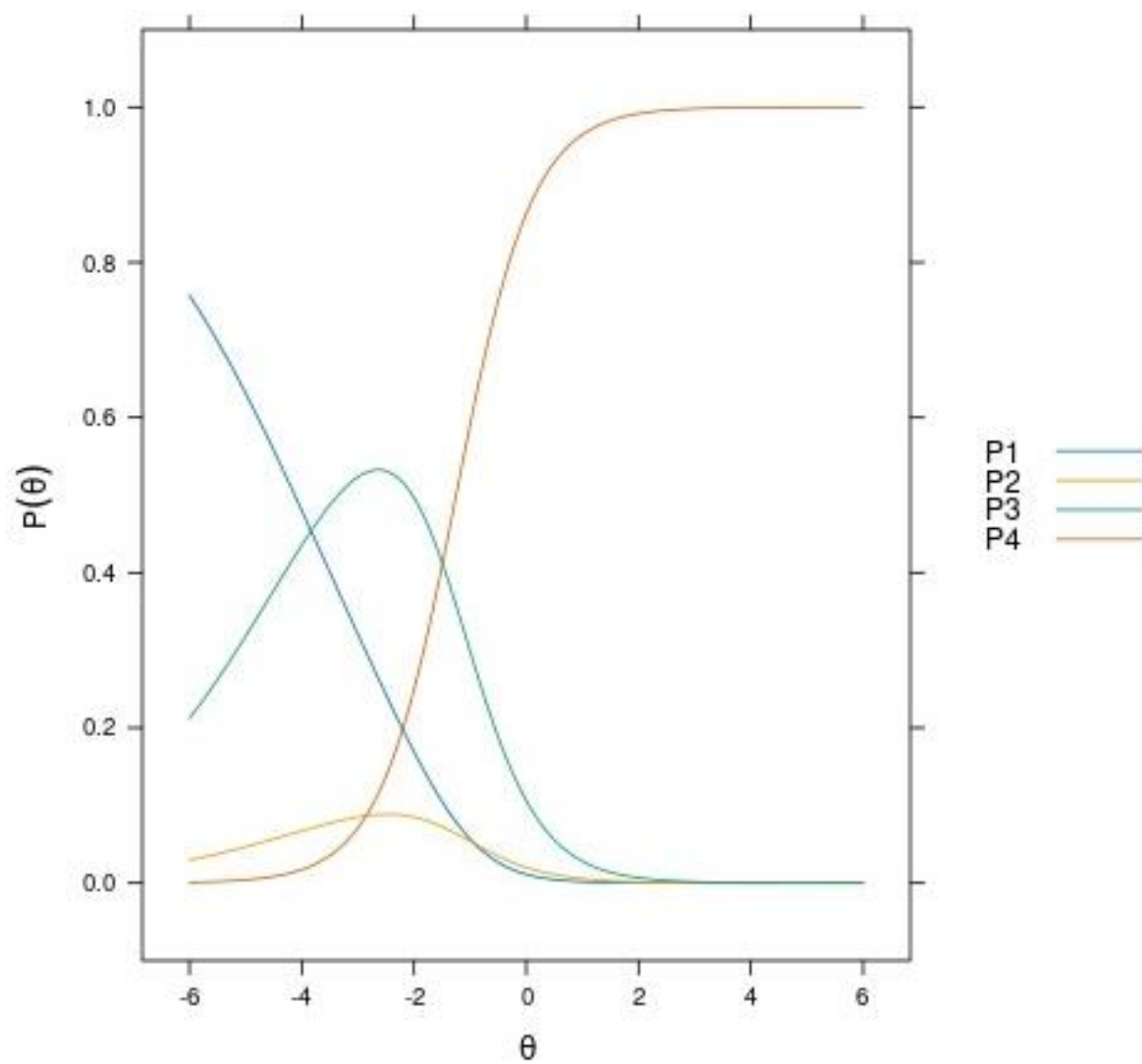
**Fonte:** Elaborado própria

**Gráfico 140** – Função de Probabilidade – Item 102

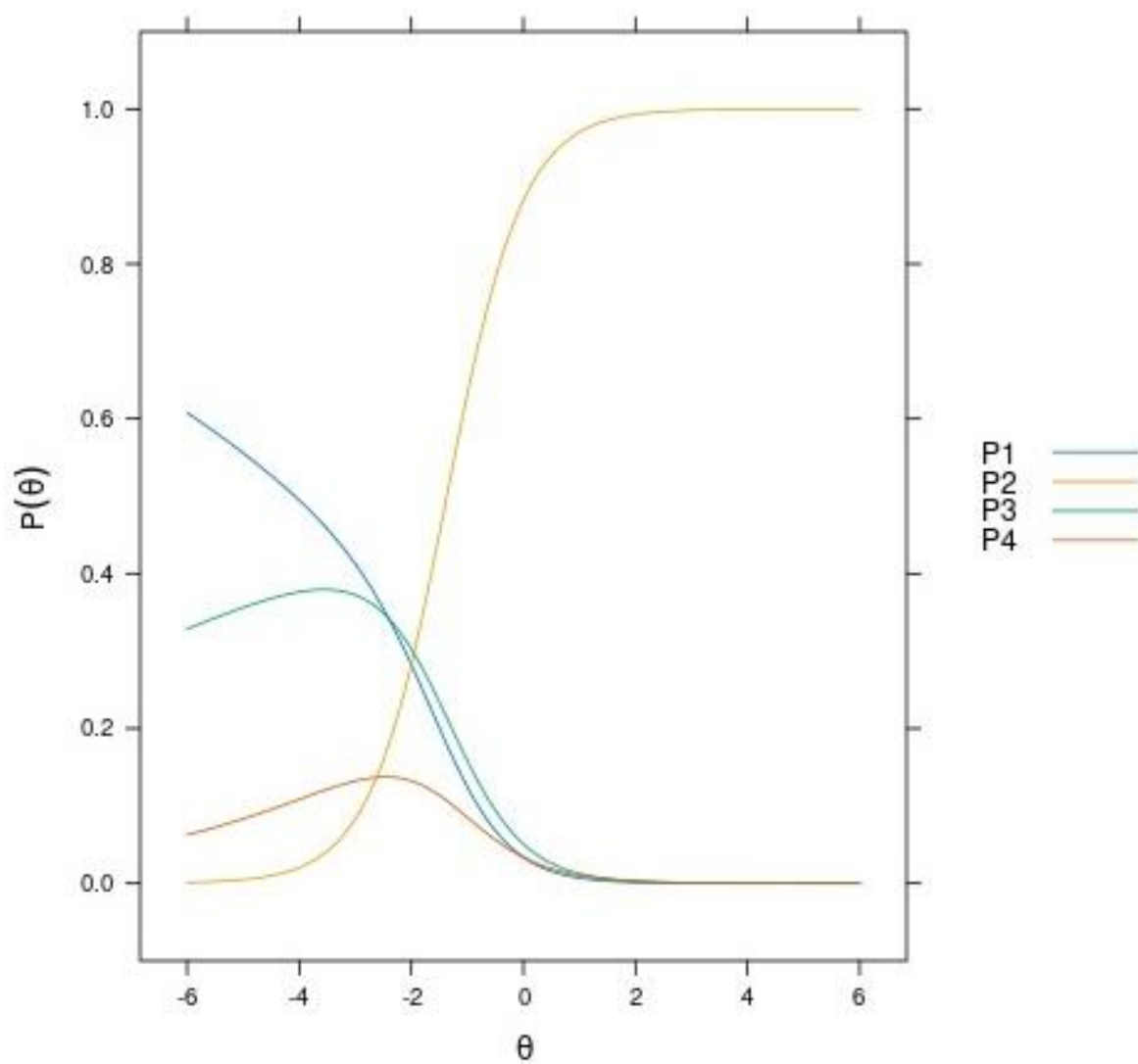
**Fonte:** Elaborado própria

**Gráfico 141** – Função de Probabilidade – Item 103

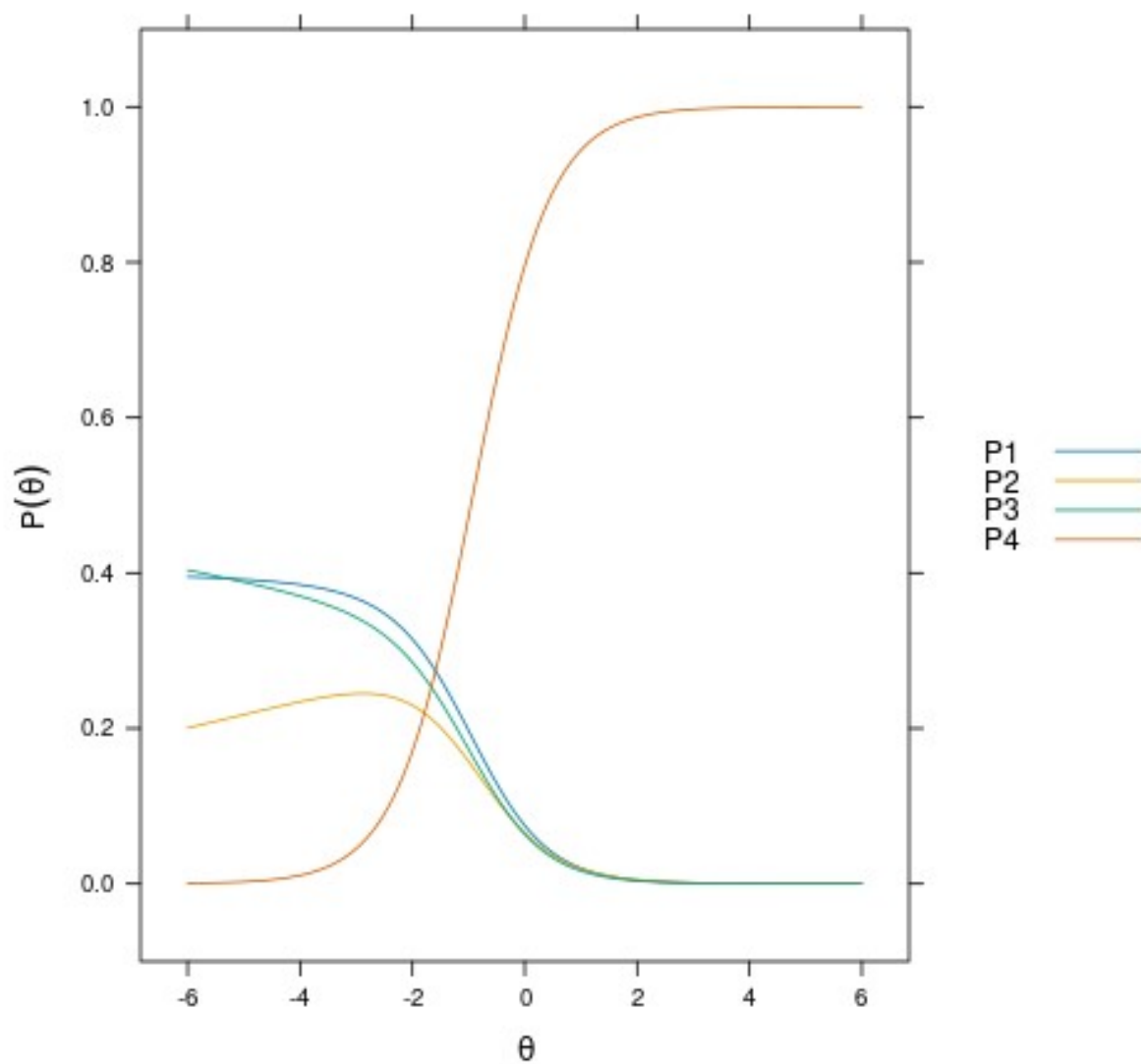
**Fonte:** Elaborado própria

**Gráfico 142** – Função de Probabilidade – Item 104

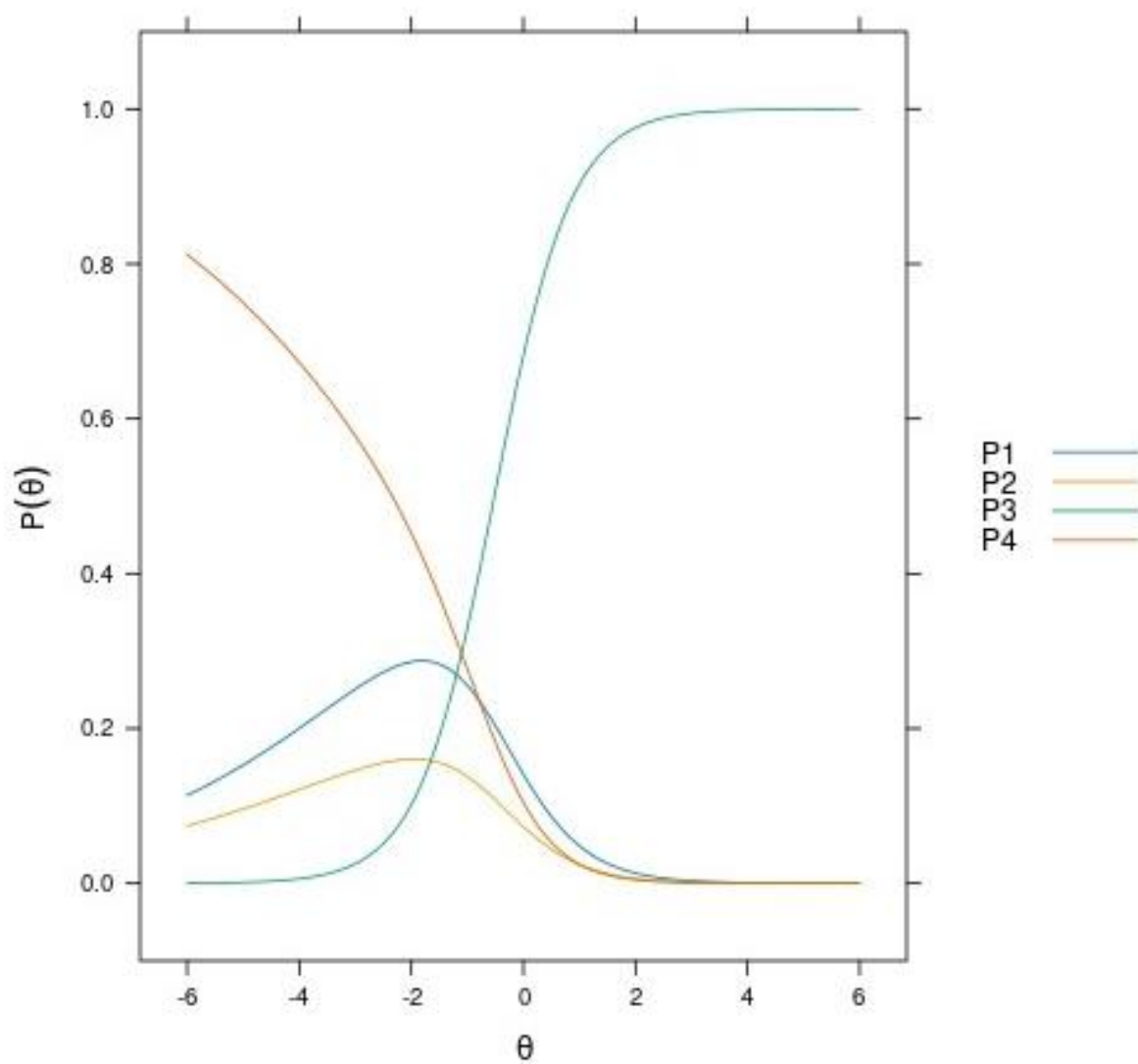
**Fonte:** Elaborado própria

**Gráfico 143** – Função de Probabilidade – Item 105

**Fonte:** Elaborado própria

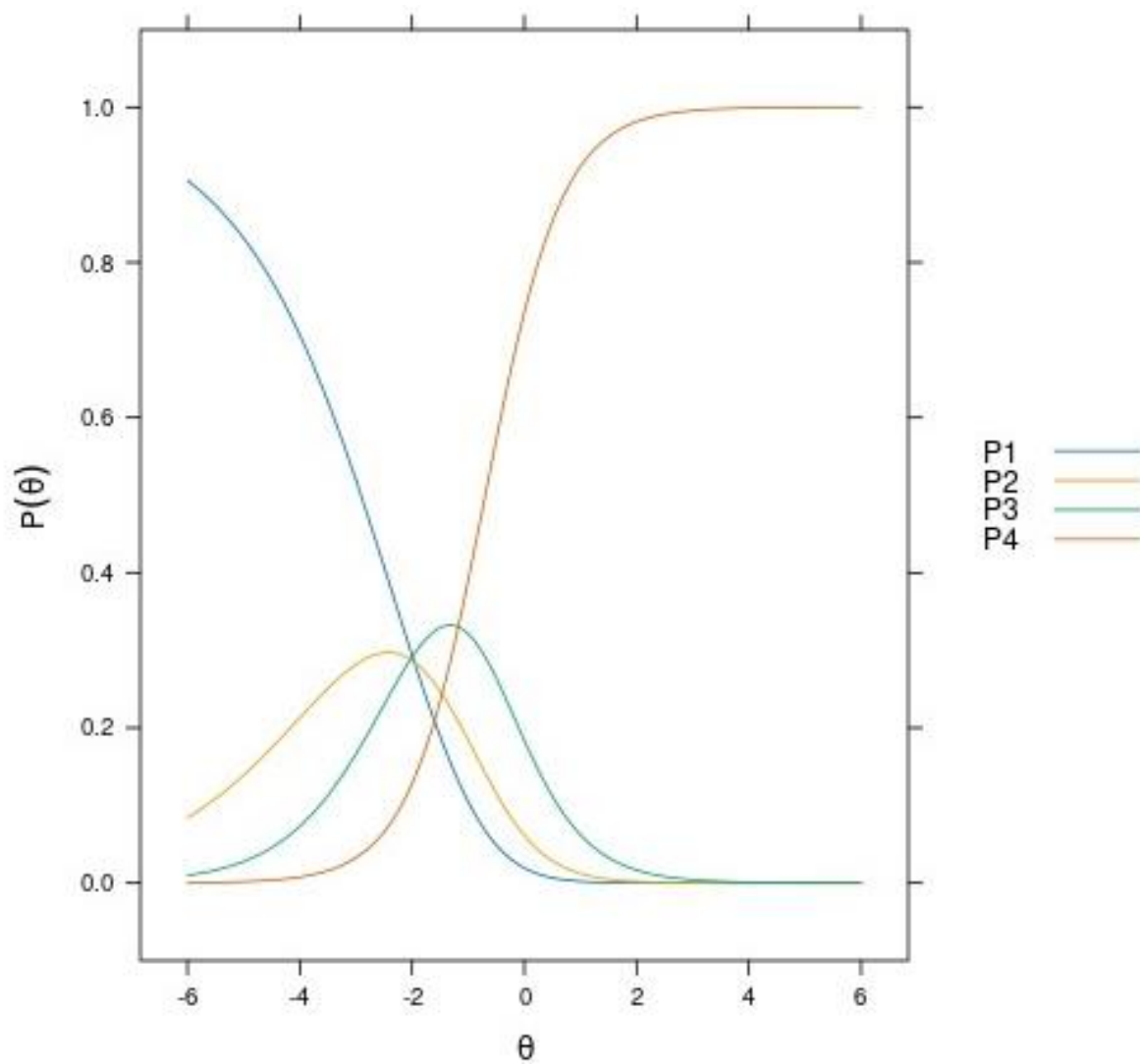
**Gráfico 144** – Função de Probabilidade – Item 106

**Fonte:** Elaborado própria

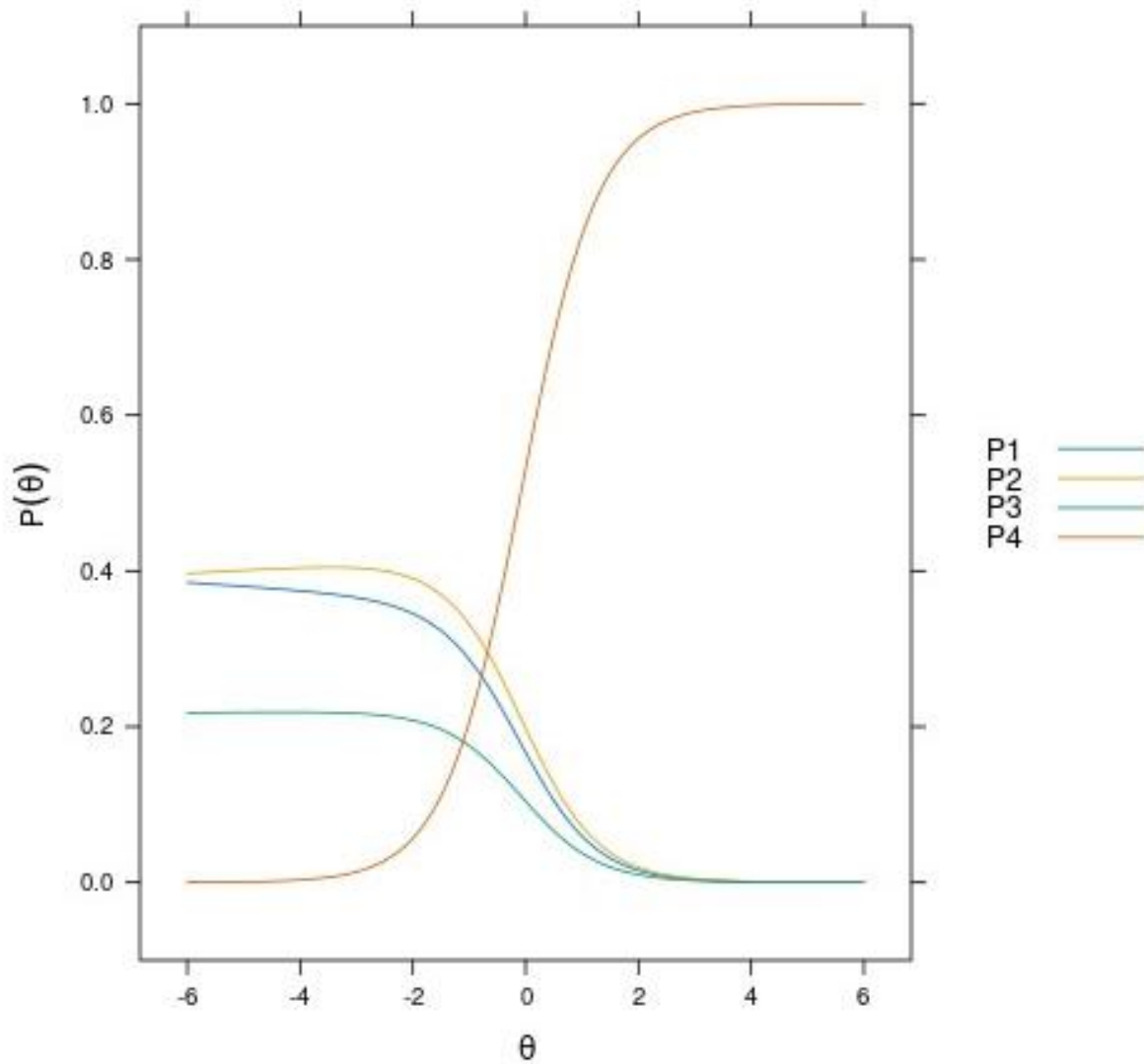
**Gráfico 145** – Função de Probabilidade – Item 107

**Fonte:** Elaborado própria

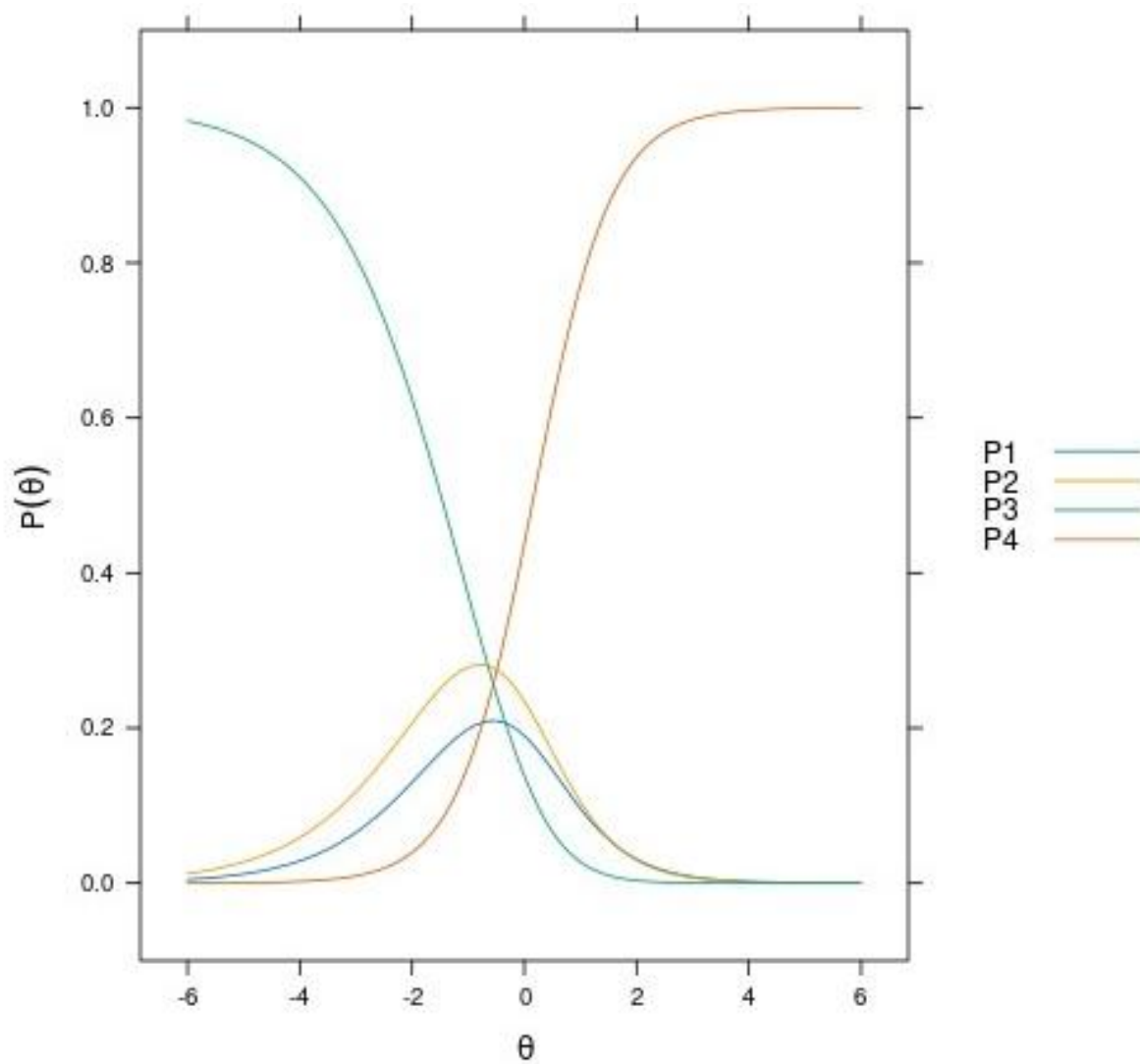


**Gráfico 146** – Função de Probabilidade – Item 108

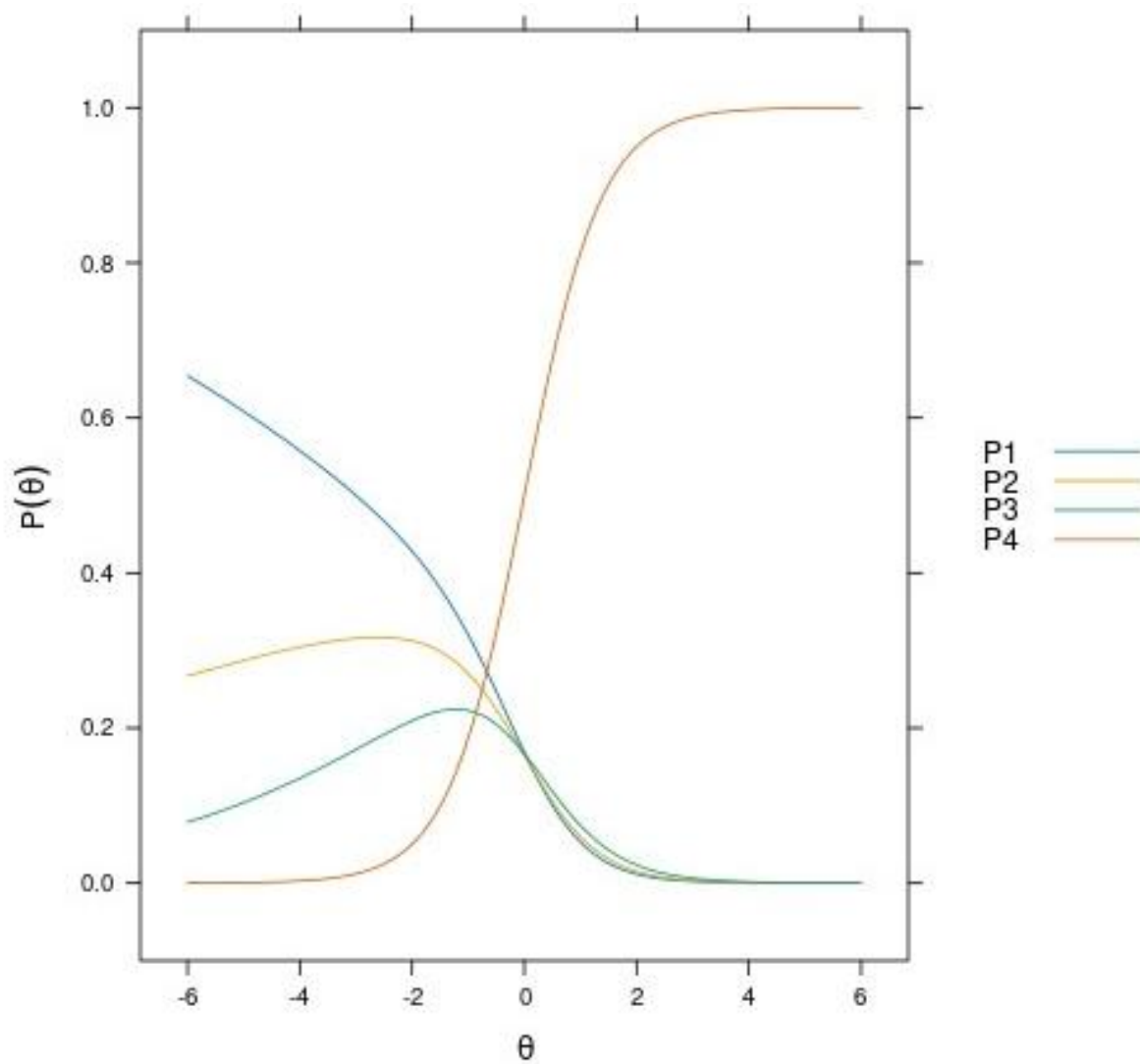
**Fonte:** Elaborado própria

**Gráfico 147** – Função de Probabilidade – Item 109

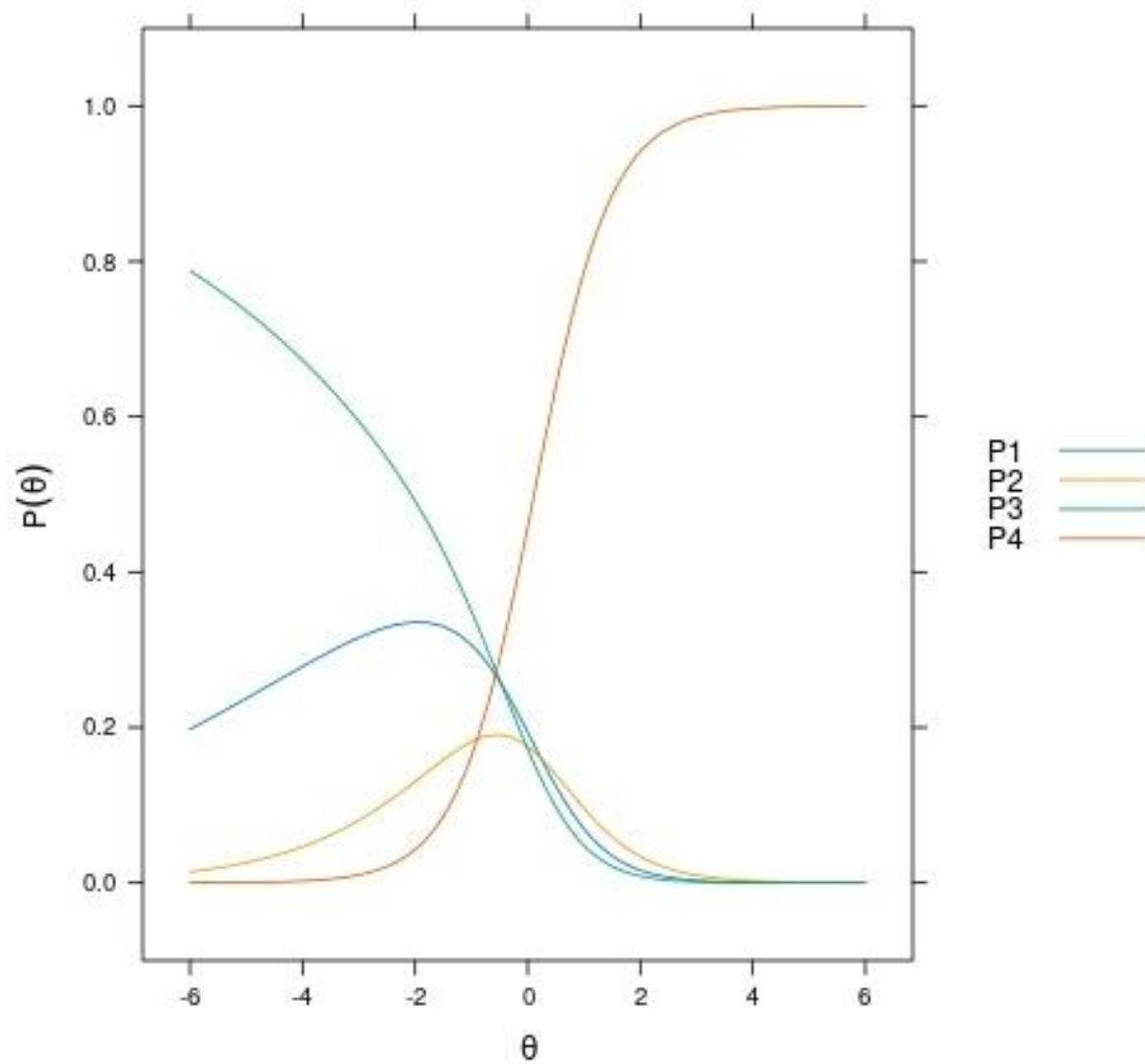
**Fonte:** Elaborado própria

**Gráfico 148** – Função de Probabilidade – Item 110

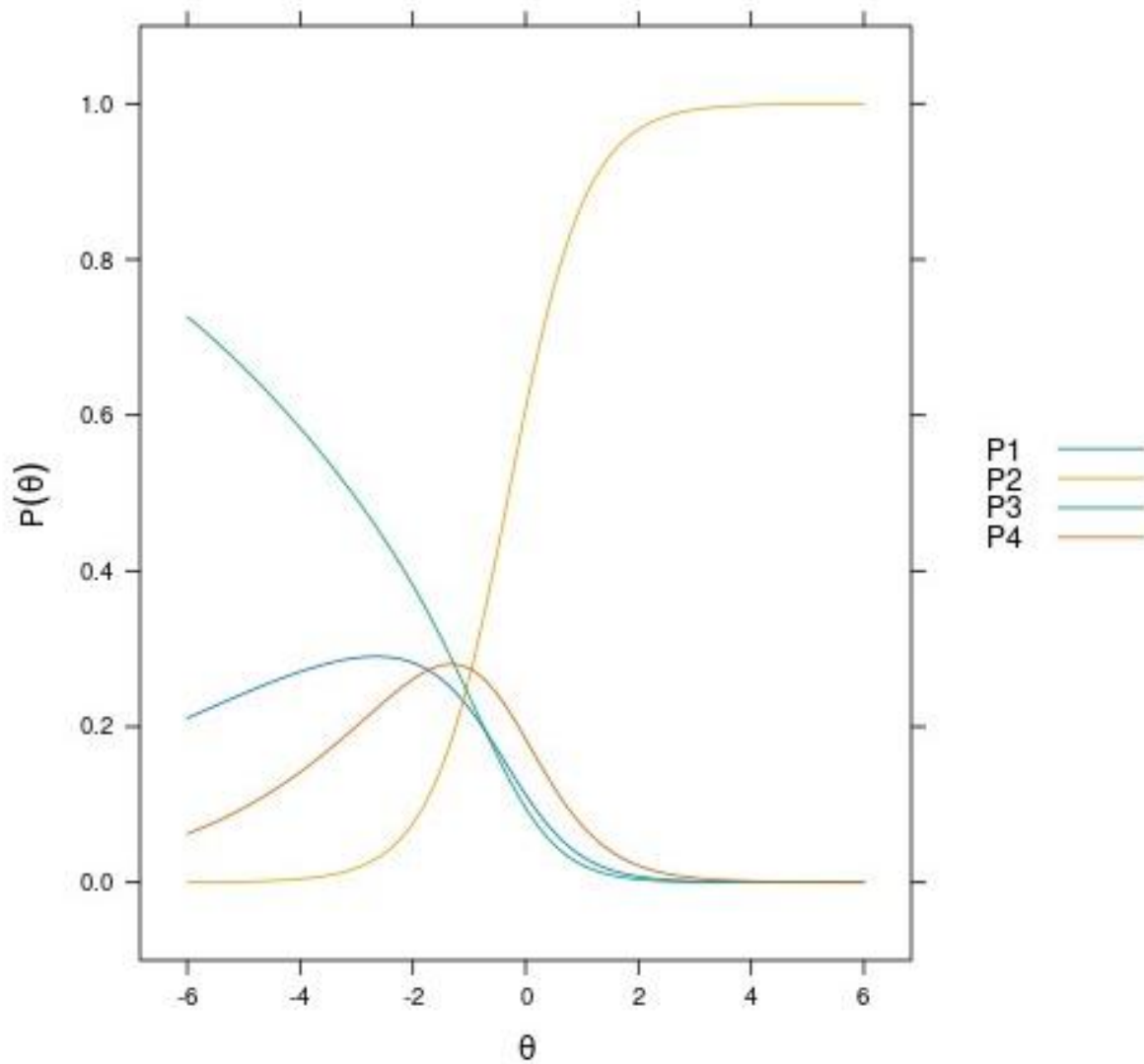
Fonte: Elaborado própria

**Gráfico 149** – Função de Probabilidade – Item 111

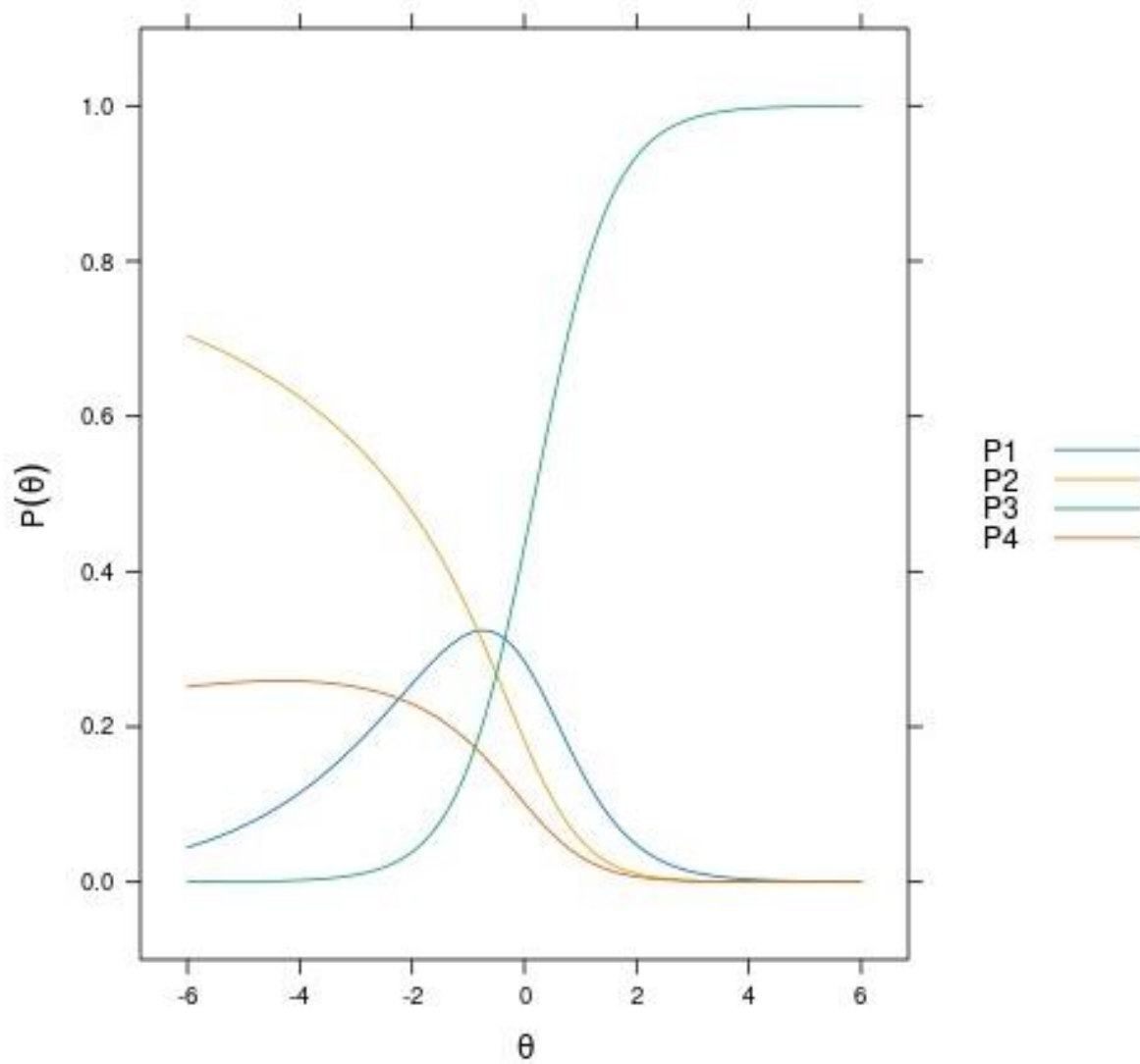
Fonte: Elaborado própria

**Gráfico 150** – Função de Probabilidade – Item 112

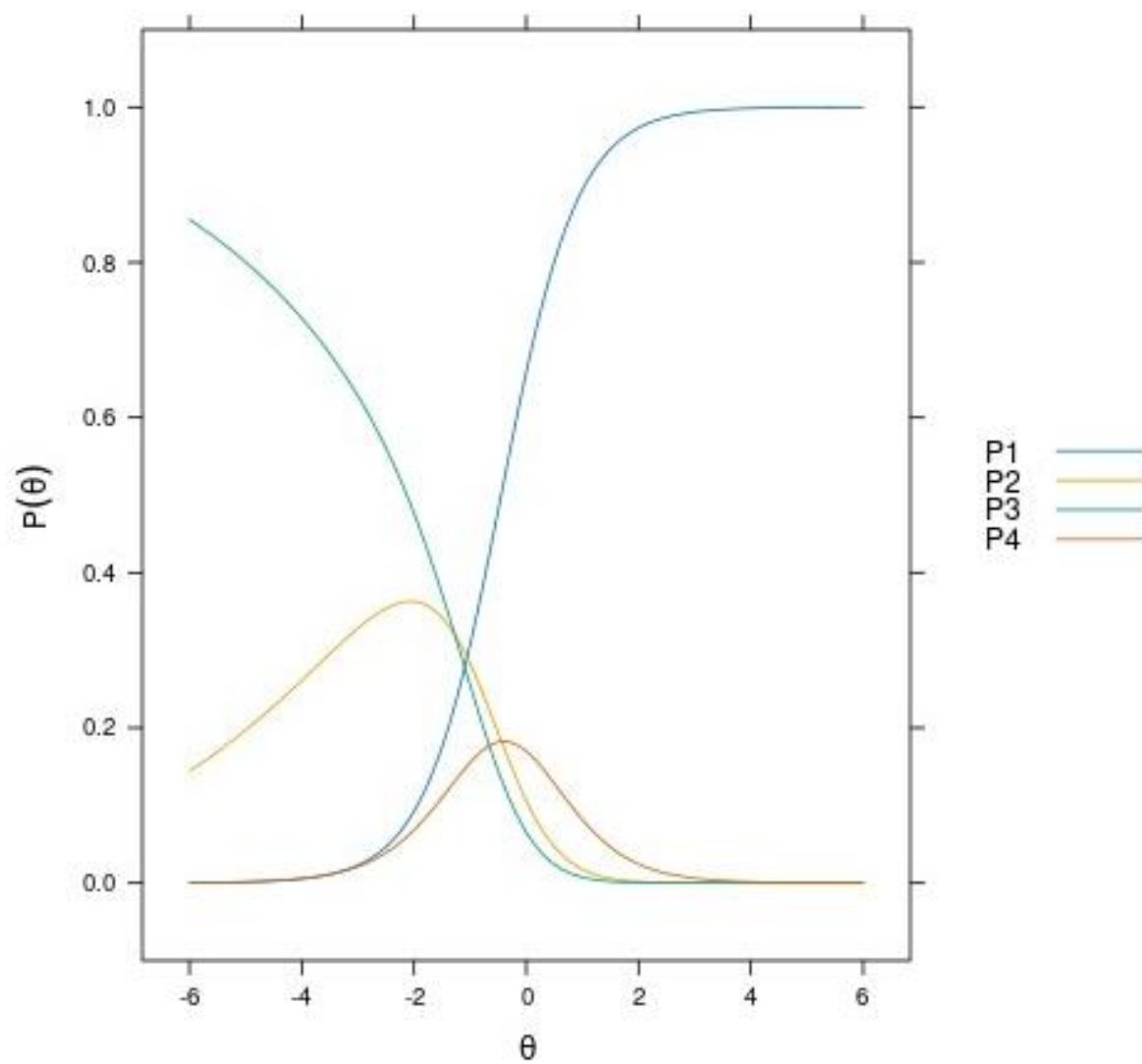
**Fonte:** Elaborado própria

**Gráfico 160** – Função de Probabilidade – Item 113

**Fonte:** Elaborado própria

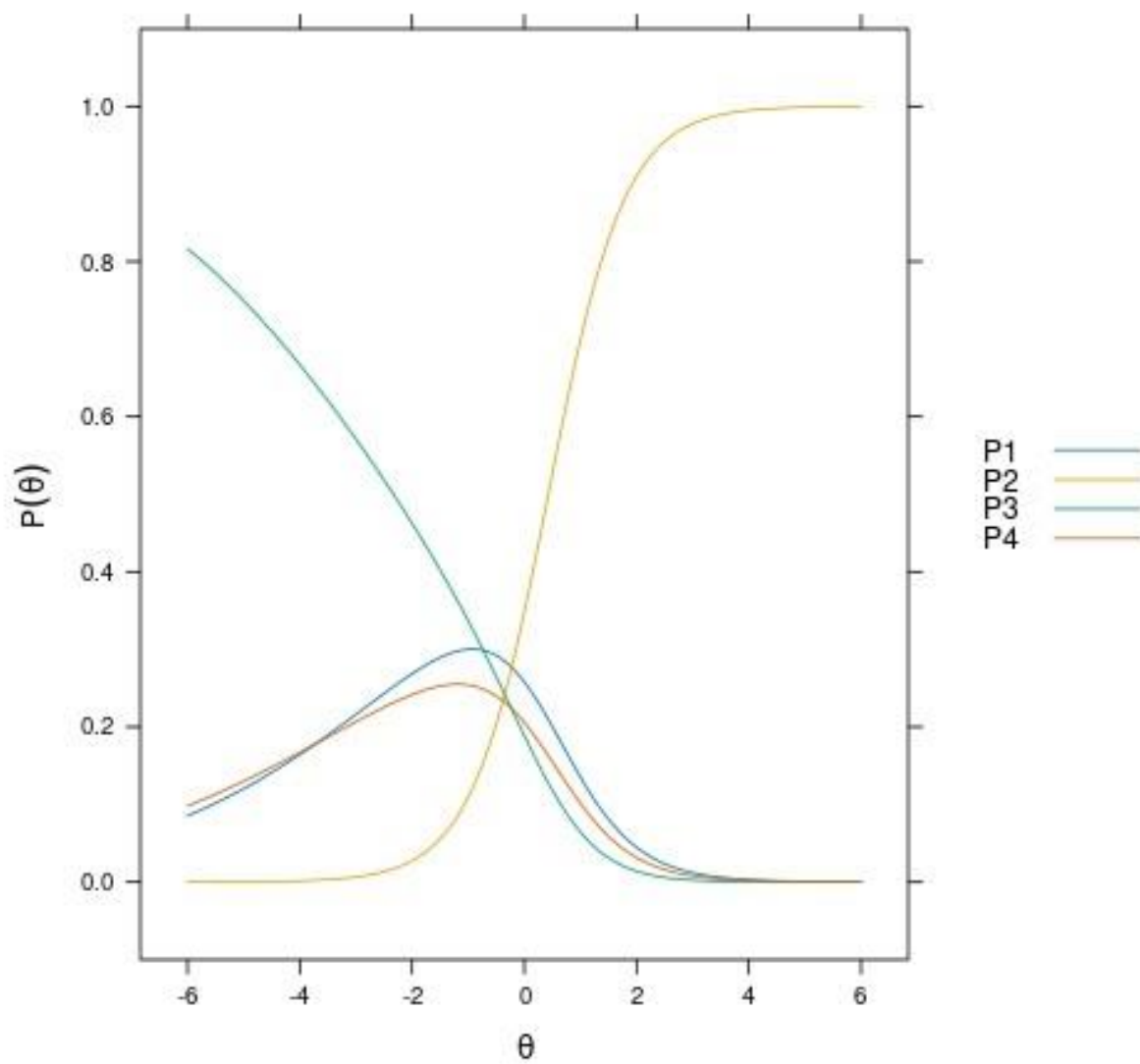
**Gráfico 161** – Função de Probabilidade – Item 114

**Fonte:** Elaborado própria

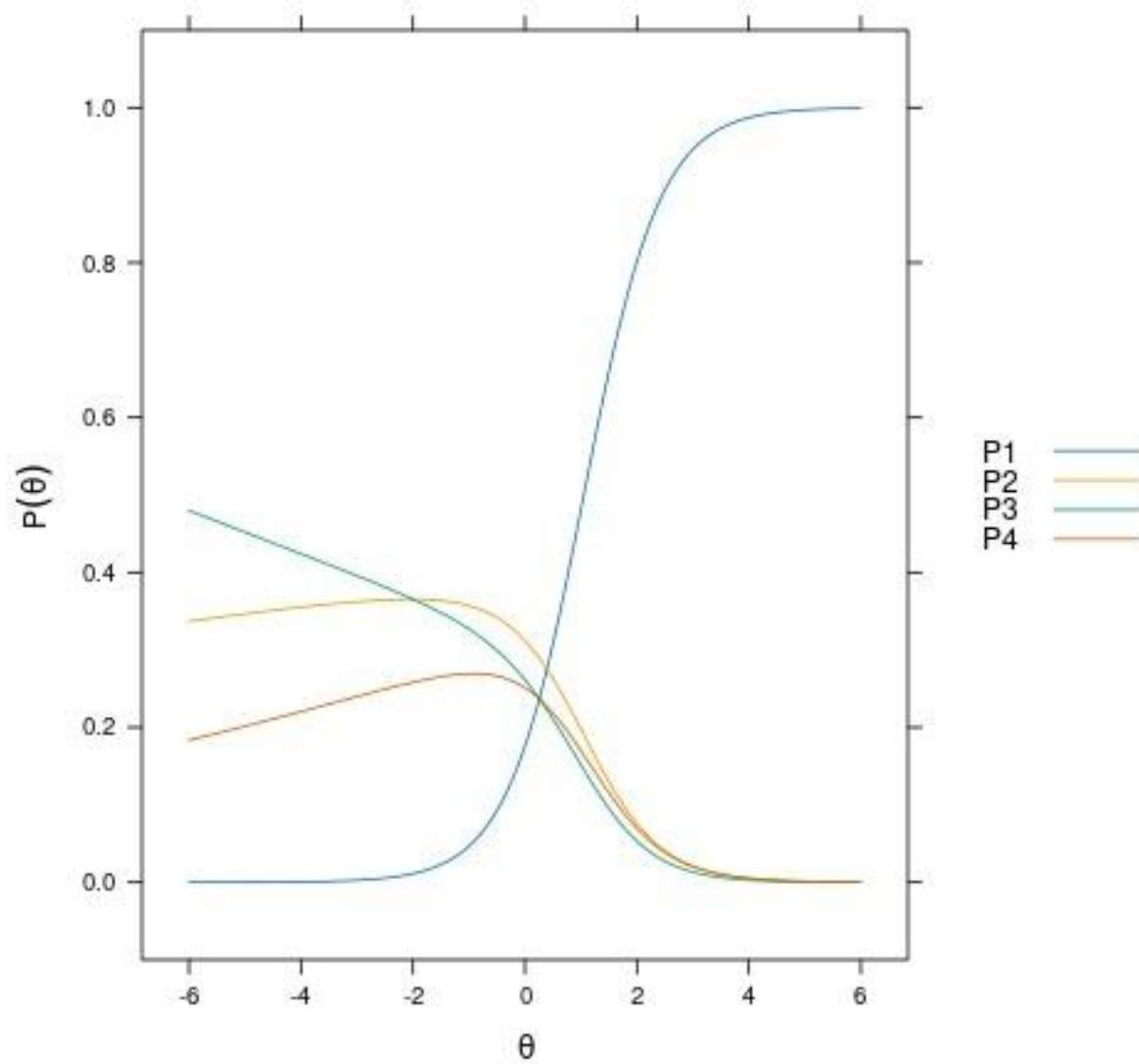
**Gráfico 162** – Função de Probabilidade – Item 115

**Fonte:** Elaborado própria

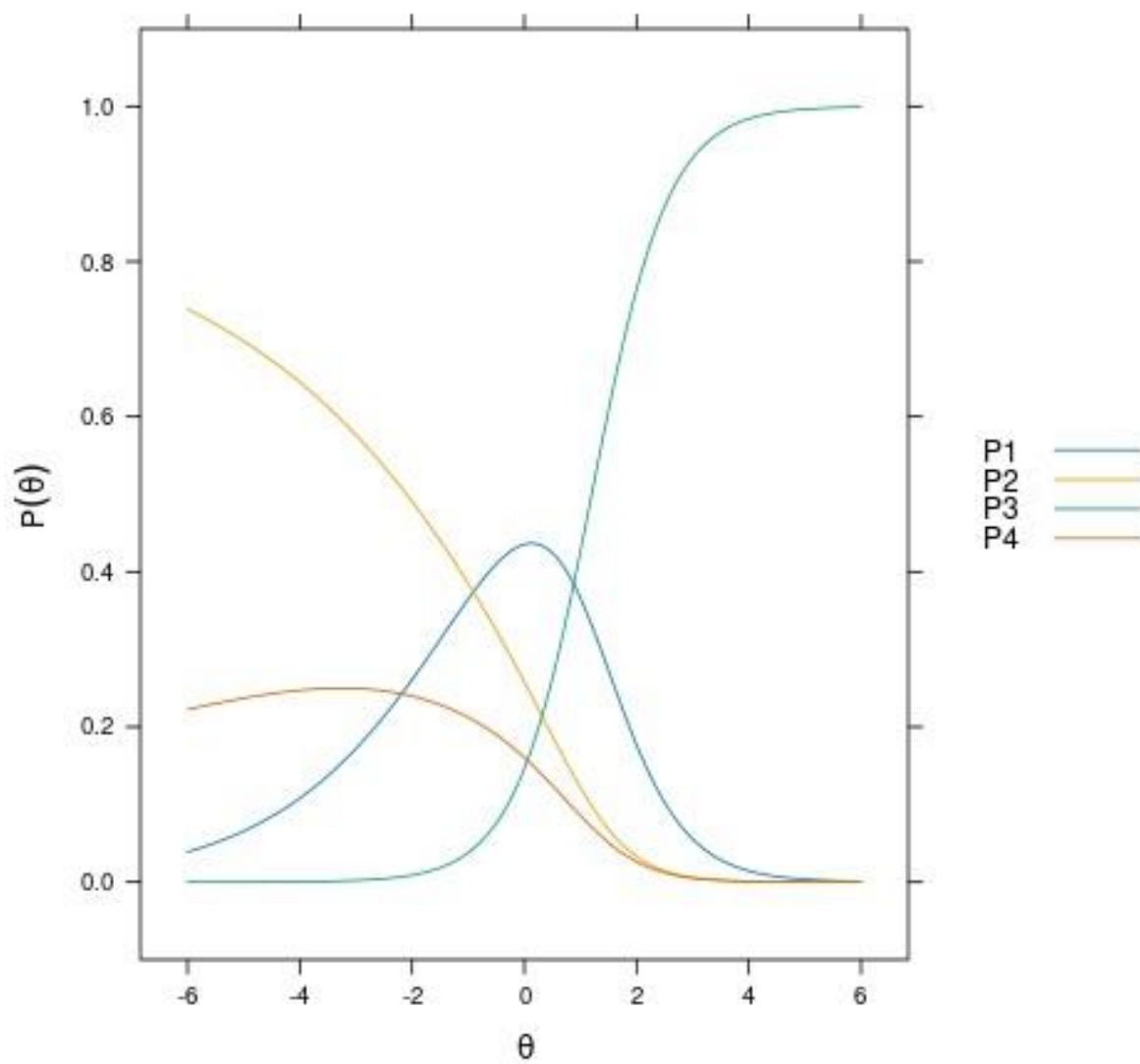


**Gráfico 163** – Função de Probabilidade – Item 116

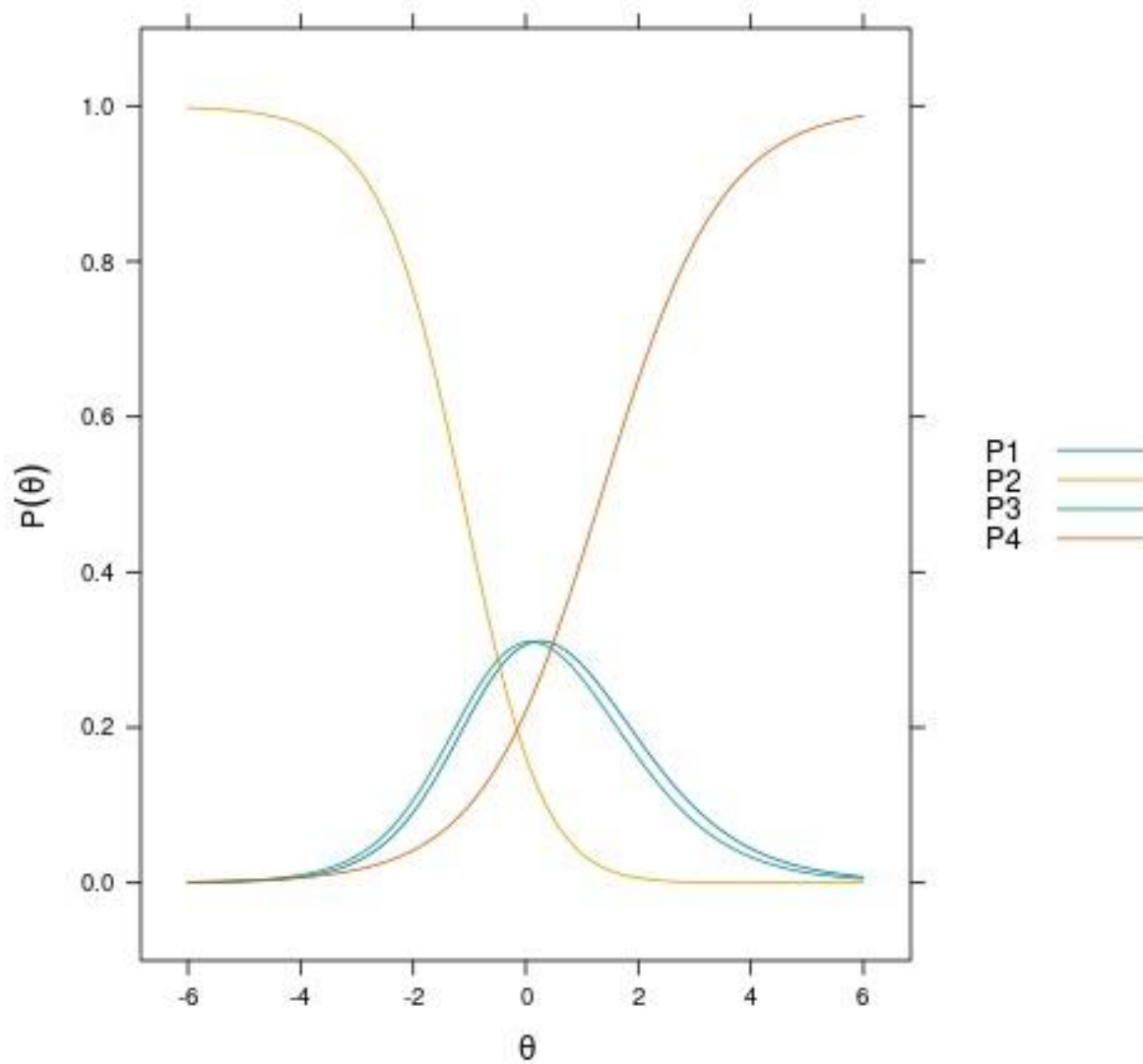
**Fonte:** Elaborado própria

**Gráfico 164** – Função de Probabilidade – Item 117

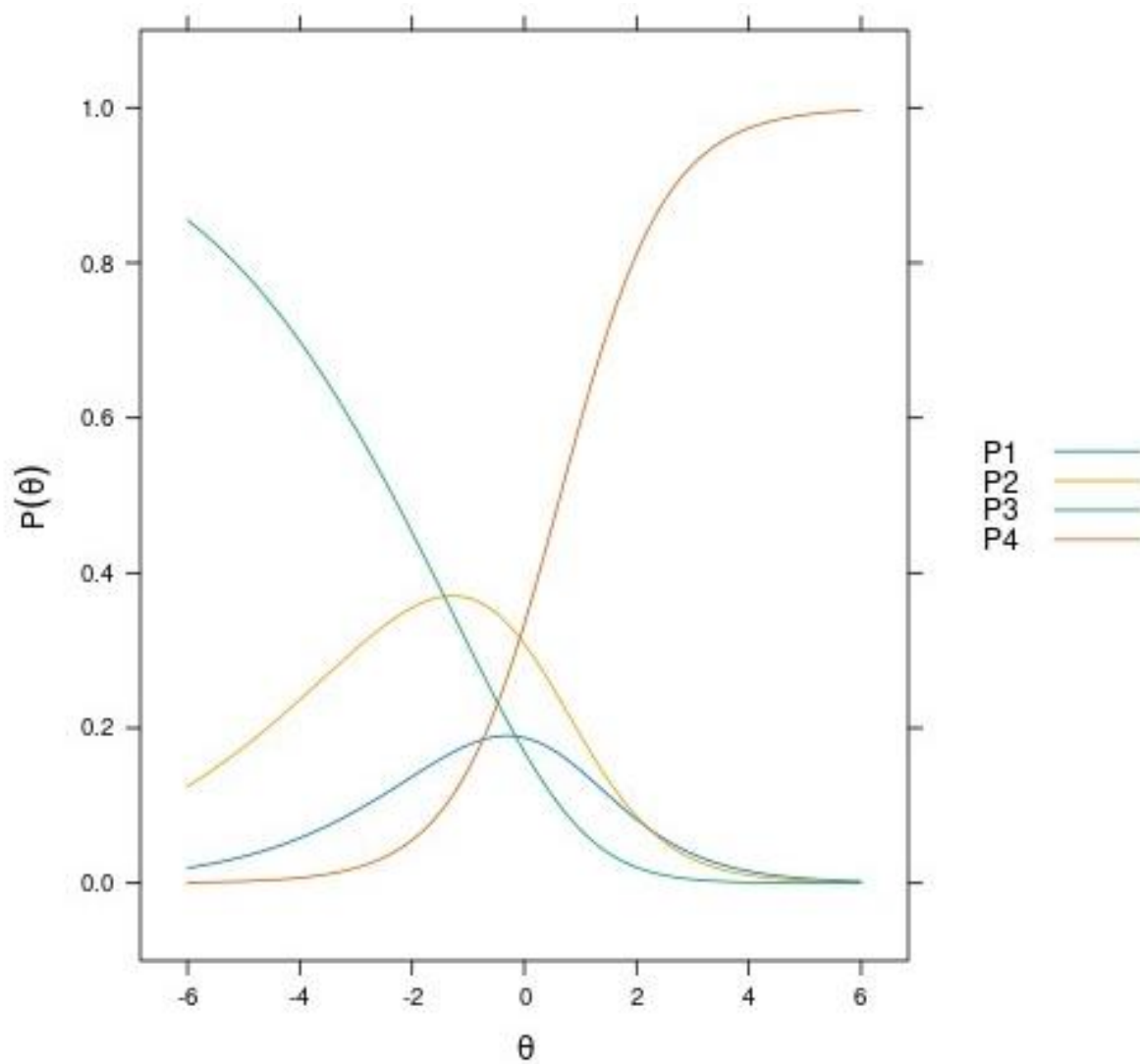
**Fonte:** Elaborado própria

**Gráfico 165** – Função de Probabilidade – Item 118

**Fonte:** Elaborado própria

**Gráfico 166** – Função de Probabilidade – Item 119

**Fonte:** Elaborado própria

**Gráfico 167** – Função de Probabilidade – Item 120

Fonte: Elaborado própria