

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Faculdade de Filosofia e Ciências Humanas
Programa de Pós-Graduação em Psicologia: Cognição e Comportamento

Kellyane Madureira Figueiredo

**INVESTIGAÇÃO DAS PROPRIEDADES PSICOMÉTRICAS E EXPLORAÇÃO DE
MODELOS DA INTELIGÊNCIA NOS TESTES DE SELEÇÃO PROFISSIONAL
(BATERIA TSP)**

Belo Horizonte

2024

Kellyane Madureira Figueiredo

**INVESTIGAÇÃO DAS PROPRIEDADES PSICOMÉTRICAS E EXPLORAÇÃO DE
MODELOS DA INTELIGÊNCIA NOS TESTES DE SELEÇÃO PROFISSIONAL
(BATERIA TSP)**

Tese apresentada ao Programa de Pós-graduação em Psicologia: Cognição e Comportamento, da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Doutorado em Psicologia.

Orientadora: Profa. Dra. Elizabeth do Nascimento.

Belo Horizonte

2024

153.4 Figueiredo, Kellyane Madureira.
F475i Investição das propriedades psicométricas e exploração
2024 de modelos da inteligência nos Testes de Seleção Profissional
(bateria TSP) [manuscrito] / Kellyane Madureira Figueiredo. -
2024.
178 f.
Orientadora: Elizabeth do Nascimento.

Tese (doutorado) - Universidade Federal de Minas Gerais,
Faculdade de Filosofia e Ciências Humanas.
Inclui bibliografia.

1. Psicologia – Teses. 2. Inteligência – Teses.
3. Psicometria - Teses. 4. Cognição – Teses. I. Nascimento,
Elizabeth do. II. Universidade Federal de Minas Gerais.
Faculdade de Filosofia e Ciências Humanas. III. Título.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FAFICH - PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA COGNIÇÃO E COMPORTAMENTO

ATA DE DEFESA DE TESE ALUNA KELLYANE MADUREIRA FIGUEIREDO

Realizou-se, no dia 26 de abril de 2024, às 14:30 horas, através de videoconferência, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada Investigação das propriedades psicométricas e exploração de modelos da inteligência nos Testes de Seleção Profissional (Bateria TSP), apresentada por **KELLYANE MADUREIRA FIGUEIREDO**, número de registro 2020679196, graduada no curso de PSICOLOGIA, como requisito parcial para a obtenção do grau de Doutor em PSICOLOGIA: COGNIÇÃO E COMPORTAMENTO, à seguinte Comissão Examinadora: Prof(a). Elizabeth do Nascimento - Orientador (FAFICH-UFMG), Prof(a). Ana Paula Porto Noronha (Universidade São Francisco), Prof(a). Rebecca de Magalhães Monteiro (Pontifícia Universidade Católica de Minas Gerais), Prof(a). Cristiano Mauro Assis Gomes (UFMG), Prof(a). Sílvia Miranda Amorim (UFMG).

A Comissão considerou a tese:

Aprovada Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 29 de abril de 2024.

Prof(a). Elizabeth do Nascimento (Doutora)

Prof(a). Ana Paula Porto Noronha (Doutora)

Prof(a). Rebecca de Magalhães Monteiro (Doutora)

Prof(a). Cristiano Mauro Assis Gomes (Doutor)

Prof(a). Sílvia Miranda Amorim (Doutora)



Documento assinado eletronicamente por **Elizabeth do Nascimento, Membro de comissão**, em 29/04/2024, às 19:35, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ana Paula Porto Noronha, Usuário Externo**, em 30/04/2024, às 09:46, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Silvia Miranda Amorim, Professora do Magistério Superior**, em 30/04/2024, às 10:09, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiano Mauro Assis Gomes, Professor do Magistério Superior**, em 30/04/2024, às 15:57, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rebecca de Magalhães Monteiro Lopes, Usuária Externa**, em 09/04/2025, às 15:35, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3212751** e o código CRC **15763654**.

AGRADECIMENTOS

Durante todo o percurso dedicado ao doutorado, contei com o apoio fundamental de pessoas cuja presença foi indispensável para alcançar esse objetivo. Primeiramente, agradeço à minha família, que ao longo desses mais de dez anos de estudo, tem me apoiado, compreendido minhas frequentes ausências e sempre oferecido palavras de incentivo. À minha companheira Luana pelo apoio incondicional e segurança emocional proporcionada nos momentos mais difíceis. Ao meu pai (em memória), cujo pensamento jamais poderia abranger essa conquista, mas que agora tem uma filha doutora. Aos amigos Ariany, Renata e Santhiago pelos muitos momentos de reflexão e crítica sobre o que permeia a vida acadêmica e, principalmente, pelo companheirismo e cumplicidade. À Fernanda Franco, minha gratidão como uma grande amiga, mas também uma eterna mestra, cuja sabedoria e trocas têm sido importantes para minha jornada.

À professora Elizabeth do Nascimento, agradeço por todos os anos de convivência e apoio. Pelo convite para participar da presente pesquisa e por ter acreditado no meu desenvolvimento como pesquisadora. É com profunda gratidão que reconheço a oportunidade ímpar de contribuir com um primeiro teste, assinado por mim, favorável pelo SATEPSI/CFP. Devo à você muitas oportunidades de trabalho e de projetos futuros na pesquisa com métodos quantitativos em Psicologia.

Agradeço à equipe de coleta de dados pelo trabalho intenso, dedicação e responsabilidade: Ariadina Cibele de O. Rocha, Camila Cristina F. Bicalho, Carlos Irineu G. Costa, Caroline de Moraes, Damares Diniz Teixeira, Dayane Camila L. da Silva, Débora S. D. Silva, Demerval Guiçarducci Bruzzi, Dione Hubner Gonçalves, Francisca Talitta M. Saboya, Luana Dutra Santiago, Maralins Lopes Rezende, Michele Ferreira da Rocha, Natane Monique G. Machado, Santhiago Bonfim de Souza e Wedja J. Granja Costa. À todas as instituições de ensino que autorizaram a coleta de dados, meus sinceros agradecimentos.

Ao Bruno Damásio, expressei meu sincero reconhecimento por criar a plataforma “Psicometria Online”, uma ferramenta essencial para estudos e aprimoramento nesta área. Em meio à jornada solitária da pós-graduação, os cursos e as interações nas caixas de perguntas abertas no *Instagram* desempenharam um papel fundamental no fortalecimento do meu conhecimento em análise de dados.

À Fundação Coordenação Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa de doutorado concedida. À Edites pela confiança no trabalho e apoio em todas as propostas para ampliação da coleta de dados.

*“Cuida com exatidão da perpendicular
e das paralelas perfeitas.
Com apurado rigor.*

(...)

*Todos os dias estarás refazendo o teu desenho.
Não te fatigues logo. Tens trabalho para toda a vida.
E nem para o teu sepulcro terás a medida certa.”*

Cecília Meireles

RESUMO

O estudo teve como objetivo principal investigar as propriedades psicométricas dos Testes de Seleção Profissional (Bateria TSP), um instrumento amplamente utilizado em avaliações psicológicas no contexto organizacional e do trabalho. Participaram 1606 indivíduos, residentes nos estados de Minas Gerais, Espírito Santo, Ceará e Distrito Federal. A amostra contemplou, portanto, as regiões Sudeste (n=1119), Nordeste (n=172) e Centro-oeste (n=312), três respondentes não informaram o local de residência. Os participantes foram homens e mulheres, com idades entre 15 e 78 anos (M=27,17; DP=11,91), com níveis de escolaridade do fundamental ao superior – incompleto e completo, com diferentes ocupações profissionais. Primeiramente, foram realizadas estatísticas descritivas para caracterizar os resultados para cada instrumento. Posteriormente, os dados foram analisados pela Teoria Clássica dos Teste (TCT) e pelos parâmetros individuais dos itens pela Teoria de Resposta ao Item (TRI). Em seguida, buscou-se evidências de validade, tanto internas quanto com sua relação com variáveis externas, como critérios de desempenho e outras medidas de inteligência convergentes. Além disso, investigou-se as estimativas de precisão do testes da Bateria TSP por meio da análise da consistência interna e estabilidade temporal. Por fim, houve a investigação do modelo de estrutura da inteligência da Bateria TSP conduzida por análises fatoriais confirmatórias para testar modelos baseados em teorias já preestabelecidas e modelos alternativos. Em geral, os resultados demonstram adequadas evidências de validade e precisão para os testes da Bateria TSP. As análises fatoriais confirmatórias e pela TRI demonstram que alguns itens de cada instrumento apresentaram coeficientes inadequados, indicando a necessidade de modificá-los ou removê-los em futuras atualizações do teste. Os resultados da TRI demonstraram que todos testes indicam melhor funcionamento para sujeitos que têm um nível de *theta* baixo à médio. Os resultados dos modelos teóricos e alternativos da inteligência apresentaram melhores ajustes para os modelos que consideraram os testes isoladamente ou agrupados em função do tipo de estímulo (verbal e não verbal). Considerando os resultados das AFC a proposição de um escore geral a partir da aplicação da Bateria TSP parece ser empiricamente viável. Recomenda-se que pesquisas futuras dos testes considerem a inclusão de estímulos mais desafiadores, a fim de oferecer uma variedade adequada de dificuldade aos participantes. Estudos futuros são relevantes para o aprimoramento da Bateria TSP, sugere-se análises de invariância em subamostras divididas por escolaridade, idade e sexo, a fim de averiguar a estabilidade dos fatores por diferentes grupos amostrais, principalmente com o controle da variável escolaridade. É fundamental destacar que, no Brasil, o instrumento ainda não dispõe de aplicação e correção informatizada, o que ressalta a importância de estudos para viabilizar essa possibilidade.

Palavras-chave: Inteligência; Psicometria; Teste psicológico; Avaliação psicológica da cognição; Testes de Seleção Profissional (Bateria TSP).

ABSTRACT

The main objective of this study was to investigate the psychometric properties of the Professional Selection Tests (TSP Battery), an instrument widely used in psychological assessments within organizational and work contexts. A total of 1.606 individuals participated, residing in the states of Minas Gerais, Espírito Santo, Ceará, and the Federal District. The sample, therefore, included the Southeast ($n = 1.119$), Northeast ($n = 172$), and Central-West ($n = 312$) regions, with three respondents not reporting their place of residence. Participants included both men and women, aged between 15 and 78 years ($M = 27.17$; $SD = 11.91$), with educational backgrounds ranging from elementary to higher education - both incomplete and complete - and representing various professional occupations. Initially, descriptive statistics were conducted to characterize the results for each instrument. Subsequently, the data were analyzed using Classical Test Theory (CTT) and the individual item parameters were examined using Item Response Theory (IRT). Evidence of validity was then sought, both internal and in relation to external variables such as performance criteria and other convergent measures of intelligence. Furthermore, the precision estimates of the TSP Battery tests were examined through analyses of internal consistency and temporal stability. Finally, the structural model of intelligence underlying the TSP Battery was investigated using confirmatory factor analyses (CFA), testing both theoretically grounded and alternative models. Overall, the results provided adequate evidence of validity and reliability for the TSP Battery tests. The CFA and IRT analyses revealed that some items within each instrument had inadequate coefficients, suggesting the need for modification or removal in future test revisions. IRT results also indicated that the tests functioned better for individuals with low to medium theta levels. The results for both theoretical and alternative intelligence models showed better fit for models that considered tests either individually or grouped by stimulus type (verbal and non-verbal). Based on the CFA results, proposing a general score derived from the application of the TSP Battery appears to be empirically viable. It is recommended that future studies consider including more challenging stimuli to ensure an appropriate range of difficulty for participants. Further research is important for improving the TSP Battery; invariance analyses are suggested for subsamples divided by education level, age, and gender to verify the stability of factors across different groups—especially with regard to the education variable. It is also essential to note that, in Brazil, the instrument has not yet been adapted for computerized administration and scoring, underscoring the need for research to make this possible.

Keywords: Intelligence; Psychometrics; Psychological test; Psychological assessment of cognition; Professional Selection Tests (TSP Battery).

LISTA DE FIGURAS

<i>Figura 1.</i> Estrutura do Modelo 1 hipotético.....	59
<i>Figura 2.</i> Estrutura do Modelo 2 hipotético.....	59
<i>Figura 3.</i> Estrutura do Modelo 3 hipotético.....	60
<i>Figura 4.</i> Estrutura do Modelo 4 hipotético.....	60
<i>Figura 5.</i> Estrutura do Modelo 5 hipotético.....	60
<i>Figura 6.</i> Estrutura do Modelo 6 hipotético.....	61
<i>Figura 7.</i> Estrutura do Modelo 7 hipotético.....	61
<i>Figura 8.</i> Estrutura do Modelo 8 hipotético.....	61
<i>Figura 9.</i> Diagrama quantil-quantil do teste Fluência.....	66
<i>Figura 10.</i> Mapa Térmico de Correlações.....	67
<i>Figura 11.</i> Diagrama Q-Q teste Números.....	72
<i>Figura 12.</i> Curva de Gauss do teste Números.....	72
<i>Figura 13.</i> Curvas Características dos Itens (CCIs) do teste Números.....	75
<i>Figura 14.</i> Curva de Informação do Teste (CIT) e erro-padrão	75
<i>Figura 15.</i> Diagrama Q-Q teste Julgamento.....	82
<i>Figura 16.</i> Curva de Gauss do teste Julgamento.....	82
<i>Figura 17.</i> Curvas Características dos Itens (CCI's) do teste Julgamento.....	84
<i>Figura 18.</i> Curva de Informação do Teste (CIT) e erro-padrão.....	85
<i>Figura 19.</i> Diagrama Q-Q teste Percepção.....	91
<i>Figura 20.</i> Curva de Gauss do teste Percepção.....	91
<i>Figura 21.</i> Curvas Características dos Itens (CCIs) do teste Percepção.....	94
<i>Figura 22.</i> Curva de Informação do Teste (CIT) e erro-padrão.....	94
<i>Figura 23.</i> Diagrama Q-Q teste Precisão.....	100
<i>Figura 24.</i> Curva de Gauss do teste Precisão.....	100
<i>Figura 25.</i> Curvas Características dos Itens (CCIs) do teste Precisão.....	102
<i>Figura 26.</i> Curva de Informação do Teste (CIT) e erro-padrão.....	102
<i>Figura 27.</i> Diagrama Q-Q teste Partes.....	107
<i>Figura 28.</i> Curva de Gauss do teste Partes.....	107
<i>Figura 29.</i> Curvas Características dos Itens (CCIs) do teste Partes.....	110
<i>Figura 30.</i> Curva de Informação do Teste (CIT) e erro-padrão.....	110
<i>Figura 31.</i> Diagrama Q-Q teste Dimensão.....	116
<i>Figura 32.</i> Curva de Gauss do teste Dimensão.....	116

<i>Figura 33.</i> Curvas Características dos Itens (CCIs) do teste Dimensão.....	118
<i>Figura 34.</i> Curva de Informação do Teste (CIT) e erro-padrão.....	119
<i>Figura 35.</i> Diagrama Q-Q teste Blocos.....	125
<i>Figura 36.</i> Curva de Gauss do teste Blocos.....	125
<i>Figura 37.</i> Curvas Características dos Itens (CCIs) do teste Blocos.....	127
<i>Figura 38.</i> Curva de Informação do Teste (CIT) e erro-padrão.....	127
<i>Figura 39.</i> Diagrama Q-Q teste Memória.....	133
<i>Figura 40.</i> Curvas Características dos Itens (CCI's) do teste Memória.....	135
<i>Figura 41.</i> Curva de Informação do Teste (CIT) e erro-padrão.....	135
<i>Figura 42.</i> Estrutura fatorial do modelo 5.....	149
<i>Figura 43.</i> Estrutura fatorial do modelo 6.....	149
<i>Figura 44.</i> Estrutura fatorial do modelo 7.....	150
<i>Figura 45.</i> Estrutura fatorial do modelo 8.....	150

LISTA DE TABELAS

Tabela 1. Descrição das habilidades amplas e específicas do modelo CHC de inteligência..	24
Tabela 2. Testes, construtos avaliados e tipo de estímulo presente na Bateria TSP.....	33
Tabela 3. Caracterização sociodemográfica da amostra (N=1599).....	49
Tabela 4. Síntese dos procedimentos de análise de dados.....	59
Tabela 5. Frequência de sujeitos por teste da Bateria TSP.....	63
Tabela 6. Estatísticas descritivas do teste Fluência.....	65
Tabela 7. Cargas fatoriais e índices de ajustes do teste Fluência da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	66
Tabela 8. Desempenho dos participantes no teste Fluência de acordo com o sexo.....	68
Tabela 9. Desempenho dos participantes no teste Fluência segundo a faixa etária.....	68
Tabela 10. Desempenho dos participantes no teste Fluência segundo o nível de escolaridade.....	69
Tabela 11. Matriz de Correlações entre as variáveis idade e escolaridade e o teste Fluência.....	70
Tabela 12. Estatísticas do estudo de estabilidade temporal do teste Fluência (n=78).....	70
Tabela 13. Estatística descritiva do teste Números.....	71
Tabela 14. Parâmetros individuais dos itens do teste Números (N=1255).....	73
Tabela 15. Cargas fatoriais e índices de ajustes do teste Números da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	76
Tabela 16. Desempenho dos participantes no teste Números de acordo com o sexo.....	78
Tabela 17. Desempenho dos participantes no teste Números segundo a faixa etária.....	79
Tabela 18. Desempenho dos participantes no teste Números segundo o nível de escolaridade.....	79
Tabela 19. Estatísticas do estudo de estabilidade temporal do teste Números (N=81).....	80
Tabela 20. Estatística descritiva do teste Julgamento.....	81
Tabela 21. Parâmetros individuais dos itens do teste Julgamento (N=1290).....	82
Tabela 22. Cargas fatoriais e índices de ajustes do teste Julgamento da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	86
Tabela 23. Desempenho dos participantes no teste Julgamento de acordo com o sexo.....	88
Tabela 24. Desempenho dos participantes no teste Julgamento segundo a faixa etária.....	88
Tabela 25. Desempenho da amostra no teste Julgamento segundo o nível de escolaridade.....	89

Tabela 26. Estatísticas do estudo de estabilidade temporal do teste Julgamento (N=87)	89
Tabela 27. Estatísticas descritivas do teste Percepção.....	90
Tabela 28. Parâmetros individuais dos itens do teste Percepção (N=1270)	92
Tabela 29. Cargas fatoriais e índices de ajustes do teste Percepção da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	95
Tabela 30. Desempenho dos participantes no teste Percepção de acordo com o sexo.....	96
Tabela 31. Desempenho dos participantes no teste Percepção segundo a faixa etária.....	97
Tabela 32. Desempenho dos participantes no teste Percepção segundo o nível de escolaridade.....	97
Tabela 33. Estatísticas do estudo de estabilidade temporal do teste Percepção (n=78).....	98
Tabela 34. Estatísticas descritivas do teste Precisão.....	99
Tabela 35. Parâmetros individuais dos itens do teste Precisão (N=1202).....	100
Tabela 36. Cargas fatoriais e índices de ajustes do teste Precisão da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	103
Tabela 37. Desempenho dos participantes no teste Precisão de acordo com o sexo.....	104
Tabela 38. Desempenho dos participantes no teste Precisão segundo a faixa etária.....	105
Tabela 39. Desempenho dos participantes no teste Precisão segundo o nível de escolaridade.....	105
Tabela 40. Estatísticas do estudo de estabilidade temporal do teste Precisão (n=82).....	106
Tabela 41. Estatísticas descritivas do teste Partes.....	107
Tabela 42. Parâmetros individuais dos itens do teste Partes (N=1268).....	108
Tabela 43. Cargas fatoriais e índices de ajustes do teste Partes da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	111
Tabela 44. Desempenho dos participantes no teste Partes de acordo com o sexo.....	113
Tabela 45. Desempenho dos participantes no teste Partes segundo a faixa etária.....	113
Tabela 46. Desempenho dos participantes no teste Partes segundo o nível de escolaridade.....	114
Tabela 47. Estatísticas do estudo de estabilidade temporal do teste Partes (n=90)	114
Tabela 48. Estatísticas descritivas do teste Dimensão.....	115
Tabela 49. Parâmetros individuais dos itens do teste Dimensão (N=1236).....	117
Tabela 50. Cargas fatoriais e índices de ajustes do teste Dimensão da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	119
Tabela 51. Matriz de correlações entre o teste Dimensão, G-38 e provas do BPR-5.....	121
Tabela 52. Desempenho dos participantes no teste Dimensão o de acordo com o sexo.....	122

Tabela 53. Desempenho dos participantes no teste Dimensão segundo a faixa etária.....	122
Tabela 54. Desempenho dos participantes no teste Dimensão segundo o nível de escolaridade.....	123
Tabela 55. Estatísticas do estudo de estabilidade temporal do teste Dimensão (n=95)	123
Tabela 56. Estatísticas descritivas do teste Blocos.....	124
Tabela 57. Parâmetros individuais dos itens do teste Blocos (N=1276)	125
Tabela 58. Cargas fatoriais e índices de ajustes do teste Blocos da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	128
Tabela 59. Matriz de correlações entre o teste Blocos, G-38 e provas de raciocínio (BPR-5)	130
Tabela 60. Desempenho dos participantes no teste Blocos de acordo com o sexo.....	130
Tabela 61. Desempenho dos participantes no teste Blocos segundo a faixa etária.....	131
Tabela 62. Desempenho dos participantes no teste Blocos segundo o nível de escolaridade.....	131
Tabela 63. Estatísticas do estudo de estabilidade temporal do teste Blocos (n=92).....	131
Tabela 64. Estatísticas descritivas do teste Memória.....	133
Tabela 65. Parâmetros individuais dos itens do teste Memória (N=1144).....	133
Tabela 66. Cargas fatoriais e índices de ajustes do teste Memória da Bateria TSP de acordo com a Análise Fatorial Confirmatória.....	136
Tabela 67. Desempenho dos participantes no teste Memória de acordo com o sexo.....	137
Tabela 68. Desempenho dos participantes no teste Memória segundo a faixa etária.....	138
Tabela 69. Desempenho dos participantes no teste Memória segundo o nível de escolaridade.....	138
Tabela 70. Estatísticas do estudo de estabilidade temporal do teste Memória (n=86)	139
Tabela 71. Matriz de correlações entre os testes da Bateria TSP.....	140
Tabela 72. Itens selecionados para AFC.....	143
Tabela 73. Ajuste dos modelos hipotéticos testados (AFC de itens)	144
Tabela 74. Confiabilidade dos fatores para os modelos da inteligência testados.....	145
Tabela 75. Estatísticas descritivas das cargas fatoriais.....	146
Tabela 76. Correlações entre os fatores e as cargas dos fatores de primeira ordem nos fatores de segunda ordem.....	147
Tabela 77. Ajuste dos modelos hipotéticos testados (AFC por escores somados).....	148
Tabela 78. Confiabilidade e estatística descritivas dos fatores.....	149

LISTA DE SIGLAS, ABREVIATURAS E SÍMBOLOS

AERA	<i>American Educational Research Association</i>
AFC	Análise Fatorial Confirmatória
APA	<i>American Psychological Association</i>
BaFaCaLo	Bateria de Fatores Cognitivos de Alta-Ordem
BPR-5	Bateria de Provas de Raciocínio
CCIs	Curvas Características dos Itens
CFI	<i>Comparative Fit Index</i>
CFP	Conselho Federal de Psicologia
CHC	Cattell-Horn-Carroll
CIT	Curva de Informação do Teste
COEP	Comitê de Ética em Pesquisa
COVID-19	Coronavirus Disease 2019
DP	Desvio Padrão
EFA	<i>Exploratory Factor Analysis</i>
EJA	Educação de Jovens e Adultos
GFI	<i>Goodness-of-fit Index</i>
gl	Graus de liberdade
IC	Intervalo de confiança
LADI	Laboratório de Avaliação das Diferenças Individuais
MIRT	<i>Multidimensional Item Response Theory</i>
MLE	<i>Maximum Likelihood Estimation</i>
NCME	<i>National Council on Measurement in Education</i>
Q-Q	Quantil-Quantil
RMSEA	<i>Root Mean Square Error of Approximation</i>
SRMR	<i>Standardized Root Mean Square Residual</i>
TALE	Termo de Assentimento Livre e Esclarecido
TCLE	Termos de Consentimento Livre e Esclarecido
TLI	<i>Tucker Lewis Index</i>
TRI	Teoria de Resposta ao Item
TSP	Teste de Seleção Profissional
UFMG	Universidade Federal de Minas Gerais

WISC-V	Escala de Inteligência Wechsler para Crianças - quinta edição
WLSMV	<i>Weighted Least Squares Means and Variance Adjusted</i>
α	alfa de Cronbach
c.c.	Confiabilidade composta
p	Probabilidade
Ω	ômega de McDonald
η^2	Tamanho do efeito
χ^2	Qui-quadrado

SUMÁRIO

1. INTRODUÇÃO	20
1.1 O modelo bifatorial de Spearman	20
1.2 O modelo das aptidões primárias de Thurstone	21
1.3 O modelo hierárquico de Vernon	21
1.4 O modelo VPR de Johnson e Bouchard.....	22
1.5 O modelo Gf-Gc estendido de Cattell-Horn.....	22
1.6 O modelo dos três estratos de Carroll	23
1.7 O modelo Cattell-Horn-Carroll (CHC)	24
1.8 A prática da avaliação psicológica da inteligência	28
2. A BATERIA TSP	30
2.1 Teste Fluência	36
2.2 Teste Números.....	37
2.3 Teste Julgamento	39
2.4 Teste Percepção.....	41
2.5 Teste Precisão.....	42
2.6 Teste Partes	43
2.7 Teste Dimensão	45
2.8 Teste Blocos	46
2.9 Teste Memória	47
3. OBJETIVOS.....	48
3.1 Objetivo geral.....	48
3.2 Objetivos específicos.....	48
4. MÉTODO	49
4.1 Amostra geral.....	49
4.2 Instrumentos para a coleta de dados	50
4.3 Procedimentos para a coleta de dados.....	54
4.4 Procedimentos para a análise dos dados	56
4.5 Modelos teóricos e alternativos da inteligência investigados.....	61
4.6 Estratégias analíticas para investigação da estrutura de inteligência da Bateria TSP.....	64

5. RESULTADOS E DISCUSSÃO DAS PROPRIEDADES PSICOMÉTRICAS DOS TESTES DA BATERIA TSP	65
5.1 Teste Fluência	65
5.1.1 Evidências de validade baseadas na estrutura interna para o teste Fluência.....	67
5.1.2 Evidências de validade baseadas nas relações com variáveis externas para o teste Fluência	68
5.1.3 Evidências de confiabilidade do teste Fluência.....	71
5.2 Teste Números.....	72
5.2.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Números	73
5.2.2 Evidências de validade baseadas na estrutura interna para o teste Números	77
5.2.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Números	78
5.2.4 Evidências de confiabilidade do teste Números.....	80
5.3 Teste Julgamento	82
5.3.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Julgamento	83
5.3.2 Evidências de validade baseadas na estrutura interna para o teste Julgamento	86
5.3.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Julgamento	88
5.3.4 Evidências de confiabilidade do teste Julgamento	90
5.4 Teste Percepção.....	91
5.4.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Percepção	92
5.4.2 Evidências de validade baseadas na estrutura interna para o teste Percepção	95
5.4.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Percepção	97
5.4.4 Evidências de confiabilidade do teste Percepção	98
5.5 Teste Precisão.....	99
5.5.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Precisão	101
5.5.2 Evidências de validade baseadas na estrutura interna do teste Precisão	104
5.5.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Precisão	105
5.5.4 Evidências de confiabilidade do teste Precisão.....	106
5.6 Teste Partes	107
5.6.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Partes	109
5.6.2 Evidências de validade baseadas na estrutura interna do teste Partes.....	111

5.6.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Partes	113
5.6.4 Evidências de confiabilidade do teste Partes.....	115
5.7 Teste Dimensão	116
5.7.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Dimensão.....	117
5.7.2 Evidências de validade baseadas na estrutura interna do teste Dimensão	120
5.7.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Dimensão.....	122
5.7.4 Evidências de confiabilidade do teste Dimensão	124
5.8 Teste Blocos	125
5.8.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Blocos.....	126
5.8.2 Evidências de validade baseadas na estrutura interna do teste Blocos	128
5.8.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Blocos.....	130
5.8.4 Evidências de confiabilidade do teste Blocos	132
5.9 Teste Memória	133
5.9.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Memória	134
5.9.2 Evidências de validade baseadas na estrutura interna do teste Memória.....	137
5.9.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Memória	138
5.9.4 Evidências de confiabilidade do teste Memória.....	139
5.10 Discussão geral dos resultados da Bateria TSP	140
6. RESULTADOS E DISCUSSÃO DOS MODELOS TEÓRICOS DE INTELIGÊNCIA TESTADOS PELA BATERIA TSP	143
6.1 Seleção dos itens a serem utilizados nas AFCs.....	143
6.2 Resultados da normalidade multivariada	144
6.3 Resultados das AFCs de itens: ajuste dos modelos, confiabilidade e cargas fatoriais	144
6.4 Resultados das AFCs considerando os escores somados para os modelos 5 a 8: ajuste dos modelos, confiabilidade e cargas fatoriais.....	149
7. CONSIDERAÇÕES FINAIS	153
REFERÊNCIAS	156
Apêndice A – Aprovação do Comitê de ética em Pesquisa.....	167
Apêndice B - Carta de autorização da Instituição	171

Apêndice C – Questionário Geral.....	172
Apêndice D – Termo de Consentimento Livre e Esclarecido 1 - Adultos.....	173
Apêndice E – Termo de Consentimento Livre e Esclarecido 2 – Adolescentes.....	175
Apêndice F – Termo de Assentimento Livre e Esclarecido (TALE)	177

1. INTRODUÇÃO

A história da Psicometria, dos testes e da avaliação psicológica está intimamente ligada e foi fundamental para a consolidação da Psicologia como um campo científico. A abordagem epistemológica do empirismo, que se baseia em dados e observações para a formulação de teorias e modelos, tem desempenhado um papel crucial no processo de investigação das diferenças individuais. Os construtos hipotéticos descritos teoricamente e observados nos processos de avaliação psicológica, são investigados por meio do uso de procedimentos padronizados e reproduzíveis, os testes psicológicos. As análises psicométricas possibilitam a interação entre dados empíricos e teoria, permitindo a compreensão dos construtos, incluindo sua natureza, estrutura, organização e seu papel na explicação de determinados comportamentos e atitudes (Ambiel & Pacanaro, 2019; Golino, 2017; Olea & Abad, 2006).

A inteligência, ou o funcionamento cognitivo em geral, foi um dos primeiros fenômenos psicológicos a serem estudados. Desde as primeiras pesquisas sobre inteligência, diferentes esforços são empregados para compreender seu funcionamento, as aptidões e capacidades relacionadas às diferenças individuais no desempenho cognitivo, bem como sua origem e desenvolvimento ao longo da vida (Andrés-Pueyo, 2006). Essa busca está intrinsecamente relacionada aos progressos alcançados na área da Psicometria e à criação de instrumentos destinados à avaliação psicológica da inteligência. Os testes psicológicos, por sua vez, têm sido elementos fundamentais nas investigações e refinamentos dos diferentes modelos teóricos da inteligência.

Existem importantes pontos de convergência entre as propostas para conceituar e definir a inteligência, embora ainda não se tenha uma definição consensual entre os pesquisadores. De maneira geral, a inteligência está relacionada à capacidade de aprender, adaptar-se ao meio e raciocinar (Andrés-Pueyo, 2006; Muniz & Lins, 2018). A contribuição de vários pesquisadores, como Spearman, Thurstone, Guilford, Vernon, Cattell, Horn, Carroll, resultou na proposição de diversos modelos da inteligência. Esses modelos possuem debates contínuos e demonstram que não são tão antagônicos entre si, conforme poderá ser observado pelas breves descrições dos principais modelos que serão apresentados.

1.1 O modelo bifatorial de Spearman

Os fundamentos iniciais dos modelos psicométricos da inteligência foram estabelecidos por Charles Spearman em 1904. Ele propôs um modelo bifatorial no qual cada desempenho cognitivo é influenciado por dois fatores distintos: um fator geral, nomeado como “g”, que representa uma atividade cognitiva comum a todos os processos mentais, e por habilidades específicas (fator *s*) determinadas pelas exigências exclusivas de cada tarefa. Nesse modelo, o fator geral é considerado como tendo a maior capacidade de explicar a variância no desempenho e possui um papel central na teoria (Laros, Jesus, & Karino, 2013). Em 1927, Spearman identificou a possibilidade de agrupar testes com base em semelhanças no conteúdo, formato ou requisitos de resposta, o que hoje é nomeado, em outros modelos da inteligência, como as capacidades amplas (Wasserman, 2018).

1.2 O modelo das aptidões primárias de Thurstone

Crítico do modelo de Spearman, o psicólogo norte-americano Louis Thurstone, em 1938, apresentou o modelo das aptidões mentais primárias, que divergia da ideia de um fator único de inteligência. O modelo de Thurstone propõe a existência de diversas capacidades e aptidões intelectuais especializadas e independentes (ortogonais). Dessa maneira, ao resolver algum problema ou responder a uma tarefa, como em testes psicológicos, é exigido apenas a aptidão intelectual relacionada à natureza da situação exposta. Esse modelo enfatiza uma estrutura básica de fatores de inteligência que podem atuar de maneira autônoma, aplicáveis a tarefas com diferentes especificidades (Andrés-Pueyo, 2006). Em 1945, Thurstone identificou oito aptidões mentais primárias que correspondem a diferentes tipos de raciocínio: compreensão verbal (V), fluência verbal (W), raciocínio (indutivo) (R), capacidade numérica (N), velocidade perceptual (P), velocidade de julgamento (J), capacidade espacial (S) e memória (M). Em 1947, Thurstone, aceita a existência de um fator geral, no entanto, recomendou o uso dos perfis cognitivos específicos para descrever as variações individuais entre as habilidades, possibilitando um julgamento das facilidades e dificuldades cognitivas (Wasserman, 2018).

1.3 O modelo hierárquico de Vernon

Philip Vernon foi pioneiro na pesquisa de fatores gerais e específicos por meio de um modelo hierárquico da inteligência. Ele propôs quatro estratos, com o fator geral no topo, seguido por dois fatores amplos ortogonais: V:ED, relacionado a habilidades verbais

educacionais, e K:M, relacionado a habilidades espaciais e mecânicas. Esses fatores refletem principalmente conhecimentos adquiridos por meio de experiências educacionais (V:ED) e habilidades desenvolvidas independentemente da formação educacional (K:M). Por sua vez, V:ED e K:M englobam fatores de grupo menores, habilidades como fluência, linguagem, matemática, ciência e habilidades espaciais. Na base do modelo, estão os subfatores específicos medidos por testes psicológicos (Andrés-Pueyo, 2006; Johnson & Bouchard, 2005).

Vernon considerou que o fator V:ED está relacionado com habilidades verbais, raciocínio matemático, cálculo, atenção, entre outros aspectos. Já o fator K:M está associado a habilidades espaciais, mecânicas, coordenação psicomotora, tempo de reação, rapidez perceptual, trabalho manual e várias outras habilidades de natureza técnica (Andrés-Pueyo, 2006; Wasserman, 2018). O modelo de Vernon reconciliou as propostas de Spearman e de Thurstone e forneceu suporte empírico para a dicotomia verbal e não verbal das habilidades cognitivas que se popularizou nas escalas Wechsler de inteligência (Woodcock et al., 2018).

1.4 O modelo VPR de Johnson e Bouchard

Johnson e Bouchard propuseram um modelo hierárquico que se acredita ser sucessor ao modelo de Vernon (Schneider & McGrew, 2018). O modelo mantém uma estrutura de quatro estratos, com o fator geral no topo, mas sugere a inclusão de um outro fator no terceiro estrato, onde estão presentes V:ED e K:M. O novo fator incluído é o de rotação de imagem (ou rotação mental). As habilidades espaciais comumente abrangem dois aspectos distintos: visualização e orientação. O fator de rotação de imagem adicionado ao modelo consiste, principalmente, na habilidade de visualização que engloba a capacidade de girar e manipular mentalmente objetos em duas ou três dimensões. A orientação espacial e a memória foram incluídas no segundo estrato. O grupo de fatores amplos recebeu uma nova nomenclatura: verbal-perceptual-rotação de imagens, e o modelo de Johnson e Bouchard passa a ser conhecido como modelo VPR (Johnson & Bouchard, 2005).

1.5 O modelo Gf-Gc estendido de Cattell-Horn

O modelo inicialmente proposto por Raymond Cattell em 1941 separou o fator geral (g) em dois fatores distintos: inteligência fluída (Gf) e inteligência cristalizada (Gc). A inteligência fluída está associada à resolução de situações por meio de raciocínio dedutivo e

indutivo, sendo mais influenciada pela maturação biológica e mecanismos neurológicos, com pouca influência do ambiente. Cattell percebia Gf como o fator essencial para a aquisição de conhecimento, refletindo a facilidade de raciocínio e a capacidade de adaptação a novas situações. Por outro lado, a inteligência cristalizada é desenvolvida por meio das experiências do indivíduo com o ambiente, principalmente mediada pelo componente verbal. Nesse caso, a resolução de problemas é realizada com base nos conhecimentos adquiridos por meio da cultura e do processo educacional do indivíduo (Brown, 2016; Flanagan & Dixon, 2014; Schelini, 2006; Wasserman, 2018).

Na década de 1960, John Horn foi responsável por expandir o modelo, incluindo ao Gf-Gc outras quatro capacidades cognitivas: processamento visual (Gv), memória de curto prazo (Gsm), armazenamento e recuperação de longo prazo (Glr), velocidade de processamento (Gs). Posteriormente, foram incluídas outras habilidades: processamento auditivo (Ga), conhecimento quantitativo (Gq), leitura e escrita (Grw) e velocidade de decisão correta (CDS) (Horn & Blankson, 2012; Horn, Donaldson & Engstrom, 1981; Horn & Stankov, 1982; Ortiz, 2015). O modelo Cattell-Horn que ficou conhecido como Gf-Gc estendido, foi aprimorado pelos autores por meio de novas pesquisas, resultando em um conjunto de 15 ou 16 capacidades gerais (Wasserman, 2018).

1.6 O modelo dos três estratos de Carroll

A teoria de três estratos das habilidades cognitivas proposta por John Carroll foi escrita em sua obra *Cognitive Abilities: A survey of factor-analytic studies* (Carroll, 1993). Carroll utilizou 460 matrizes de dados publicadas entre aproximadamente 1930 e 1985, relacionadas a testes de avaliação de capacidades cognitivas. Aplicando técnicas de análise fatorial exploratória com uma variedade de critérios quantitativos buscou investigar o número de fatores a serem fixados. Muitos desses conjuntos de dados eram provenientes de pesquisas que exploraram teorias das capacidades cognitivas propostas por autores como Spearman, Thurstone, Vernon, Cattell e Horn. Por este motivo, há semelhanças entre esses modelos, e a teoria de Carroll é considerada uma expansão e extensão das teorias anteriores (Carroll, 2005; Ortiz, 2015).

A teoria de Carroll propõe uma estrutura hierárquica das capacidades intelectuais em três diferentes camadas, construída a partir de técnicas de análise fatorial e pela matriz de correlação das variáveis estudadas. Os estratos coincidem com o grau de amplitude do fator no âmbito das capacidades cognitivas. No primeiro estrato estão as aptidões específicas que

surtem diretamente pela aplicaçaõ da matriz às variáveis originais dos testes psicológicos. Nesse estrato mais baixo, existem numerosos fatores de primeira ordem (cerca de 69 habilidades) que representam especializações agrupadas sob os fatores de segunda ordem. O segundo estrato possui oito fatores amplos semelhantes ao que foi proposto por Cattell-Horn: inteligência fluída, inteligência cristalizada, memória geral e aprendizagem, percepção visual geral, percepção auditiva geral, capacidade de recuperação geral, rapidez cognitiva geral e velocidade de processamento. No topo da estrutura, o terceiro estrato, está a aptidão geral, conceitualmente semelhante ao *g* de Spearman. Com a estrutura estabelecida por Carroll, qualquer habilidade importante identificada, poderia ser encaixada em um dos três estratos. Ele mesmo, em 2003, incluiu outros dois fatores no segundo estrato, sugerindo que este modelo pode ser expandido (Carroll, 1993, 2005; Ortiz, 2015; Wasserman, 2018).

1.7 O modelo Cattell-Horn-Carroll (CHC)

Ao avaliar as concordâncias gerais e as discrepâncias observadas apenas nos detalhes dos modelos teóricos de inteligência, entre diálogos, acordos e desacordos, em 1998, McGrew e Flanagan propuseram uma taxonomia cognitiva integrando dois modelos já existentes. O primeiro é o Gf-Gc estendido, desenvolvido por Cattell e Horn e o segundo é a teoria dos três estratos de Carroll. Este modelo ficou conhecido como a teoria Cattell-Horn-Carroll (CHC), considerado uma das teorias de inteligência mais abrangentes e reconhecidas atualmente (Flanagan & Dixon, 2014; McGrew, 2005, 2009).

Há um consenso amplamente difundido de que o modelo CHC oferece uma taxonomia adequada para representar as diferenças individuais referentes às habilidades cognitivas humanas (Flanagan & Dixon, 2014; McGrew, 2009; Primi, 2003). É fundamentado por uma teoria psicométrica da inteligência e desenvolvido por pesquisas baseadas em programas estatísticos de análise fatorial. O uso da estratégia empírica para a identificação da estrutura de inteligência é um componente importante para o estudo dos fatores latentes que podem explicar os resultados em tarefas, como nos testes psicológicos. Essa abordagem tem contribuído significativamente para os estudos que estabelecem os fundamentos conceituais da inteligência (McGill & Dombrowski, 2019; McGrew, 2009; Reynolds et al., 2013).

O modelo CHC de inteligência propõe que as habilidades cognitivas são dispostas em níveis diferentes em função do seu grau de abrangência e especialização e, em essência, a inteligência é multidimensional e integrada (Flanagan & Dixon, 2014). As dimensões de habilidade estão integradas em uma estrutura hierárquica, não se trata de categorias rígidas,

mas sim de uma maneira de descrever as habilidades em categorias em função de seus diferentes graus de amplitude. No primeiro nível hierárquico, as habilidades específicas são as únicas habilidades que podem ser medidas diretamente, relacionadas a uma atividade ou um teste psicológico (parte inferior da hierarquia). Os demais níveis superiores indicam, progressivamente, habilidades mais genéricas e amplas, são classificações teóricas inferidas pelas observações das relações entre as habilidades específicas (Schneider & McGrew, 2018).

A proposta inicial do modelo CHC apresentava o fator *g* como representante da inteligência geral, 10 habilidades amplas e, aproximadamente, 70 habilidades específicas. Em 2009, seis habilidades amplas foram acrescentadas (McGrew, 2009). Embora o fator *g* seja por vezes criticado e questionado, ele se mantém no modelo atual, com 17 habilidades amplas e 90 habilidades específicas. As atualizações do modelo não geram apenas novas habilidades, há revisões, como por exemplo, dos nomes utilizados como foi o caso da Memória de curto prazo (*Gsm*) renomeada para Capacidade de memória de trabalho (*Gwm*); divisão de uma habilidade ampla, Armazenamento e recuperação de longo prazo (*Glr*), em duas habilidades amplas: Aprendizagem de longo prazo (*Gl*) e Fluência de recuperação (*Gr*) (Schneider & McGrew, 2018).

A Tabela 1, baseada em Campos, Zaia e Primi (2019), oferece uma síntese dos fatores amplos (estrato II) e nomeia os fatores específicos (estrato I) do modelo CHC de inteligência. Foram realizadas algumas breves correções nas nomeações e siglas e acrescentadas informações para refletir a atualização feita por Schneider e McGrew em 2018. Optou-se por seguir a apresentação dos fatores estabelecida por Schneider e McGrew e as principais mudanças foram destacadas em letra itálica.

Tabela 1. Descrição das habilidades amplas e específicas do modelo CHC de inteligência

Fatores amplos (estrato II)	Descrição	Habilidades específicas (estrato I)
Inteligência fluída (<i>Gf</i>)	Capacidade para raciocinar em situações novas ou inesperadas, envolve a capacidade de resolver problemas novos, relacionar ideias, induzir conceitos abstratos.	Indução (<i>I</i>); Raciocínio sequencial geral (<i>RG</i>); Raciocínio quantitativo (<i>RQ</i>); <i>Fatores provisórios</i> : Velocidade de raciocínio (<i>RE</i>); Raciocínio piagetiano (<i>RP</i>).
Capacidade de memória de trabalho (<i>Gwm</i>)	Envolve a capacidade de manutenção de informações por curto período de tempo, ocorrendo a recuperação logo em seguida.	<i>Armazenamento auditivo de curto prazo (Wa)</i> ; <i>Armazenamento visual-espacial de curto prazo (Wv)</i> ; <i>Controle da atenção (CA)</i> ; Capacidade da memória de trabalho (<i>Wc</i>).
Eficiência de aprendizagem (<i>Gl</i>)	Relacionada ao armazenamento e consolidação da informação a longo prazo.	Memória associativa (<i>MA</i>); Memória significativa (<i>MM</i>); Memória de recuperação livre (<i>M6</i>).

Fluência de recuperação (Gr)	Envolve a extensão e fluência para recuperar informações por associações a outros itens, podendo envolver a criatividade e produção de ideias.	<i>Produção de ideias:</i> Fluência ideacional (FI); Fluência de expressão (FE); Fluência associativa (FA); Sensibilidade para problemas/ Fluência de solução (SP); Originalidade/ Criatividade (FO); <i>Recordação fluente de palavras:</i> Velocidade de acesso lexical (LA); Facilidade de nomear (NA); Fluência de palavras (FW); <i>Relacionados à figuras:</i> Fluência figural (FF); Flexibilidade figural (FX).
Velocidade de processamento (Gs)	Capacidade de realizar rapidamente, tarefas comuns dentro de um limite de tempo.	<i>Tarefas cognitivas:</i> Velocidade perceptual (P); Velocidade de pesquisa perceptual (Ps); Comparação de velocidade perceptual (Pc). <i>Área acadêmica:</i> Facilidade numérica (N); Velocidade de leitura (RS); Velocidade de escrita (WS).
Velocidade de reação e decisão (Gt)	Definida pela rapidez em fornecer respostas corretas em problemas de compreensão e raciocínio e tomar decisões envolvendo processamentos complexos.	Tempo de reação simples (R1); Tempo de reação para escolha (R2); <i>Tempo de inspeção (IT)</i> ; Velocidade de processamento semântico (R4); Velocidade de comparação mental (R7).
<i>Velocidade psicomotora (Gps)</i>	<i>Capacidade de realizar movimentos motores do corpo físico (dedos, mãos, pernas, por exemplo) com velocidade e fluidez</i>	Velocidade de movimento do membro (R3); Velocidade de escrita (WS); Velocidade de articulação (PT); Tempo de movimento (TM).
Inteligência cristalizada (Gc)	Definida a partir de experiência e conhecimentos aprendidos, inclui a compreensão da comunicação, estando associada ao conhecimento declarativo e de procedimentos.	Desenvolvimento da linguagem (LD); Conhecimento lexical (VL); Conhecimento geral (K0); Capacidade auditiva (LS); Capacidade de comunicação (CM); Sensibilidade gramatical (MY);
Conhecimento de domínios específicos (Gkn)	Conhecimento declarativo e processual especializado, estando associado a conhecimentos adquiridos ao longo da vida em determinadas áreas específicas.	Informações científica gerais (K1); Conhecimento cultural (K2); Conhecimento mecânico (MK); Proficiência de língua estrangeira (KL); Conhecimento de assinatura (KF); Habilidade na leitura labial (LP).
Leitura e escrita (Grw)	Habilidade básica de leitura e escrita utilizadas na compreensão da linguagem escrita e na expressão de pensamento pelo ato de escrever, estando associada à capacidade de decodificação, composição e compreensão de textos.	Compreensão da leitura (RC); Decodificação da leitura (RD); Velocidade de leitura (RS); <i>Velocidade de escrita (WS)</i> ; Uso da língua inglesa (EU).

Conhecimento quantitativo (Gq)	Compreensão de conceitos e relações quantitativas, manipulação de símbolos numéricos, estando associado à capacidade de utilizar informações quantitativas.	Conhecimento matemático (KM); Desempenho em matemática (A3)
Processamento visual (Gv)	Capacidade de sintetizar estímulos e gerar, perceber, armazenar e transformar imagens visuais.	Visualização (VZ); Relações espaciais/ <i>Rotação acelerada</i> (SR); Imagens (IM); Flexibilidade de finalização (CF); Velocidade de finalização (CS); Memória visual (MV); Análise espacial (SS); Integração perceptual em série (PI); Estimção de comprimento (LE); Percepção de ilusões (IL); Alternações perceptivas (PN).
Processamento auditivo (Ga)	Compreensão e síntese da configuração auditiva, percepção, análise e síntese sonora e a capacidade discriminativa de sons e de nuances em estruturas musicais complexas.	<i>Auditiva verbal</i> : Codificação fonética (PC); Discriminação da linguagem sonora (US); Resistência a estímulos auditivamente distorcidos (UR); <i>Auditiva não verbal</i> : Manutenção e avaliação do ritmo (U8); Memória para padrões de sons (UM); Avaliação e discriminação musical (U1, U9); Tom absoluto (UP); Localização sonora (UL).
Habilidades olfatórias (Go)	Habilidade para detectar e processar informações significativas do sistema olfativo relacionado a odores.	Memória olfativa (MO).
Habilidades táteis (Gh)	Habilidades para perceber e processar informações significativas em sensações hápticas, não apenas referente ao toque, mas à cognição realizada com sensações táteis.	Embora a Sensibilidade tátil (TS) tenha sido apresentada como uma habilidade estreita de Gh, ela recebe crítica por se uma habilidade sensorial e não cognitiva. As pesquisas para este fator ainda são limitadas para a definição de evidências estruturais comprovadas cientificamente.
Habilidades cinestésicas (Gk)	Habilidade para perceber e processar informações significativas ao posicionamento do corpo, peso, movimento, entre outros.	Ainda não há evidência estrutural significativa apresentada em estudos para esta habilidade. A Sensibilidade cinestésica (KS) trata-se de uma habilidade de acuidade sensorial relacionada a discriminação final em sensação proprioceptiva, como se e quanto um membro foi movido).

Habilidades psicomotoras (Gp)	Habilidades para realizar movimentos motores do corpo físico com precisão, coordenação ou força.	Destreza manual (P1); Destreza dos dedos (P2); Força estática (P3); Equilíbrio corporal bruto (P4); Coordenação de múltiplos membros (P6); Estabilidade das mãos e braços (P7); Precisão de controle (P8); Sequência de movimentos (mira/objetivo) (AI).
Inteligência Emocional (Gei)	Capacidade de percepção de expressões emocionais, estando associada a compreensão do comportamento emocional e de resolução de problemas por meio das emoções.	<i>Habilidades provisórias:</i> Percepção de emoção (Ep); Conhecimento da emoção (Ek); Gerenciamento das emoções (Em); Utilização da emoção (Eu).

Nos últimos anos, o modelo CHC de inteligência tem sido utilizado como base teórica para uma série de instrumentos criados, além de servir para atualizar os principais testes psicológicos. Mesmo nos casos em que os testes não foram originalmente desenvolvidos com base no modelo CHC, os pesquisadores têm se empenhado em realizar estudos de atualização, buscando alinhar as tarefas propostas nos testes com os princípios desse modelo (Caemmerer, Keith, & Reynolds, 2020). Por exemplo, a Escala de Inteligência Wechsler para Crianças quinta edição - WISC-V (Keith et al., 2006; McGill & Dombrowski, 2019) e a Bateria de Provas de Raciocínio (BPR-5), instrumento construído no Brasil, inicialmente com base na teoria de Thurstone (Primi, Nakano, & Wechsler, 2012).

1.8 A prática da avaliação psicológica da inteligência

A avaliação psicológica da inteligência é uma ciência aplicada e assume um papel de grande relevância. Através de métodos, técnicas e instrumentos, busca-se compreender as características psicológicas dos indivíduos e auxilia os psicólogos na tomada de decisão em diversos contextos (Reppold, Zanini, & Noronha, 2019; Wasserman, 2018). É evidente que tais avaliações não apenas impactam a experiência individual, mas também desempenham um papel vital no funcionamento das estruturas sociais mais amplas. Isso ressalta a importância que a avaliação psicológica possui ao longo do tempo como uma prática de proteção para a sociedade brasileira (Zanin et al., 2022).

A inteligência e, de maneira mais ampliada, as habilidades cognitivas, ao longo do tempo tornaram-se um dos construtos mais estudados e pesquisados no contexto educacional, industrial, do trabalho e organizacional (Chiodi, 2012; Larson, 2019). Como resultado, foi demonstrado o poder da inteligência para prever o desempenho humano em várias situações,

entre elas, no local de trabalho, na capacidade de produção e desenvoltura profissional. Esses achados trazem implicações para a importância de se ter instrumentos para mensuração da inteligência que podem ser utilizados nos processos de recrutamento e seleção a fim de gerar melhores resultados individuais e coletivos à organização (Hunter & Schmidt, 1996; Hunter, Schmidt, & Judiesch, 1990).

O fator geral, historicamente, é uma das características mais estudadas e operacionalizadas. Alguns resultados apontam sobre como ele influencia na vida diária dos indivíduos com associações entre desempenho escolar e profissional, grau de instrução e renda (Gottfredson, 2006). Em geral, um *g* alto pode prever sucesso em atividades com maior complexidade, no entanto, especificamente no contexto laboral, os indivíduos para executar as funções profissionais necessitam de diferentes capacidades. Em tarefas mais operacionais e rotineiras, o grau de predição da capacidade mental geral tende a ser menor (Lang et al., 2010; Schneider & Newman, 2015). Kell e Lang (2017) propõem que deve-se buscar um alinhamento entre a extensão de uma característica psicológica preditora (capacidade do indivíduo) e a extensão do critério de interesse do desempenho profissional. Dessa maneira, uma característica ampla como o fator *g*, pode ser melhor preditor de desempenho no trabalho quando relacionado à critérios de trabalho mais amplos, como por exemplo, desempenho geral no trabalho. As habilidades cognitivas específicas seriam melhores predictoras quando se tem como foco uma variável de trabalho mais especializada, como na situação de investigação de comportamento contraproducente no trabalho.

Uma empresa que busca selecionar funcionários deve considerar o tipo de trabalho a ser desempenhado pois isso poderá influenciar sobre qual seria a melhor característica preditora, uma habilidade mental geral ou uma habilidade específica, para avaliar o desempenho laboral. Uma empresa que busca por um engenheiro de aviação pode considerar que a habilidade espacial seja mais importante neste contexto de trabalho e ter maior validade ecológica do que uma habilidade mental geral (Kell & Lang, 2017). Com o reconhecimento das habilidades mentais específicas e a sua relação com o desempenho dos indivíduos, como exemplo, a aptidão numérica para a atividade de um contador, capacidade percepto-motora e de atenção para o operador de máquinas, memória visual para seguranças, rapidez perceptual para o investigador de incêndio, fica demonstrada a relevância dos instrumentos para mensuração das habilidades amplas, mas também das específicas (King, 1957; Robbins, Judge, & Sobral, 2010).

Nesse sentido, as habilidades cognitivas avaliadas pelos testes da Bateria TSP desempenham um papel crucial em uma variedade de atividades profissionais. Isso ressalta a

relevância desses instrumentos para avaliações psicológicas em especial no ambiente organizacional e de trabalho, e explica seu contínuo uso e comercialização, especialmente para seleção de pessoas para concursos e empresas privadas. O objetivo da pesquisa foi realizar estudos para atualização dos Testes de Seleção Profissional (Bateria TSP), seguindo as diretrizes específicas da Comissão Internacional de Teste e os padrões para testes educacionais e psicológicos (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Com a autorização da Edites – Empresa Distribuidora de Testes Ltda., responsável pela comercialização no Brasil, a pesquisa foi conduzida para contribuir com estudos recentes que investigam a qualidade da Bateria TSP. Desta maneira, busca-se assegurar sua contínua utilização pela comunidade de psicólogos em suas práticas profissionais, bem como contribuir para o aprimoramento dos modelos teóricos da inteligência.

2. A BATERIA TSP

Com nome original “Série Fatorial de Aptidões”, a Bateria TSP, foi criada em 1947, nos Estados Unidos, pelo norte americano Joseph E. King. O instrumento foi desenvolvido para atender às demandas do contexto organizacional e do trabalho, como exemplo, uso em processos de seleção de pessoas, atribuições de cargos, treinamentos, promoções, principalmente, para a indústria e o comércio. O autor propôs utilidade do teste também para sujeitos em formação no ensino médio e superior com o objetivo de fornecer informações para a orientação profissional e o planejamento de carreira. A Bateria teve como base os modelos teóricos propostos por Thurstone e Guilford, na qual compreende que a inteligência envolve fatores e aptidões específicos diferentes e não pode ser reduzida a uma capacidade geral única. Entendia-se que os fatores específicos eram mais relevantes para avaliar os pontos fortes e fracos das habilidades intelectuais, visando prever a eficiência no desempenho laboral (King, 1957).

Foram criados 14 testes para avaliar oito fatores da inteligência, cobrindo habilidades com conteúdo verbal e não verbal: compreensão, raciocínio, sistemas, percepção, fluência, memória, relações espaciais e coordenação. Alguns fatores se repetiam em diferentes testes em função de sua importância para as atividades de trabalho (King, 1957). O Quadro 1 apresenta os oito fatores, juntamente com os testes correspondentes e a descrição do construto avaliado.

Quadro 1. Testes originais da Série Fatorial de Aptidões (King, 1957)

Série Fatorial de Aptidões		
Fator	Testes	Descrição do construto
Compreensão	Terminologia de Escritório Terminologia de Vendas (verbais) Objetos (não verbal)	Habilidade para compreender ideias, vocabulário, imagens e outros tipos de informação nos setores empresarial e industrial.
Raciocínio	Julgamento (verbal)	Habilidade para abordar e resolver problemas complexos, pensar de forma lógica, prever, planejar e lidar com conceitos abstratos.
Sistemas	Números (verbal)	Habilidade para trabalhar de forma rápida e precisa com números, arquivos, códigos, símbolos e procedimentos padrão - no qual um conjunto de símbolos é manipulado por regras específicas.
Percepção	Percepção (verbal) Precisão (não verbal)	Habilidade de examinar e localizar detalhes rapidamente, reconhecendo semelhanças e diferenças.
Fluência	Fluência (verbal)	Habilidade para usar as palavras com facilidade, escrever e falar sem hesitação ou dificuldade em encontrar a palavra certa.
Memória	Memória (verbal)	Habilidade para reconhecer e recordar associações, tais como nomes e rostos.
Relações espaciais	Partes, Blocos e Dimensão (não verbais)	Habilidade para perceber tamanhos, formas e relações espaciais de objetos em duas e três dimensões: capacidade de discernir disposição e organização.
Coordenação	Destreza e Motor (não verbais)	Habilidade para atividade motora fina e grossa, destreza e coordenação dos olhos e das mãos.

Na construção da Bateria, os itens foram formulados considerando a pertinência do conteúdo para os contextos industrial e comercial, bem como a área de trabalho mais associada a cada fator em questão. Aproximadamente 70 itens foram criados para cada teste e submetidos a análises estatísticas para verificação da homogeneidade, do nível de dificuldade e da correlação do item com o fator que se pretendia medir. O autor da Bateria tinha como objetivo limitar o tempo de cada teste a aproximadamente cinco minutos. Esse tempo foi considerado apropriado devido à homogeneidade dos fatores e à necessidade de usar um conjunto de instrumentos para prever a eficiência do candidato diante das diversas tarefas de

um cargo. Também foi considerado que o tempo total para a avaliação não deveria ultrapassar 30 minutos para cargos comuns (King, 1957).

Entre 1947 e 1950, foram realizados estudos iniciais de validação do instrumento. A confiabilidade foi avaliada através da consistência interna, utilizando o método das metades e o coeficiente Kuder-Richardson. Os resultados para os 14 testes variaram de 0,89 a 0,96. Um estudo de verificação da estabilidade temporal foi realizado com 100 participantes, que responderam aos mesmos testes novamente após seis meses. Os resultados variaram de 0,79 para o teste Fluência a 0,91 para o teste Números (King, 1957).

Outro aspecto ressaltado pelo autor em relação à confiabilidade é o uso comum de um conjunto de testes, em oposição à utilização de apenas um teste isolado. Isso possibilita calcular uma pontuação ponderada, levando em consideração todos os testes aplicados. Dois estudos foram conduzidos sobre o escore composto ponderado, com resultados de confiabilidade superiores a 0,90. Em um estudo com 100 participantes submetidos aos testes Terminologias de Escritório, Números, Percepção e Julgamento, a pontuação composta ponderada foi obtida a partir desses quatro testes. Posteriormente, os mesmos sujeitos foram reavaliados após três meses, resultando na correlação de 0,92 entre as duas pontuações ponderadas obtidas no teste e reteste (King, 1957).

No que diz respeito aos resultados da validade da pesquisa americana, as intercorrelações mais baixas foram observadas entre o teste Motor e todos os demais (0,05 a 0,15), seguido pelos testes em função da natureza do estímulo, verbal e não verbal (0,15 a 0,35), e entre os testes verbais com fatores diferentes (0,20 a 0,50) (King, 1957). Resultados semelhantes foram observados em um estudo realizado com marinheiros da Marinha Real Holandesa em 1961. O estudo teve a participação de 500 sujeitos, a maioria com 16 anos de idade. Os resultados indicaram correlações mais baixas entre o teste Destreza (fator Coordenação) e todos os demais testes (0,05 a 0,29), seguido pelos testes com estímulo verbal e não verbal (0,02 a 0,41). As maiores correlações em função do tipo estímulo foram observadas entre o teste Julgamento e os testes não verbais (Johnson, te Nijenhuis, & Bouchard Jr, 2008), resultado semelhante com os dados da pesquisa brasileira (Nascimento & Flores-Mendoza, 2007).

King (1957) também conduziu estudos de validade convergente com alguns testes da Bateria comparados com outros instrumentos utilizados naquele período que seguiam o mesmo padrão. A amostra variou entre 129 e 428 participantes, os resultados da correlação entre os testes foram de 0,63 a 0,84. Por exemplo, foi encontrada uma correlação de 0,76 entre

o teste Percepção e o *Minnesota Clerical Test*. O teste Partes e o *Minnesota Paper Form Board Test* tiveram correlação de 0,78.

Embora tenha sido concebido em 1948, o programa “*Job-Tests Program*” foi lançado em 1951, com o objetivo de validar um grupo de testes para identificação das habilidades intelectuais requeridas nas áreas básicas do trabalho na indústria e no comércio. Primeiramente, foram identificadas 24 áreas comuns desses contextos e os cargos específicos que os compõem, sendo algumas áreas: escriturário, mecânico, vendas, técnico e supervisor. Foram estudados e definidos quais os grupos de testes avaliavam as habilidades necessárias para aprender e exercer as atividades de cada área. Foram atribuídos pesos a cada fator de aptidão envolvido, definidos a partir dos coeficientes de correlação, bem como nos resultados de regressão para cada teste e na intercorrelação com os demais testes da Bateria. Com o objetivo de orientar as empresas foi desenvolvido um manual intitulado “*The 24 Job-Test Areas*”, contendo aproximadamente 800 cargos relacionados às 24 áreas mapeadas e descritas. O manual permitia a identificação do cargo que a empresa pretendia preencher e recomendava os testes adequados para essa função. Além disso, oferecia orientações simples para que a pontuação e os escores ponderados pudessem ser calculados por quem utilizasse a Bateria (King, 1957).

Por fim, um aspecto relevante nos estudos originais da Bateria é que os dados normativos foram obtidos a partir da participação de 5.000 a 10.000 candidatos e indivíduos empregados. Foi publicado um único conjunto de normas, convertendo as pontuações brutas em percentis e estatinos, sem criar tabelas específicas com base nas características individuais como: idade, sexo, escolaridade ou classificação ocupacional. O autor considerou que os empregos requerem habilidades básicas, independentemente das características sociodemográficas dos candidatos. Em outras palavras, o grau de aptidão intelectual necessário para desempenhar uma atividade profissional deve prevalecer sobre algumas diferenças observadas nos grupos de variáveis individuais (King, 1957).

No Brasil, os primeiros estudos de tradução e adaptação da bateria TSP foram realizados pelo psicólogo Glauco Piovani, pelos interesses da Editora Edites, entre as décadas de 1960 e 1970. Houve uma revisão em 1987, com a condução do psicólogo César Piovani. Esses estudos mantinham a base teórica do instrumento original proposta por Thurstone que haviam sete construtos correlacionados (capacidade espacial, numérica, verbal, velocidade perceptual, memória, fluência verbal e raciocínio). A revisão e reavaliação mais recente dos testes e aprovada pelo CFP foi conduzida pelo Laboratório de Avaliação das Diferenças Individuais (LADI), do Departamento de Psicologia da UFMG, comandada pelas professoras

doutoras Elizabeth do Nascimento e Carmen Flores-Mendoza. Nesta revisão houve uma primeira aproximação com a teoria CHC de inteligência em sua fundamentação teórica (Nascimento & Flores-Mendoza, 2007).

É válido esclarecer que nos estudos brasileiros, os três testes para o fator Compreensão foram excluídos da Bateria TSP em 2004, pois os resultados de algumas pesquisas indicaram que os itens estavam defasados. O teste Destreza, após estudos de revisão e atualização, em 2006, tornou-se independente, possuindo manual próprio e com aprovação isolada no CFP (Flores-Mendoza & Nascimento, 2007). Por fim, o teste Motor presente na avaliação do fator Coordenação, nem ao menos foi numericamente mencionado. Observa-se nos registros da Edites (Edites, s.d.) que o instrumento original era composta por 13 testes, e não 14. Acredita-se que isso pode ser em função do formato do teste, que consistia em um aparelho de manipulação de objetos, como porcas, parafusos e arruelas (King, 1957).

Na última versão aprovada pelo CFP, em 2007, a Bateria TSP era composta por nove testes: Fluência, Números, Julgamento, Memória, Percepção Partes, Blocos, Dimensão e Precisão. O público-alvo eram pessoas com idades entre 16 e 50 anos. Embora houvesse uma ênfase de aplicação no contexto Organizacional e do Trabalho, assim como na Psicologia do Trânsito, a Bateria TSP podia ser utilizada em qualquer área que necessitasse de uma avaliação rápida das funções mentais. Na Tabela 2, os testes são descritos em relação a habilidade cognitiva avaliada e ao tipo de estímulo predominante em cada um (Nascimento & Flores-Mendoza, 2007).

Tabela 2. Testes, construtos avaliados e tipo de estímulo presente na Bateria TSP

BATERIA TSP		
Teste	Habilidade cognitiva avaliada	Tipo de estímulo
Fluência	Fluência verbal	Verbal
Números	Habilidade numérica	Verbal
Julgamento	Raciocínio lógico	Verbal
Memória	Memória	Verbal / Não-verbal
Percepção	Habilidade perceptual	Verbal
Precisão	Habilidade perceptual	Não verbal
Partes, Dimensão e Blocos	Habilidade espacial	Não verbal

Os testes da Bateria TSP são independentes, é possível que o psicólogo avaliador escolha qual ou quais serão utilizados de acordo com o seu objetivo. Diferentemente do estudo original, não existem pesquisas em função das áreas de trabalho para o contexto

brasileiro. O tempo de aplicação da Bateria completa varia de 90 a 120 minutos de acordo com o grau de escolaridade do examinando. Todos os testes são cronometrados. Exceto pelo teste Fluência e Memória, os demais têm duração de cinco minutos para a execução (Nascimento & Flores-Mendoza, 2007).

A Bateria TSP caracteriza-se pela facilidade de aplicação e correção, podendo ser realizada de forma coletiva ou individual. Oferece instruções claras e objetivas, com exemplos resolvidos e problemas semelhantes aos encontrados no teste propriamente dito, para que sejam respondidos após a instrução, como método de treinamento. A correção é feita com crivos, o que tende a minimizar possíveis erros. Essas características contribuem para a utilização eficiente da Bateria, promovendo agilidade no processo de avaliação realizado no contexto organizacional e do trabalho, conforme proposto em sua criação (Nascimento & Flores-Mendoza, 2007).

Sobre as análises psicométricas conduzidas no Brasil, a precisão avaliada pela consistência interna os resultados foram satisfatórios. O teste Fluência foi o único avaliado pelo coeficiente de *Guttman* em função do número pequeno de itens. O resultado foi 0,74, menor índice alcançado pela Bateria. Os demais testes apresentaram resultados considerados adequados, com valores para o coeficiente alfa de Cronbach variando entre 0,86 e 0,94 (CFP, 2022). A estabilidade temporal feita por meio do teste e reteste (intervalo de 21 dias), apresentou coeficientes de correlação de *Pearson* que variaram entre 0,54 e 0,88 ($p < 0,001$).

Para as evidências de validade foram feitas análises exploratórias em que demonstraram a unidimensionalidade de todos os testes da Bateria, com cargas fatoriais entre 0,60 e 0,86, índices considerados moderados e fortes (Dancey & Reidy, 2013). A correlação com testes de inteligência foi feita com o Matrizes Progressivas de *Raven* - Escala Geral e o Teste não-verbal de inteligência (G-36). Os resultados apontaram que os testes Números, Percepção e Fluência foram os únicos que não compartilham recursos cognitivos de natureza figurativa-espacial, com índices de correlação não significativos que variaram entre 0,0 e 0,19, magnitude nula a fraca (Dancey & Reidy, 2013). Foi feito um estudo com as variáveis sexo, faixa etária e escolaridade, em função das diferenças encontradas a partir dos resultados do Teste *t* de *Student*, as autoras estabeleceram normas controlando essas características (Nascimento & Flores-Mendoza, 2007).

Serão apresentados os testes que compõem a Bateria TSP, demonstrando as características específicas de estímulo, a tarefa solicitada ao examinando, a duração da execução e o método de contabilização da pontuação.

2.1 Teste Fluência

O teste Fluência é subdividido em quatro seções de tarefas com estímulos verbais (letra, sufixo, prefixo, por exemplo), que envolvem a evocação e a escrita de um maior número de palavras em um tempo limite de dois minutos para cada seção. No item de treinamento do teste apresentado abaixo é solicitado ao examinando que escreva todas as palavras terminadas em TAR, ele poderá escrever quantas palavras recordar e em qualquer ordem (Nascimento & Flores-Mendoza, 2007).

Escreva todas as palavras que conhece terminadas em: TAR

ajuntar

O escore é formado sendo atribuído 0 para as respostas erradas e 1 para as palavras que obedeceram ao comando solicitado. Uma observação importante é que erros gramaticais não são considerados, então, desde que o examinando tenha atendido ao comando do item, será atribuído um ponto mesmo à escrita incorreta da palavra. No entanto, não poderá usar variações de uma mesma palavra já escrita, como aumentativo, diminutivo, palavras ou tempos verbais, nem mesmo nomes próprios. Neste teste não há limite máximo de pontuação (Nascimento & Flores-Mendoza, 2007).

O objetivo do teste Fluência é o de avaliar a capacidade de vocabulário simples do examinando, representando um aspecto da aptidão verbal, pois indica a facilidade em utilizar palavras (Nascimento & Flores-Mendoza, 2007). Este tipo de tarefa verifica a amplitude de palavras adquiridas pelo examinando relacionadas às habilidades básicas de escrever palavras isoladas com características fonéticas específicas (como demonstrando pelo item exemplo). O teste Fluência demanda a rapidez e a capacidade de evocação de informações estabelecidas na memória de longo prazo, relacionada ao vocabulário adquirido pela pessoa. As pessoas diferem quanto ao armazenamento de informações na memória de longo prazo e, este aspecto da habilidade tem se tornado cada vez mais reconhecido devido ao seu papel na compreensão da leitura, da escrita e da comunicação (Schneider & McGrew, 2012).

O teste Fluência pode ser associado ao fator de primeira ordem Fluência de palavras (FW) no modelo Cattell-Horn-Carroll (CHC) de inteligência, definida como a capacidade de

rapidamente pensar e produzir palavras isoladas que compartilham uma característica fonêmica, estrutural ou ortográfica específica (McGrew, 2005; Schneider & McGrew, 2012). Esta aptidão pertencia originalmente à habilidade ampla, Armazenamento e recuperação de longo prazo (Glr), no entanto a partir da evolução da teoria CHC e de suas melhorias graduais ao longo do tempo, à habilidade cognitiva Glr foi separada formalmente em 2018, formando duas categorias amplas: Eficiência de aprendizagem (Gl) e Fluência de recuperação (Gr). Esta última, relacionada à velocidade de recuperação seletiva e estratégica de informação verbais e não verbais armazenadas na memória de longo prazo. Em Gr ficaram presentes as habilidades estreitas relacionadas à fluência em relação à produção de ideias e figuras, recordação fluente de palavras, como exemplo: Fluência ideacional (FI), Fluência de palavras (FW), Fluência figural (FF), entre outras (Schneider & McGrew, 2018). Dessa maneira, supõe-se que o teste Fluência da bateria TSP, esteja relacionado ao fator amplo do modelo CHC de inteligência, Fluência de recuperação (Gr), na habilidade estreita FW.

O teste Fluência pode também aparecer associado a outra aptidão ampla do modelo CHC, Inteligência cristalizada (Gc), como no estudo da Bateria de Provas de Raciocínio (BPR-5), em que o teste de Raciocínio verbal (RV) apresentou cargas fatoriais interpretadas para Gc (Almeida et al., 2010; Primi, Correia & Almeida, 2018). No entanto, a hipótese de trabalho é que o teste Fluência da bateria TSP será associado a Gr em função dessa habilidade estar teoricamente mais relacionada ao estímulo proposto pela tarefa, pois o mesmo requer do respondente velocidade para produzir palavras a partir de um recurso fonológico solicitado, condizente com FW, não sendo necessária a ortografia correta da palavra (o que poderia ser esperado em Gc).

2.2 Teste Números

O teste Números é composto por 54 problemas numéricos que envolvem duas das operações básicas de divisão, soma, multiplicação e/ou subtração formados por três números com um, dois até três dígitos. O teste é de múltipla escolha com quatro alternativas possíveis de serem assinaladas. Há um espaço para rascunho caso o examinando julgue ser necessário. O tempo limite para a resolução dos problemas aritméticos é de 5 minutos. As respostas são pontuadas em 0 (incorretas) e 1 (corretas), gerando o escore máximo de 54 pontos (Nascimento & Flores-Mendoza, 2007). Veja o item de treinamento presente no teste Números.

Abaixo há três exemplos de questões iguais às que resolverá no teste propriamente dito. Estude-as para compreender como deve escolher as respostas corretas.

	①	②	③	④	
$4 + 7 + 3 =$	12	13	14	15	3
$24 \div 6 + 3 =$	7	8	6	10	1
$5 \times 9 + 4 =$	45	49	51	54	2

Conforme pode ser visto no item de treinamento, existem pelo menos duas operações matemáticas exigidas em cada problema. Um dado importante é que diferente das operações matemáticas em que divisões e multiplicações devem ser realizadas primeiro, neste caso do teste, a resolução deve seguir a sequência que foi apresentada ao examinando. Então, veja o primeiro problema 4 mais 7 igual a 11; 11 mais 3 igual a 14. O número 3 foi assinalado no quadrado porque corresponde ao resultado da operação (14) que está na coluna de número 3. No teste Números, vale a regra do próprio teste, por isso as operações são feitas na sequência em que aparecem, inicia o cálculo entre o primeiro e o segundo número e, em seguida, é feita a próxima operação solicitada com o terceiro número (Nascimento & Flores-Mendoza, 2007).

O teste Números tem como objetivo verificar a rapidez e precisão no manejo e na compreensão de sistemas numéricos de cálculos básicos descontextualizados. Revela a capacidade de conhecimento obtido geralmente na escola em termos matemáticos e a velocidade para operá-la (Nascimento & Flores-Mendoza, 2007). Este procedimento parece seguir o cerne do fator específico Desempenho em matemática (A3), presente no fator de segunda ordem Conhecimento quantitativo (Gq), do modelo CHC de inteligência, em que o examinando precisa resolver problemas de cálculo matemático utilizando do raciocínio quantitativo aprendido (Schneider & McGrew, 2012, 2018). Em função disso, acredita-se que o teste Números da bateria TSP não esteja relacionado à inteligência fluida, pois o raciocínio quantitativo relacionado à Gf é entendido como a capacidade de raciocinar de maneira indutiva e dedutiva frente a um problema matemático, este não é o caso do teste Números.

O desempenho em tarefas matemáticas tendem a aparecer em diferentes habilidades amplas do modelo CHC de inteligência. No estudo da Bateria Multidimensional de Inteligência Infantil, o teste Desempenho Matemático foi inicialmente pensado para representar a capacidade específica A3, presente em Gq (Schelini & Wechsler, 2005). No entanto, no estudo da estrutura fatorial da bateria, o teste Desempenho Matemático obteve

associação com o fator específico Raciocínio quantitativo (RQ), presente na Inteligência fluída (Gf) (Schelini & Wechsler, 2006). A prova de Raciocínio Numérico, da BPR-5, apresentou cargas fatoriais para o fator de inteligência cristalizada do modelo CHC de inteligência (Almeida et al., 2010). O subteste Aritmética da Escala Wechsler de Inteligência (WISC) inicialmente foi associado à Memória de curto prazo (Gsm), mas em função de requerer escuta e linguagem receptiva também foi hipotetizado para estar relacionado à Gc e, também foi demonstrada associação com Gf (Keith et al., 2006). Essa mistura complexa de habilidades relacionadas ao subteste Aritmética do WISC está em consonância com o que foi apontado por Schneider e McGrew (2012) sobre a necessidade de pesquisa para determinar o posicionamento mais apropriado do sentido de número no modelo CHC de inteligência.

2.3 Teste Julgamento

No teste Julgamento o examinando deve observar a relação estabelecida entre a série de letras ou de números e deverá assinalar uma alternativa que responde corretamente o item entre as quatro disponíveis. É composto por 59 itens divididos em 6 blocos com 9 questões cada. O tempo limite para a resolução é de cinco minutos. As respostas são pontuadas em 0 (incorreta) e 1 (correta), com pontuação máxima de 54 pontos (Nascimento & Flores-Mendoza, 2007).

Existem itens que apresentam uma série com letras ou números que envolvem identificar igualdades e diferenças. Estes itens exigem a capacidade de identificar a característica subjacente apresentada combinando informações separadas para a formação de inferências (raciocínio de indução), como por exemplo:

Qual o grupo diverge dos demais?

1	2	3	4
EXHM	LMNO	ABCD	RSTU

Este item de treinamento do teste exige a identificação da resposta correta (EXHM) em que requer do examinando a descoberta da regra subjacente aos três grupos de letras que apresentam um mesmo padrão, no caso a ordem alfabética e, assinalar o número 1 no quadrado porque corresponde ao resultado (EXHM) que está na coluna de número 1.

Ou, itens com seriação de números e letras em que o examinando deverá analisar logicamente o princípio adotado nos itens para deduzir a resposta correta para a sequência

esperada (raciocínio de dedução). Pode-se ver pelo exemplo abaixo:

O número seguinte é: 4 5 5 6 6 6 ?

1	2	3	4
6	7	8	9

O item exige a capacidade de raciocínio hipotético-dedutivo para responder corretamente (7) a partir das conclusões tiradas das condições apresentadas pela sequência numérica determinada, no item exemplo, houve o aumento da quantidade de números e do número. Dessa maneira, o número 2 deverá ser colocado no campo de resposta, pois a resposta correta (7) está na coluna de número 2.

Ou, itens que predominam um raciocínio quantitativo que demandam do examinando a capacidade de raciocinar dedutivamente com conceitos envolvendo relações e propriedades matemáticas básicas. Veja um dos exemplos de treinamento do teste:

O número seguinte é: 3 5 7 9 11 13 ?

1	2	3	4
13	14	15	16

O item de treinamento do teste exige a capacidade de raciocínio dedutivo para solucionar corretamente (15) o problema matemático a partir das conclusões envolvendo as relações e as propriedades numéricas estabelecidas. No item exemplificado acima, é necessário ter um raciocínio análogo numérico de adição, em que é somado dois a cada um dos números.

O objetivo do teste Julgamento é o de verificar a aptidão do examinando para abordar e resolver problemas em termos de raciocínio lógico, mais evidente no raciocínio abstrato que depende menos do aprendizado anterior, característica central da inteligência fluída (Gf). Algumas habilidades específicas são mais essenciais para os fatores amplos, este é o caso da indução (I) e do raciocínio sequencial geral (dedutivo) para Gf, sendo o raciocínio quantitativo (RQ) considerado uma prioridade mais baixa desde que fatores de dificuldades matemáticas não estejam sendo focalizadas no exame (Schneider & McGrew, 2012, 2018). O teste Julgamento demanda do examinando a capacidade de reconhecer e compreender as relações entre as letras ou números e, apesar dessas diferenças de conteúdo o raciocínio dedutivo, o de indução e o raciocínio quantitativo são utilizados para que as conclusões sejam tomadas a partir da exposição de determinadas premissas em cada item, por este motivo

acredita-se estar associado à inteligência fluída. O Teste de Indução estudado por Gomes e Borges (2009b) apresenta características similares ao teste Julgamento e os resultados dos estudos psicométricos o associou à Gf do modelo CHC de inteligência (Gomes, Nascimento, & Araújo, 2021).

2.4 Teste Percepção

No teste Percepção são apresentados 27 problemas com dois pares (A e B) de sentenças cada um. A tarefa consiste na identificação das igualdades e diferenças comparando o item A de cima ao item A de baixo (o mesmo acontecerá para o item B). O examinando deverá registrar a quantidade de diferenças existente em cada par de sentenças, podendo ser de um até no máximo quatro erros. As diferenças podem ser relacionadas à falta ou troca de letra(s) e/ou número(s). O tempo limite para a resolução do teste é de cinco minutos. A pontuação poderá ser 0 (incorreto) ou 1 (correto) em cada par de itens analisado, com escore máximo de 54 pontos (Nascimento & Flores-Mendoza, 2007). Veja tarefa de treinamento contida no teste.

1º Exemplo:

A - Vinho branco Decova - grf. R\$ 12,50 - 1 l. R\$ 25,30

B - Vinho tinto Mascato - grf. R\$ 15,25 - 1 l. R\$ 35,25

A - Vinho branco Decova - grf. R\$ 11,50 - 1 l. R\$ 25,30

B - Vinho tinto Masceto - grf. R\$ 15,20 - 1 l. R\$ 33,25

1
3

O examinando ao comparar item A de cima com item A de baixo deverá perceber que existe um número de preço trocado (R\$12,50 e R\$ 11,50). Ao comparar item B de cima com item B de baixo, há também dois números de preços trocados (R\$15,25 e R\$15,20; R\$ 35,35 e R\$33,35) e uma letra trocada (Mascato e Masceto), por este motivo as respostas escritas nos quadrinhos ao lado foram 1 e 3, sequencialmente a resposta para as questões A e B. Nesta tarefa não constituem erros a falta de vírgula ou ponto ou qualquer outro sinal (Nascimento & Flores-Mendoza, 2007).

O objetivo do teste é o de verificar a aptidão para distinguir diferenças e semelhanças de natureza verbal, utilizado em trabalhos administrativos e outras atividades que envolvem grande quantidade de papéis e detalhes (King, 1957). Esta tarefa pressupõe fatores como atenção e perseverança, intrínsecos, porém não implicados ao objetivo principal (Nascimento & Flores-Mendoza, 2007). Existem diferenças individuais na habilidade perceptual que no

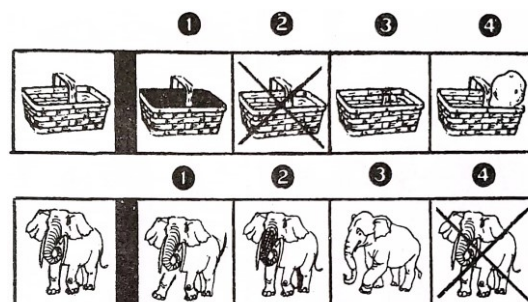
modelo CHC de inteligência está relacionada ao fator de segunda ordem Velocidade de processamento (Gs), definido como a capacidade de realizar, fluentemente, tarefas cognitivas simples e repetitivas (ou automatizadas), por meio do controle e da rapidez atencional, durante um intervalo definido de tempo (McGrew, 2005).

Acredita-se que a tarefa solicitada pelo teste Percepção esteja relacionada à habilidade estreita, Velocidade perceptiva (P). Esta habilidade cognitiva é considerada uma característica central de Gs (Schneider & McGrew, 2012), pois relaciona-se a velocidade e fluência em que estímulos visuais podem ser comparados quanto às suas semelhanças ou diferenças. Houve uma diferenciação proposta para a habilidade P, subdividida em Velocidade de pesquisa perceptual (Ps) e Comparação de velocidade perceptual (Pc). Alguns exemplos de teste Pc são: identificar pares de nomes ou números idênticos ou incompatíveis. Exemplo de teste Ps: examinar as palavras e riscar todas as palavras que tenham a letra “a” (Schneider & McGrew, 2018). O teste Percepção apresenta uma maneira de medir este fator em que pares de frases são apresentadas ao examinando para a identificação de falta ou troca de letra(s) ou número(s). A hipótese é de que esse teste tenha conteúdo relacionado à Pc.

2.5 Teste Precisão

O teste Precisão é caracterizado por conter uma série de figuras (em preto e branco) de objetos e de pessoas em atividade. Cada item apresenta uma figura modelo (à esquerda de uma tarja preta vertical) e quatro opções de resposta, onde só uma delas é idêntica a figura modelo. O examinando deverá responder a cada um dos 48 itens e assinalar com um X sobre a própria figura que considerar idêntica ao modelo. O tempo limite para a resolução do teste é de cinco minutos. As respostas são pontuadas em 0 (incorreta) ou 1 (correta) e a pontuação máxima é de 48 pontos (Nascimento & Flores-Mendoza, 2007). A questão abaixo, mostra uma tarefa de treinamento apresentada no teste.

Olhe bem a figura à sua esquerda. Qual das outras quatro figuras é exatamente igual a da esquerda?



O objetivo do teste é o de avaliar a aptidão para percepção rápida de semelhanças ou diferenças de natureza não verbal, adequado para tarefas de inspeção ou design (King, 1957). Assim como para o teste Percepção, nesta tarefa também pressupõe fatores como atenção e perseverança, intrínsecos, porém não implicados ao objetivo principal (Nascimento & Flores-Mendoza, 2007). Este teste também pressupõe associação com o fator Velocidade de processamento (Gs), do modelo CHC de inteligência. No entanto, considera-se que esteja relacionado, especificamente, à Comparação de velocidade perceptual (Pc). De acordo com o item de treinamento apresentado, a atividade requerida exige que o examinando analise a figura modelo presente à esquerda da tarja preta e julgue os quatro demais estímulos do item para descobrir qual deles é a figura idêntica ao modelo. Essa característica do teste Precisão está em conformidade com os exemplos de testes de Pc apresentados por Schneider e McGrew (2018), em que foi citada a tarefa de verificação de formas idênticas.

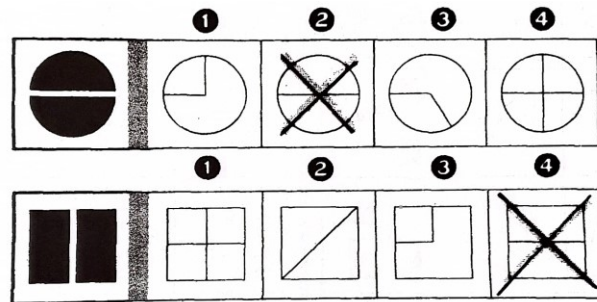
Schneider e McGrew (2018) sugerem que sejam classificados o conteúdo principal e o conteúdo secundário do teste, em função de um melhor entendimento das diferentes pontuações entre os vários testes de Ps e Pc. Neste caso, acredita-se que o teste Precisão seja categorizado como Pc-Gv, em função das figuras apresentarem secundariamente um processamento visual. Em uma pesquisa de validação realizada com o Teste de Figuras Idênticas (P3), semelhante ao teste Precisão da bateria TSP, foram feitas análises fatoriais exploratórias e o resultado foi maior para o fator cognitivo Gv (0,32) e g (0,30) (Gomes & Borges, 2007). Em um estudo da estrutura fatorial de um modelo ortogonalizado da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCaLo), o Teste de Velocidade Perceptiva (nomeado no estudo anterior como Teste de Figuras Idênticas), apresentou cargas fatoriais tanto em Gs (0,20), quanto em Gv (0,16). No entanto, o teste foi melhor explicado por g (0,66), situação em que o autor sugeriu estudos futuros para avaliar a relação entre Gs e g encontrados (Gomes, 2010). Dessa maneira, é válido investigar para o teste Precisão sua relação com Gv e g, para além do fator de segunda ordem Gs.

2.6 Teste Partes

No teste Partes são apresentadas uma série de figuras geométricas em que solicita ao examinando a junção de partes para formar um todo. Conforme poderá ver no item de treinamento, existe uma figura geométrica modelo escurecida e separada em suas partes a esquerda de uma tarja preta vertical. Na sequência, são apresentadas quatro opções de resposta em que o examinando deverá assinalar com um X sobre a alternativa em que há a

junção correta das partes para formar um todo da figura modelo. São 48 itens para resolver em um tempo limite de cinco minutos. As respostas são pontuadas em 0 (incorreta) ou 1 (correta), com pontuação máxima de 48 pontos (Nascimento & Flores-Mendoza, 2007).

Item de treinamento: Estude esses exemplos e suas respostas, procurando entender porque essas são as respostas corretas.



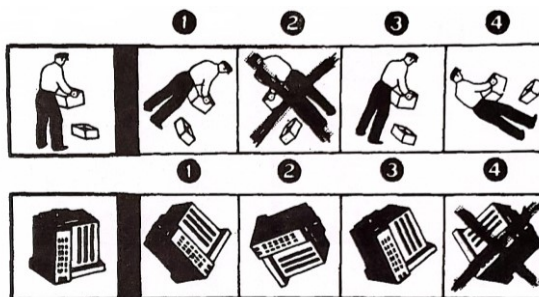
No primeiro item de treinamento, a figura modelo mostra duas partes escurecidas. Ao examinar as quatro alternativas de resposta, a única figura que mostra a maneira correta de as partes ficarem unidas é a de número 2, por este motivo existe um X assinalado sobre a figura. O mesmo ocorreu para o segundo item de treinamento, embora a alternativa 4 esteja virada quando comparada à figura modelo escurecida, ela é a única figura que representa corretamente a união das partes da figura modelo. Percebe-se, então, que as partes integradas podem estar rotadas (item 2) ou não (item 1) (Nascimento & Flores-Mendoza, 2007).

O objetivo do teste é o de verificar a aptidão para visualizar tamanhos, formas e a relação do todo com as partes, qualidades relacionadas à percepção espacial (Nascimento & Flores-Mendoza, 2007). Por este motivo, acredita-se que o teste Partes esteja associado à habilidade específica Relações espaciais/rotação acelerada (SR), presente no fator de segunda ordem Processamento visual (Gv), do modelo CHC de inteligência. Relações espaciais é definida como a capacidade de resolver problemas por meio da percepção e manipulação rápida de padrões visuais (rotação mental, transformações, reflexão, por exemplo), envolvendo imagens simples que podem ser giradas. Dessa maneira, SR tem mais a ver com rapidez com que a rotação mental acontece do que com a dificuldade de visualizar e girar uma imagem. SR pode exigir a identificação de uma figura quando vista por diferentes ângulos ou posições (McGrew, 2005; Schneider & McGrew, 2018). O teste Partes da bateria TSP apresenta uma maneira de medir as relações espaciais, pois solicita que o examinando analise partes de uma figura geométrica que estão separadas e, identifique dentre as alternativas (rotadas ou não) a figura completa, considerando o seu tamanho e formato.

2.7 Teste Dimensão

O teste Dimensão é composto por uma série de figuras de pessoas em atividade e de objetos. A tarefa solicita identificar a mudança de posição da figura modelo, localizada à esquerda de uma tarja preta vertical. Quando a figura modelo está virada para o lado esquerdo, o examinando deverá identificar entre as alternativas a figura virada para o lado direito. Somente uma figura entre quatro alternativas possíveis estará invertida (virada ao contrário). O examinando deve assinalar com um X sobre a resposta que apresenta corretamente uma inversão da figura modelo. Os 48 itens poderão ser respondidos em um período máximo de cinco minutos. As respostas são pontuadas em 0 (incorreta) ou 1 (correta) e a pontuação máxima é de 48 pontos (Nascimento & Flores-Mendoza, 2007). Veja abaixo item de treinamento do teste Dimensão.

Olhe bem a figura à sua esquerda. Qual das outras quatro figuras está desenhada de maneira invertida à figura esquerda?



O item de treinamento mostra que todas as alternativas de respostas estão inclinadas mas, apenas uma entre elas está invertida (virada ao contrário) em relação à figura modelo. O item um mostra uma pessoa em atividade virada para o lado direito. A figura de número 2 é a única alternativa que mostra uma pessoa virada para o lado esquerdo (ao contrário), por este motivo ela foi assinalada com um X em cima dela (Nascimento & Flores-Mendoza, 2007).

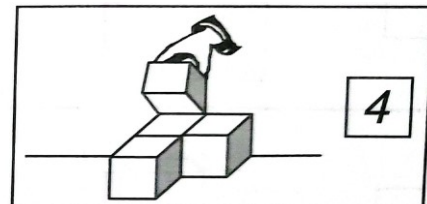
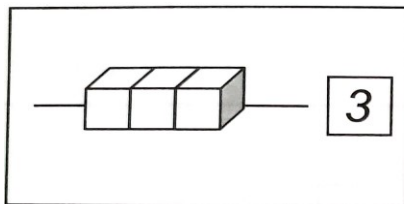
O teste Dimensão tem como objetivo avaliar a aptidão de relação espacial vinculada aos aspectos de motricidade, como por exemplo, a lateralidade. Ele também permite verificar a capacidade de visualização de objetos em duas dimensões (Nascimento & Flores-Mendoza, 2007). Em função do teste Dimensão demandar que o examinando realize uma rotação mental da posição da figura e reconheça, entre as quatro alternativas de resposta, a figura que se encontra em posição invertida à figura modelo, acredita-se que o teste Dimensão, esteja relacionado à habilidade específica Relações espaciais/rotação acelerada (SR), presente no fator de segunda ordem Processamento visual (Gv), do modelo CHC de inteligência, assim

como, o teste Partes da bateria TSP.

2.8 Teste Blocos

O teste Blocos é formado por um conjunto de figuras geométricas que envolvem a decomposição e rotação das partes. Cada um dos 32 itens que compõe o teste apresenta uma pilha de blocos que deverão ser decompostas para que o examinando possa identificar quantos blocos tem a pilha (inclusive os blocos escondidos), em um limite de cinco minutos. O conjunto de blocos de um item tem tamanhos e formatos idênticos, mas de um item para o outro, os blocos podem variar de tamanho e formato. A resposta deverá ser registrada no campo apropriado e são pontuadas em 0 (incorreta) ou 1 (correta), sendo o escore máximo possível é de 32 pontos (Nascimento & Flores-Mendoza, 2007). Veja abaixo o item de treinamento presente no teste.

Preste atenção a pilha de blocos abaixo. Quantos blocos há em cada pilha?



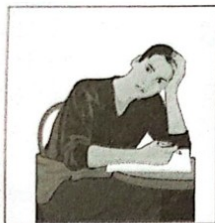
Conforme visualizado no exemplo, o teste demanda do examinando a soma de elementos pela representação mental da figura completa para responder acertadamente. O objetivo do teste Blocos é o de testar o aspecto das relações espaciais ligado à capacidade de estimar quantidades. (Nascimento & Flores-Mendoza, 2007).

Neste caso, acredita-se que o teste Blocos esteja associado ao fator específico Imagens (IM) que compõe a habilidade de segunda ordem Processamento visual (Gv), da teoria CHC de inteligência. A habilidade Imagens é definida como a capacidade de representar (codificar) mentalmente e/ou manipular, por exemplo, objetos, ideias e eventos, de uma forma abstrata, sem que estes estejam realmente presentes (McGrew, 2005; Schneider & McGrew, 2018). O teste Blocos da bateria TSP está em consonância com a definição de IM, pois sua tarefa demanda uma capacidade para formar representações mentais de padrões visuais e utilizá-las para resolver problemas espaciais, exigindo para isso, integração perceptiva de blocos ausentes.

2.9 Teste Memória

No teste Memória o examinando terá dois minutos para memorização das figuras que constam na última página formada por um conjunto de 16 figuras de pessoas, sendo a maioria em situações de trabalho (em preto e branco) com os nomes e sobrenomes. A resolução da folha de exercício que contém 36 figuras deverá ser realizada no período de três minutos. Cada figura tem cinco alternativas de resposta, sendo quatro nomes e sobrenomes e, a quinta opção “não foi mostrado”, ou seja, o respondente deverá sempre assinalar uma alternativa em cada figura (Nascimento & Flores-Mendoza, 2007). Veja a tarefa de treinamento apresentada no teste Memória.

As figuras abaixo devem ser memorizadas quanto aos nomes e rostos (detalhes, em geral).



Ana Morgado



Jaime Silva



Dolores Cunha



Ricardo Bonfim

Abaixo, há quatro figuras com cinco alternativas de resposta cada uma. Caso o examinando tenha visto a figura no momento de memorização (acima), então ela deverá assinalar corretamente o nome da pessoa. Caso ela não tenha sido mostrada, deverá assinalar no campo de resposta a alternativa “não foi mostrado”.



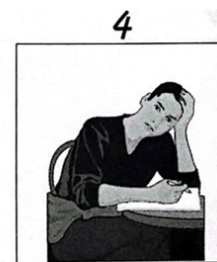
1. Haroldo Abreu
2. Ricardo Bonfim
3. Francisco Elias
4. Jaime Silva
5. Não foi mostrado



1. Margarida Arruda
2. Claudia Aragão
3. Natalia Campos
4. Dolores Cunha
5. Não foi mostrado



1. Ricardo Bonfim
2. Arnaldo Fonseca
3. Jaime Silva
4. Raimundo Morais
5. Não foi mostrado



1. Dolores Cunha
2. Maria Godoy
3. Ana Morgado
4. Nancy Morgado
5. Não foi mostrado

As respostas são pontuadas em 0 (incorreta), 1 (nome e sobrenome corretos) ou 2 (não foi mostrado). A pontuação máxima é de 56 pontos. O testes é o único que apresenta conteúdo verbal e não verbal.

O teste Memória tem como objetivo verificar a aptidão para reter e evocar, de forma direta e imediata, nome, fisionomias e detalhes, no modo de associação. A memorização se dá pelos detalhes em termos de suas diferenças e igualdades (Nascimento & Flores-Mendoza, 2007). No caso do teste Memória, ele pode estar relacionado ao fator de primeira ordem, Memória visual (MV), contido nas habilidades de Processamento visual (Gv), do modelo Cattell-Horn-Carroll (CHC) de inteligência, como ocorreu no Teste de Memória de Mapas (Memória Visual – nomenclatura atual) traduzido, adaptado e validado do *Kit of Factor-Referenced Cognitive Testes* do *Educational Testing Service*, em que foi associado à MV (Gomes & Borges, 2007; Gomes & Borges, 2009a). Memória visual é definida como a capacidade de formar e armazenar uma imagem durante um breve período de estudo e, posteriormente, recuperá-la e reconhecê-la durante o período de resposta a uma demanda (McGrew, 2005; Schneider & McGrew, 2018). Conforme demonstrado pelo item exemplo, este procedimento é abordado na resolução do teste Memória da bateria TSP, em que o sujeito tem um período de estudo e memorização de figuras e nomes de pessoas e, posteriormente, deve evoca-las respondendo a folha teste.

3. OBJETIVOS

3.1 Objetivo geral

Investigar novas evidências de validade e precisão para os Testes de Seleção Profissional (Bateria TSP) e, testar alguns modelos teóricos e alternativos da inteligência para avaliação do modelo mais parcimonioso.

3.2 Objetivos específicos

- Analisar os parâmetros individuais dos itens por meio da Teoria de Resposta ao Item (TRI).
- Buscar evidências de validade baseadas na estrutura interna e nas relações com variáveis externas (critério e convergente).
- Investigar as estimativas de precisão pela consistência interna e estabilidade temporal.
- Avaliar a estrutura interna da Bateria TSP segundo alguns modelos teóricos e alternativos da inteligência.

4. MÉTODO

4.1 Amostra geral

A amostra do processo de atualização da Bateria TSP que resultou no manual aprovado em 2007 foi composta por pessoas entre 16 e 50 anos, com escolaridade mínima cursando o Ensino Médio (Nascimento & Flores-Mendoza, 2007). Os critérios para caracterização da amostra da presente pesquisa foram expandidos, adotou-se a idade a partir de 15 anos como idade mínima e a inclusão de pessoas idosas (60 anos ou mais). A escolaridade dos participantes também foi ampliada. No ano de 2015 foi realizada uma coleta de dados com estudantes do Ensino Fundamental da Educação de Jovens e Adultos (EJA), com idade a partir de 15 anos. Esta coleta foi coordenada e realizada pela autora da presente pesquisa, na ocasião foi bolsista de apoio técnico do projeto de extensão (pesquisa encomendada) intitulado “Normatização da Bateria TSP, Teste de Destreza e MPM para adultos com ensino fundamental”¹. Esse banco de dados foi incluído na composição da amostra do presente estudo.

A coleta de dados para pessoas a partir do Ensino Médio (cursando ou concluído) iniciou em maio de 2019 (após aprovação do comitê de ética em pesquisa), coordenada pela autora da presente pesquisa, em uma parceria de trabalho com a editora Edites. Esta coleta foi interrompida em fevereiro de 2020, em função da pandemia do COVID-2019 e foi retomada no primeiro semestre de 2022 para a ampliação da amostra. Toda a composição da amostra foi por conveniência, considerando a facilidade de acesso da pesquisadora e de parceiros da Edites em Fortaleza/CE e em Brasília/DF.

O planejamento da amostra incluiu um número mínimo de representantes de cada faixa etária nos diferentes subgrupos formados em função do sexo e escolaridade. O tamanho amostral tem como objetivo obter respostas que produzam uma variância satisfatória que permita captar as diferenças individuais (Pasquali, 2010).

Participaram desse estudo 1606 indivíduos, residentes nos estados de Minas Gerais, Espírito Santo, Ceará e Distrito Federal. A amostra contemplou, portanto, as regiões Sudeste (n=1119), Nordeste (n=172) e Centro-oeste (n=312), três respondentes não informaram o local de residência. Os participantes foram homens e mulheres, com idades entre 15 e 78 anos (M=27,17; DP=11,91), com níveis de escolaridade do fundamental ao superior e com diferentes ocupações (Tabela 3).

¹ O projeto foi aprovado pela Câmara do Departamento de Psicologia com registro no SIEX (nº301870).

Tabela 3. Caracterização sociodemográfica da amostra (N=1599)

Faixas etárias	Escolaridade						Total
	Homem			Mulher			
	Fundamental	Médio	Superior	Fundamental	Médio	Superior	
15-16	33	24	00	16	22	00	95
17-18	27	88	07	13	123	36	294
19-25	26	31	139	15	44	368	623
26-39	33	09	47	39	37	129	295
≥40	37	12	27	99	27	91	293
Total	156	164	220	182	253	624	1599

Nota. Sete participantes não informaram a escolaridade e/ou a idade e sexo.

Para investigar a estabilidade temporal, 120 participantes se submeteram aos testes da Bateria TSP em dois momentos. O intervalo entre as duas aplicações variou entre 38 e 89 dias. A amostra alcançada ficou composta por indivíduos com idade entre 18 e 76 anos ($M = 23$ anos; $DP = 7,16$), sendo a maioria do sexo feminino (73,3%) e todos com ensino superior incompleto ou completo.

Para realizar estudos de evidências de validade baseada na relação com outras variáveis também houve uma subamostra de pessoas com ensino superior (curando ou concluído) submetida ao Teste Não Verbal de Inteligência G-38 ($n=27$), à Bateria de Provas de Raciocínio-BPR-5 ($n=84$) e à Bateria Psicológica para Avaliação da Atenção-BPA ($n=17$).

4.2 Instrumentos para a coleta de dados

Cada participante foi submetido à aplicação da Bateria TSP, instrumento foco da pesquisa de doutorado. Conforme informado, para realizar estudos de evidências de validade baseada na relação com outras variáveis foram utilizados os instrumentos de inteligência geral e específica (G-38 e BPR-5) e uma bateria atencional (BPA) em uma subamostra. Considera-se essa seleção de testes adequada para reunir diferentes tipos de evidência para contribuir para melhor interpretação dos escores da Bateria TSP.

As orientações de Muñiz e Fonseca-Pedrero (2019) sobre os passos para a seleção de outros instrumentos de medida foram consideradas para coletar evidências da validade dos resultados. Essa escolha foi definida tanto por questões substantivas quanto pragmáticas, a saber: testes indicados para a população de interesse; apresentam qualidade psicométrica e são aprovados no CFP, pelo SATEPSI; apresentam coerência substantiva para estabelecer relações entre as variáveis estudadas pela Bateria TSP, dentro de sua rede nomológica; como

também, questões econômicas: testes disponíveis ao grupo de pesquisa, possibilidade de aplicação referentes a demanda de tempo e lugar, fatores estes que geraram menor custo. Os instrumentos que foram aplicados na subamostra para as análises de validade serão apresentados brevemente por meio da (1) descrição geral do instrumento, (2) objetivo e (3) análises psicométricas que estão presentes nos manuais aprovados no CFP.

a) Teste Não Verbal de Inteligência G-38

- *Descrição geral:* O livro de exercício contém 38 itens, cada um deles tem uma parte faltando que deve ser resolvida por meio de raciocínios com níveis de dificuldades diferentes. Para cada item existem seis opções de respostas. A avaliação é feita quantitativamente pelo total de acertos e há avaliação qualitativa para o tipo de erro cometido pelo examinando. Trata-se de um teste do tipo papel e caneta, com aplicação individual ou coletiva. Pode ser aplicado em indivíduos de 18 a 57 anos. A última atualização das propriedades psicométricas e atualização das normas foi no estudo aprovado pelo CFP em 31 de agosto de 2018 (Boccalandro, 2018).

- *Objetivo:* avaliar o fator Geral de Inteligência por diferentes raciocínios, como: compreensão de relação de identidade e raciocínios por analogia (tipo numérico, mudança de posição, espacial).

- *Análises psicométricas:* O estudo aprovado na Plenária em 2018 contou com amostras que representaram as regiões Centro-Oeste, Nordeste, Sudeste e Sul do Brasil (n=4.627). Apresenta tabelas normativas para as quatro regiões estudadas e amostra total considerando a característica escolaridade em função dos resultados demonstrando diferença estatisticamente significativa em função desse critério, no qual as pessoas com Ensino Superior tendem a obter melhores pontuações quando comparadas às pessoas com Ensino Médio. Os índices de precisão foram avaliados pelo método da consistência interna (alfa de *Cronbach*) para cada região e amostra total considerando o critério escolaridade, o coeficiente mínimo encontrado foi 0,80 e o valor máximo 0,92. O método das metades com correção Spearman-Brown também foi utilizado para avaliação da consistência interna do instrumento. Foram feitos estudos por região e com o critério de escolaridade, os resultados foram entre 0,81 e 0,92. Todos os índices encontrados atestam para a precisão do instrumento de acordo com a Resolução 31/2022 (Boccalandro, 2018).

b) Bateria de Provas de Raciocínio – BPR-5

- *Descrição geral:* A primeira versão da BPR-5 adaptada para o Brasil foi publicada no ano

de 2000 (com aprovação pelo CFP em 2003). A bateria possui duas formas: Forma A, para alunos de sétimo a nono ano do Ensino Fundamental e Forma B, para alunos da primeira à terceira séries do Ensino Médio. O instrumento é composto por cinco subtestes de raciocínio, é um teste do tipo objetivo e de caneta e papel, com aplicação coletiva ou individual, com possibilidade de correção informatizada. Fundamenta-se em aspectos teóricos consistentes e mais recentes da inteligência (modelo CHC de inteligência). Avalia a capacidade de raciocínio geral e habilidades específicas: raciocínio numérico (RN), raciocínio mecânico (RM), raciocínio espacial (RE), raciocínio abstrato (RA) e raciocínio verbal (RV). Pode ser feita a aplicação completa da bateria ou dos raciocínios específicos de maneira isolada, de acordo com a necessidade de investigação. A última atualização das propriedades psicométricas foi no estudo aprovado pelo CFP em 11 de abril de 2003 e a atualização das normas foi aprovada na Plenária de 19 de maio de 2018 (Primi & Almeida, 2018; Primi, Correia, & Almeida, 2018).

- *Objetivo*: O subteste RA está relacionado à inteligência fluída (Gf - I) que diz respeito a capacidade de o indivíduo raciocinar frente a situações problema que independem de experiências prévias. O subteste RM se relaciona com conhecimentos práticos mecânicos e processamento visual. O subteste RN se relaciona a habilidade de análise quantitativa, memória de trabalho e processamento visual. O subteste RE associa-se à capacidade de processamento visual (Gv - Vz) e em parte com a Gf (I). O subteste RV associa-se com a inteligência cristalizada (Gc - K0), relacionado ao conhecimento verbal e conceitos aprendidos previamente (Campos, Zaia, & Primi, 2019; Primi & Almeida, 2018; Primi, Nakano, & Wechsler, 2012).

- *Análises psicométricas*: Na recente pesquisa de atualização das normas da BPR-5, foi possível contar com dados armazenados pela editora em função da correção informatizada que possibilita ao psicólogos usuários a autorizar a utilização dos dados em pesquisa. Isso favorece para a possibilidade de acessar diferentes contextos e finalidades de aplicação do instrumento. A precisão evidenciada pelos coeficientes de consistência interna avaliados pela fórmula de Lâmbda 6 de Guttman foi apresentado por idade e variaram entre 0,78 a 0,95 (RA), 0,84 a 0,94 (RE), 0,76 a 0,91 (RM), 0,88 a 0,98 (RN) e, 0,72 a 0,95 (RV) entre as formas A e B estudadas, com idade mínima de 12 e máxima de 23 anos. Os coeficientes apresentados demonstram uma alta consistência dos escores (Primi & Almeida, 2018). Primi, Correia e Almeida (2018) apresentam um capítulo sobre a BPR-5 e agrupam diferentes resultados para a precisão, evidências de validade baseadas no conteúdo, na estrutura interna e na relação com outras variáveis com resultados favoráveis quanto à qualidade do instrumento.

Desde sua adaptação para o Brasil, a BPR-5 é um dos instrumentos mais utilizados na avaliação das capacidades cognitivas e diversos estudos foram realizados apontando as suas qualidades psicométricas (Júnior, Nascimento, & Roazzi, 2019; Primi & Almeida, 2018; Primi & Almeida, 2000)

c) Bateria Psicológica para Avaliação da Atenção (BPA)

- *Descrição geral:* A bateria é composta por três testes que avaliam capacidades específicas da atenção: atenção concentrada (AC), atenção dividida (AD) e atenção alternada (AA), que devem ser aplicados nesta sequência. Pela análise dos três testes é possível obter o resultado geral da atenção. Nas folhas de aplicação são apresentados estímulos abstratos, com igual quantidade em cada instrumento, sendo 20 linhas e 20 estímulos distribuídos em cada teste, totalizando 400 quadradinhos com desenhos abstratos produzidos em preto e branco. Cada folha de aplicação possui 120 estímulos-alvo e 280 distratores, ou seja, o escore bruto máximo em cada teste é de 120 pontos e, para o resultado da atenção geral, 360 pontos (AC+AD+AA). O resultado em cada teste é obtido pelo total de figuras-alvo assinaladas corretamente, menos os erros e as omissões ocorridas durante a produção de respostas do indivíduo. É indicado ao examinando que inicie os testes do lado esquerdo da folha e siga em direção ao lado direito, ao final da linha, passe para a linha de baixo e continue a responder. O teste de atenção concentrada apresenta no início da folha uma única imagem que deverá ser assinalada com um traço quando for visualizada. Em cada linha ímpar existem sete alvos e 13 distratores e para cada linha par, cinco estímulos-alvos e 13 distratores. O tempo de resposta é de dois minutos. O teste AC diz respeito a capacidade de o indivíduo selecionar uma única informação e inibir os estímulos não correspondentes dentro de um período de tempo predeterminado. O teste de atenção dividida apresenta no início da folha os três estímulos-alvo que deverão ser procurados simultaneamente, sem que precisam estar na sequência mostrada, dessa maneira, cada vez que um estímulo-alvo aparecer é necessário realizar um traço em cima dele. Em cada linha existem seis figuras que devem ser assinaladas e 14 distratores. O tempo de aplicação é de quatro minutos. O teste AD avalia a capacidade de o indivíduo procurar dois ou mais estímulos de maneira concomitante e inibir os distratores presentes, em um período predeterminado. O teste de atenção alternada apresenta no início de cada linha uma figura-alvo, ou seja, para cada linha tem uma figura diferente a ser procurada e assinalada. Em cada linha ímpar existem cinco alvos e 15 distratores e para cada linha par são sete figuras-alvo e 13 distratores. O tempo de aplicação é de dois minutos e 30 segundos. O teste AA diz respeito a capacidade de o indivíduo focar sua atenção e selecionar ora um

estímulo, ora outro, com a presença de vários distratores, em um tempo predeterminado. Trata-se de um teste do tipo papel e caneta, com aplicação individual ou coletiva. Pode ser aplicado em indivíduos de 6 a 82 anos (Rueda, 2013).

- *Objetivo*: avaliar a capacidade geral de atenção, assim como, três diferentes tipos atencionais específicos, a saber, atenção concentrada, atenção dividida e atenção alternada.

- *Análises psicométricas*: A última atualização das propriedades psicométricas foi no estudo aprovado pelo CFP em 31 de janeiro de 2013 e a atualização das normas foi aprovada na Plenária de 09 de julho de 2021, com vigência até 09 de julho de 2036. Ressalta-se que atualmente existe a BPA-2 (segunda edição), mas na ocasião da coleta o teste não era comercializada, por este motivo utilizou-se a primeira edição válida. No estudo aprovado em 2013, participaram 1759 sujeitos com escolaridade referente a Educação de Jovens e Adultos (EJA) e Ensino Superior. Evidências de validade de critério foram obtidas pelas variáveis idade e escolaridade, em que os resultados foram coerentes com a literatura da área. Evidências de validade convergente foram realizadas a partir da correlação com os testes Atenção Concentrada (TEACO-FF), Atenção Dividida (TEADI) e Atenção Alternada (TEALT), valores de correlação encontrados foram entre 0,51 e 0,66, indicando evidências de validade convergente entre os testes e os testes atencionais da BPA. Evidências de validade com construtos relacionados foram obtidas entre a BPA e o Teste Não Verbal de Inteligência (R1). A precisão do teste foi evidenciada pelo método teste e reteste, com resultados de correção variando entre 0,68 e 0,89, indicando satisfatória estabilidade temporal do instrumento (Rueda, 2013).

Os participantes preencheram também um questionário para levantamento de informações sociodemográficas, de saúde e dados ocupacionais. Com esta ficha é pretendida uma breve caracterização dos indivíduos submetidos à pesquisa e realizar estudos de validade com base em alguns critérios considerados no último estudo realizado com o teste: escolaridade, sexo e faixa etária (Nascimento & Flores-Mendoza, 2007). A seleção adequada de outras variáveis de interesse permite maior qualidade dos estudos a partir do qual é possível aprofundar sobre o significado das pontuações obtida pelo teste (Muñiz & Fonseca-Pedrero, 2019).

4.3 Procedimentos para a coleta de dados

Em relação aos aspectos éticos referentes à pesquisa envolvendo seres humanos, o presente estudo foi submetido à avaliação do Comitê de Ética em Pesquisa da Universidade

Federal de Minas Gerais (COEP/UFMG) e, recebeu parecer favorável sob o número 3.358.718 (Apêndice A). A Editora que detém os direitos autorais da Bateria TSP autorizou a condução dos estudos para atualização das propriedades psicométricas do teste objeto de estudo da presente pesquisa.

Para o desenvolvimento da coleta de dados, a pesquisadora responsável formou uma equipe de pesquisa composta por estudantes de Psicologia e psicólogas formadas que foram treinadas para a aplicação, correção dos testes e lançamento dados em um banco. Essa equipe foi responsável por entrar em contato com as instituições a fim de obter autorização formal para a realização dos procedimentos descritos neste projeto junto aos estudantes (Apêndice B). Participaram escolas com ensino fundamental e médio estaduais e municipais, instituto federal, escolas técnicas e instituições de ensino superior público e privado.

Uma vez autorizado o contato com os estudantes das instituições, a equipe esclareceu sobre o objetivo, a natureza e a relevância do trabalho a ser desenvolvido, assim como, sobre o sigilo do material coletado e, por fim, os convidou para participar do estudo. Aqueles que concordaram preencheram uma ficha de dados para caracterização (Apêndice C), assinaram o Termo de Consentimento Livre e Esclarecido (TCLE – Apêndice D) quando maiores de 18 anos e, responderam aos instrumentos. No caso de participantes menores de idade, estes foram submetidos à pesquisa somente após a autorização dos pais por meio da assinatura TCLE (Apêndice E) e do Termo de Assentimento Livre e Esclarecido (TALE – Apêndice F) para participação de adolescentes em pesquisa. Ressalta-se que toda a participação na pesquisa foi voluntária e podia ser interrompida a qualquer momento pelo participante.

A aplicação dos instrumentos foi coletiva, realizada em uma ou duas sessões de acordo com a disponibilidade oferecida pela instituição e dos participantes. Os procedimentos de aplicação seguiram as recomendações internacionais para testes educacionais e psicológicos (AERA, APA, & NCME, 2014) e as orientações de Muñiz e Fonseca-Pedrero (2019), entende-se que seguir um procedimento padrão contribui para que todos os participantes tenham iguais condições para demonstrar suas competências.

Dessa maneira, a equipe responsável pela aplicação foi treinada quanto às condições físicas adequadas para a aplicação, por exemplo, conforto do local, luz, controle de ruído e interrupções; necessidade de estabelecer um bom relacionamento com os participantes; seguir rigorosamente as diretrizes de padronização da administração dos instrumentos contidas no manual, como: dar as instruções corretamente, exemplificar com clareza como as questões são resolvidas de acordo com os itens de exemplo e treino presentes em cada teste da Bateria; supervisionar a administração e minimizar possíveis fontes de erro. Ressalta-se que para os

participantes do EJA, em função de uma maior e significativa dificuldade no entendimento das instruções quando comparado aos participantes de maior escolaridade, houve a necessidade de adaptar as instruções para ajudar a garantir que todos os participantes entendessem o que se esperava deles. Esses casos também são previstos nas recomendações internacionais, em que se faz necessário entender que mesmo com as alterações o objetivo permanece o mesmo, que é o de ter medições precisas, justas e comparáveis, na busca de manter um tratamento justo, não discriminatório e garantir a equidade (AERA, APA, & NCME, 2014).

4.4 Procedimentos para a análise dos dados

Os dados coletados foram organizados em planilha para serem devidamente processados e analisados. O banco de dados do público EJA foi novamente digitado para atender aos critérios de análise e inclusão na pesquisa atual. Posteriormente, foram realizadas análises utilizando os *softwares* JAMOVI – versão 2.3 (2022), JASP – versão 0.18.3. (2024) e pela linguagem R de programação (R Core Team, 2022). Em função dos dados terem sido digitados por diferentes pessoas da equipe, a pesquisadora seguiu a recomendação de Muñiz e Fonseca-Pedrero (2019) para verificação da qualidade do lançamento. Para isso, utilizou-se da estratégia de extrair aleatoriamente uma determinada quantidade de participantes e examinou a correspondência entre as pontuações anotadas no teste físico e o dado que foi digitado. O resultado demonstrou a qualidade do banco de dados, em função da correspondência exata das pontuações investigadas dos participantes.

Para avaliar a forma como os dados se distribuem, foi realizado o teste de regressão/correlação Shapiro-Wilk (W). Este teste verifica o ajuste dos dados para distribuição normal, em que espera-se um valor de $p > 0,05$ e o valor de W próximo de 1,0 para afirmar o pressuposto da normalidade. Os valores da curtose e assimetria consideram os índices centrados no valor zero para evidência de uma distribuição normal. Foi feito também um teste gráfico pelo diagrama quantil-quantil (Q-Q), nele é feita uma comparação das proporções dos dados ordenados da amostra e os quantis para uma distribuição normal padrão. Observa-se uma distribuição normal pelo diagrama Q-Q quando os pontos representantes da amostra situam-se ao longo de uma linha diagonal reta. Dados não seguem uma distribuição normal são detectados quando for observada uma distribuição em curva S ou em formas de banana. A curva de Gauss, também conhecida como distribuição normal, também foi analisada. Ressalta-se a importância de se conciliar estratégias de julgamento diferentes para a

verificação da suposição da distribuição dos dados e tornar essa interpretação mais precisa. Isso se faz relevante em função do teste estatístico mostrar-se sensível ao erro tipo II na situação de amostras grandes, como na presente pesquisa. Desse maneira, conciliar com um métodos de inspeção visual pode tornar o julgamento mais confiável. A escolha do teste estatístico Shapiro-Wilk foi por demonstrar, entre os testes formais, um alto grau de discriminação e por ser bem apoiado na literatura. Os diagramas quantil-quantil, em comparação com outros métodos de inspeção (histogramas e gráficos) demonstram maior acurácia e são mais indicados para amostras grandes, contrapondo os testes estatísticas de normalidade mais suscetíveis ao erro tipo II (Bishara, Li, & Conley, 2021; Henderson, 2006; Miot, 2017; Vetter, 2017).

Foram conduzidas análises pela Teoria de Resposta ao Item, em que os dados foram analisados pela linguagem R de programação (R Core Team, 2022), na qual utilizou-se das análises dos parâmetros individuais dos itens, pelo pacote *Multidimensional Item Response Theory* - MIRT (Chalmers et al., 2015). Investigou-se os parâmetros de discriminação (a), dificuldade (b) e acerto ao acaso ou ‘chute’ (c). A discriminação é a capacidade de um item diferenciar sujeitos com diferentes níveis de traço latente (*theta* - θ); enquanto a dificuldade está relacionada ao tamanho de *theta* necessário para as pessoas acertarem um determinado item (Sousa & Braga, 2020); já o parâmetro de chute está relacionado a probabilidade de um sujeito com nível de θ muito baixo acertar um item (Andrade, Laros, & Gouveia, 2010). Foi considerado o pressuposto da unidimensionalidade da TRI para as análises (Pasquali, 2020).

Quanto à interpretação dos parâmetros individuais dos itens, Pasquali (2020) sugere faixas para a avaliação da discriminação de um item, sendo entre 0,01 e 0,64 indicando discriminação baixa; valores entre 0,65 e 1,34 são moderados; enquanto valores entre 1,35 e 1,69 são altos; por fim, valores acima de 1,70 são consideradas discriminações muito altas. O parâmetro de dificuldade apresenta a mesma métrica do nível de *theta*, mesmo que na prática seus valores sejam medidos entre -3 e +3, geralmente (Pasquali & Primi, 2003), seus valores estão associados a probabilidade de 50% de um sujeito acertar o item (Couto & Primi, 2011; Peres, Lessa, & Valentini, 2019). O parâmetro (c) deve ser interpretado como a probabilidade de acerto ao acaso, por isso, sua métrica varia entre 0 e 1 (de 0% a 100%), sendo que para itens dicotômicos, valores abaixo de 0,5 indicam baixa probabilidade de acerto ao acaso (Andrade, Laros, & Gouveia, 2010).

A TRI revela que a confiabilidade do teste varia entre diferentes grupos de sujeitos, indicando para qual grupo os itens do instrumento funcionam de maneira mais eficaz. A Curva de Informação do Teste (CIT) indica o quão útil o instrumento é para diferentes níveis

de traço latente. A CIT foi avaliada para oito dos testes da Bateria TSP, com o objetivo de identificar os intervalos de habilidade para os quais os testes são mais úteis, bem como determinar em qual faixa de *theta* o teste fornece pouca informação, gerando mais erro para o resultado (Pasquali, 2020).

O suporte empírico é mais forte quando são conduzidos diferentes estudos psicométricos em que são apresentadas diferentes estratégias de evidência de validade. Dois métodos mais populares para a investigação da validade dos testes são os de estrutura interna e o da relação de pontuações em diferentes instrumentos, com medidas teoricamente relacionadas (validade convergente) e/ou teoricamente divergentes (validade discriminante). Na situação de investigação fazendo uso da análise fatorial é possível verificar a dimensionalidade do instrumento, mas essa estratégia não aponta sobre o que o teste está medindo, apenas detecta os itens que se agrupam. Uma abordagem complementar seria testar o instrumento fazendo uso da validade convergente, pois ao correlacionar os resultados do teste objeto de estudo a um outro instrumento de qualidade estabelecida que avalia o mesmo construto, tem-se uma evidência sobre o construto pretendido (Kalkbrenner, 2021). Considerando o exposto, a presente pesquisa buscou métodos multifacetados para a validação do construto, pela estrutura interna e a investigação pelos resultados dos sujeitos em outros instrumentos.

A investigação da estrutura interna se deu por meio de análises fatoriais confirmatórias (AFC). O modelo testado em cada teste foi o unidimensional. Índices de ajuste preconizados na literatura foram inspecionados, assim como as cargas fatoriais e os índices de modificação (Brown, 2006). A estatística raiz quadrada média do erro de aproximação (*Root Mean Square Error of Approximation* - RMSEA) e a estatística raiz quadrática média Residual Estandartizada (*Standardized Root Mean Square Residual* - SRMR) apontam a qualidade de ajuste ao estimar os parâmetros do modelo hipotético à matriz de covariância (Pilati & Laros, 2007). Valores menores que 0,05 indicam um bom ajuste e, entre 0,06 e 0,08 ajuste razoável dos dados (Marsh, Hau, & Wen, 2004). Os índices relativos de ajuste *Comparative Fit Index* (CFI) e o de Tucker-Lewis Index (TLI) são índices comparativos entre o modelo hipotético e os dados observados e o índice *Goodness-of-fit Index* (GFI ou índice de qualidade do ajuste) são análises que estão próximas à complexidade do modelo. Seus valores variam entre 0 e 1, considera-se que quanto mais próximo de 1 melhor ajuste e, valores acima de 0,95 indicam um excelente ajuste para TLI, CFI e GFI, sendo o índice de 0,90 aceitável (Hu & Bentler, 1998; Morin, Arens, & Marsh, 2016).

Foram realizados estudos de validade convergente entre os testes da Bateria e os testes de inteligência (G-38 e BPR-5) e atenção (BPA). Foi utilizado o r de *Pearson* para as análises correlacionais e o ρ de Spearman para o teste Memória em função da não normalidade dos dados. A magnitude foi avaliada segundo os critérios de Dancey e Reidy (2019), espera-se resultados superiores a 0,40 para a evidência de validade convergente.

Para avaliar as evidências de validade de critério foram investigadas diferenças entre o desempenho nos testes e as variáveis sociodemográficas (sexo, faixa etária e escolaridade). As interpretações foram fundamentadas na literatura sobre habilidades cognitivas. As suposições foram as seguintes: 1) as médias de desempenhos entre homens e mulheres seriam semelhantes (Burgaleta et al., 2012; Flores-Mendoza, 2000); 2) as médias de desempenho difeririam entre os grupos etários (Keith et al., 2008; Tucker-Drob, Brandmaier, & Lindenberger, 2019); 3) as médias de desempenho entre os grupos difeririam conforme o nível de escolaridade (Kaufman et al., 2009).

A variável idade foi agrupada em cinco faixas etárias distintas: 15 a 16 anos; 17 a 18 anos; 19 a 25 anos; 26 a 39 anos e ≥ 40 anos em todos os testes da Bateria TSP. Quanto à escolaridade, foram considerados os níveis de ensino fundamental, médio e superior. Como os grupos possuíam números diferentes de participantes, as análises segundo a faixa etária e a escolaridade foram realizadas utilizando a correção *Welch* para a ANOVA e as comparações *post hoc* que não apresentaram homogeneidade de variância no teste de *Levene* ($p > 0,05$), foi utilizada a correção pelo teste *Games-Howell*. Para o teste Memória foram utilizadas análises não paramétricas, o teste U de Mann Whitney (Kerby, 2014) para comparação entre os sexo e ANOVA com *Bootstrapping* para estudo das variáveis idade e escolaridade (Haukoos & Lewis, 2005). As estatísticas usadas para comparação entre médias incluíram o Teste *t* de *Student* e a ANOVA, com nível de significância ($p < 0,05$) e considerado um tamanho de efeito médio ou grande ($d > 0,50$, $\eta^2 > 0,02$) (Cohen, 1988; Durlak, 2009).

O coeficiente alfa de Cronbach é uma das medidas de consistência interna mais amplamente utilizada. Apesar disso, tem sido questionado em função de subestimar a estimativa de confiabilidade, principalmente, em função dos pesquisadores o aplicarem de maneira indiscriminada, sem verificar as condições necessárias dos dados para o seu uso. Como exemplo, a suposição de *tau-equivalência*: cada item deve contribuir de igual maneira para o escore composto total, ou seja, todos os itens devem possuir cargas fatoriais iguais (covariância e intercorrelações) para que o alfa seja utilizado e tenha uma representação correta da precisão do instrumento. A estimativa ômega (Ω) de McDonald pode ser uma escolha viável quando a condição de *tau-equivalência* não é alcançada, pois possui a mesma

capacidade de estimativa que o alfa, no entanto, por seu modelo estatístico ser *cogênico* permite que a associação em cada item e a escala total varie (Hauck-Filho & Valentini, 2020; Hayes & Coutts, 2020; McNeish, 2018; Taber, 2018). Em função disso, a verificação das evidências de precisão foi realizada pelo método de consistência interna por meio dos coeficientes alfa de Cronbach e ômega (Ω) de McDonald, resultados superiores a 0,60 serão considerados adequados. Outra evidência será pelo estudo da estabilidade temporal (teste-reteste), verificado pelo coeficiente de correlação de *Pearson* para os testes com distribuição normal e o ρ de Spearman para o teste Memória, para comparação entre o desempenho dos sujeitos nas duas aplicações. A magnitude das correlações foi avaliada segundo os critérios de Dancey e Reidy (2019).

Na Tabela 4 são apresentadas as análises realizadas para o estudo das propriedades psicométricas dos testes da Bateria TSP. Nela são descritas, resumidamente, o que se pretendeu investigar, a estratégia estatística escolhida, o programa estatístico utilizado para as análises e os resultados esperados.

Tabela 4. Síntese dos procedimentos de análise de dados

Investigação	Estratégia estatística	Programa	Resultado esperado
Características da amostra	Frequência; Medidas de tendência central e dispersão;	JAMOVI	
Distribuição dos dados	Assimetria; Curtose; W de Shapiro-Wilk; Quantil-Quantil; Curva de Gauss	JAMOVI JASP	Próximo a zero W próximo a 1; $p < 0,001$ Diagrama em linha reta
Precisão (consistência interna)	alfa de Cronbach ômega (Ω_t) de McDonald	JAMOVI	$\geq 0,70$
Estabilidade temporal	Correlação simples (<i>Pearson</i>) ρ de Spearman	JAMOVI	$\geq 0,60$
Validade (estrutura interna)	Análise Fatorial Confirmatória	JAMOVI JASP	Unifatorial
Validade convergente	Correlação simples (<i>Pearson</i>) ρ de Spearman	JAMOVI	$\geq 0,40$
Validade de critério (sexo, idade, escolaridade)	Teste t de Student Teste U de Mann Whitney ANOVA (<i>Bootstrapping</i>) Índice d	JAMOVI JASP	Diferença significativa para todos os testes em função da escolaridade; variável sexo para o teste Números, Dimensão, Partes, Blocos e Fluência; Para a faixa etária nos testes com conteúdo verbal
Parâmetros individuais dos itens (TRI)	Discriminação Dificuldade Acerto ao acaso ou 'chute'	R (MIRT)	$\geq 0,65$ entre -3 e +3 $< 0,50$

4.5 Modelos teóricos e alternativos da inteligência investigados

Para investigar os modelos teóricos e alternativos da inteligência, foram empiricamente testados oito estruturas distintas. Cada modelo construído segue parâmetros específicos, podendo ser: 1) considerado todos os testes da Bateria TSP individualmente; 2) agrupando alguns dos testes com base na característica psicológica comum avaliada; 3) agrupando os testes com base na natureza do estímulo, verbal ou não verbal. Todos os modelos foram testados em sua validade estrutural usando análise fatorial confirmatória pela seleção de itens dos instrumentos.

O primeiro modelo testado considera a sua teoria de origem, ou seja, o modelo das habilidades mentais primárias de Thurstone. O modelo 1 é composto por nove fatores correlacionados, cada um explicando os itens dos nove testes da Bateria TSP em separado (Figura 1). O modelo 2 testado, foi um modelo alternativo ao modelo 1 (Figura 2). Consistiu em seis fatores correlacionados da seguinte forma: F1 explicando os itens dos testes Partes, Blocos e Dimensão; F2 os itens de Memória; F3 os itens de Julgamento; F4 os itens de Fluência; F5 os de Números; e, F6 explicando os itens dos testes Precisão e Percepção. Os agrupamentos realizados para F1 e F6, consideraram que os testes avaliam um mesmo construto cognitivo, habilidade espacial e habilidade perceptual, respectivamente. Os demais testes não foram agrupados devido às suas habilidades específicas distintas.

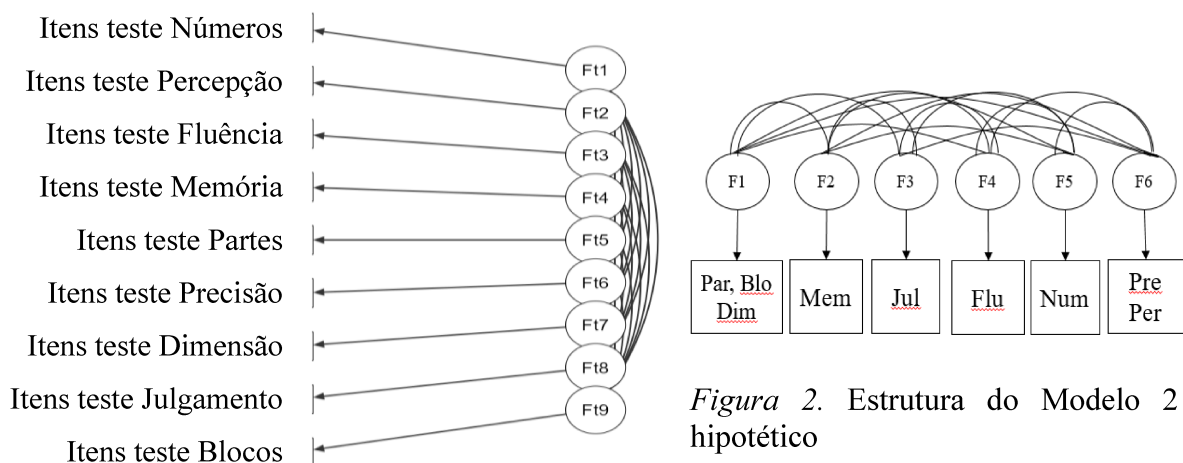


Figura 1. Estrutura do Modelo 1 hipotético

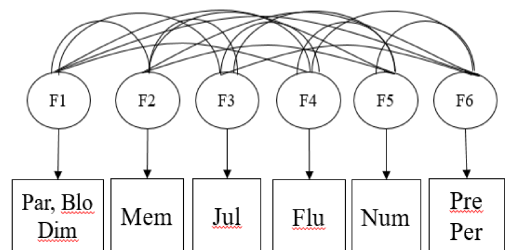


Figura 2. Estrutura do Modelo 2 hipotético

O modelo 3 foi formulado a partir da teoria de Carroll, no qual um fator geral foi adicionado para explicar a variância dos nove testes individualmente (Figura 3). Enquanto que, o modelo 4 é um modelo alternativo ao modelo 3 (Figura 4), no qual um fator geral

explica seis habilidades. Neste modelo, alguns testes foram novamente agrupados com base nas habilidades comuns avaliadas, espacial (F1) e perceptual (F6).

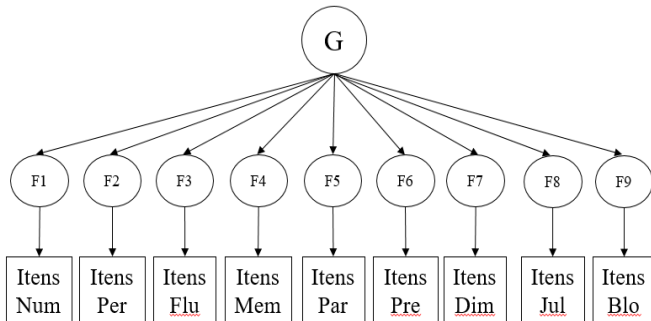


Figura 3. Estrutura do Modelo 3 hipotético

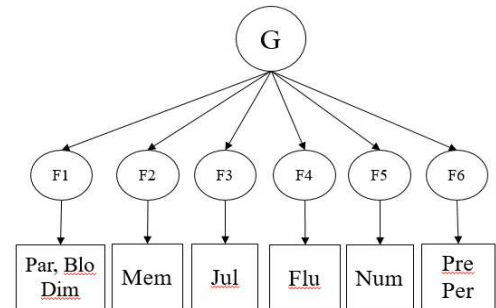
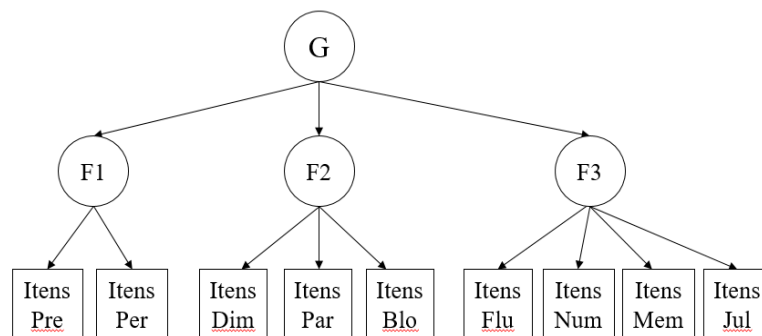


Figura 4. Estrutura do Modelo 4 hipotético

O modelo 5 foi elaborado com base em uma estrutura hierárquica da inteligência (Figura 5). Foram propostos três fatores: F1 (Percepção e Precisão), F2 (Dimensão, Partes e Blocos) e F3 (Fluência, Números, Memória e Julgamento). Foi considerado F1 habilidades perceptuais, para F2 habilidade de rotação de imagem e, para o F3, testes que possuem como fator comum a natureza do estímulo verbal. Todos os fatores de segunda ordem são explicados por um fator geral. Essa estrutura de inteligência apresenta semelhanças com o modelo VPR proposto por Johnson e Bouchard.

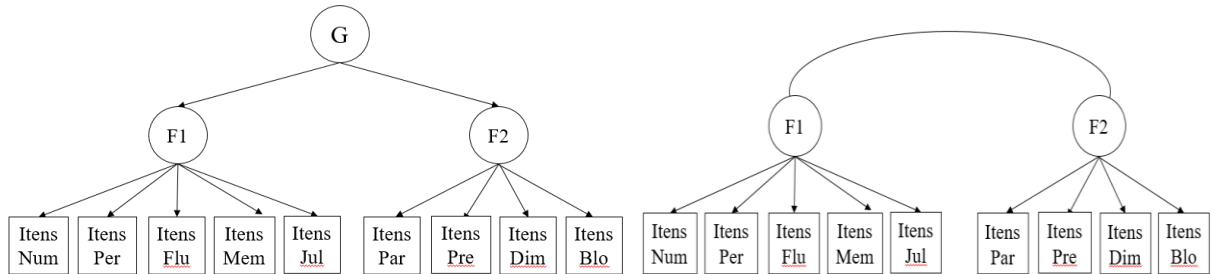


Nota: F1: Perceptual; F2: Rotação de imagem; F3: Verbal

Figura 5. Estrutura do Modelo 5 hipotético

Para os modelos 6, 7 e 8, foram considerados dois fatores, cada um agrupado com base na natureza do estímulo dos testes: aqueles com componente verbal foram alocados em F1 (Números, Percepção, Fluência, Memória e Julgamento), enquanto os testes com estímulo não verbal foram colocados em F2 (Partes, Precisão, Dimensão e Blocos). O modelo 6 (Figura 6) apresenta uma estrutura hierárquica semelhante ao modelo verbal-perceptual de Vernon, com

dois fatores amplos: F1 (V:ED) e F2 (K:M), ambos explicados por um fator geral. O modelo 7 (Figura 7) sugere uma estrutura semelhante à teoria de Cattell e Horn, com dois fatores gerais correlacionados (Gc e Gf).



Nota: F1 (V:ED); F2 (K:M)

Nota: F1: Gc; F2: Gf.

Figura 6. Estrutura do Modelo 6 hipotético

Figura 7. Estrutura do Modelo 7 hipotético

O modelo 8 (Figura 8) representa uma abordagem alternativa projetada para analisar a variância explicada pelo fator geral e pelos fatores específicos. O modelo segue uma estrutura bifatorial da inteligência, com dois fatores de primeira ordem (F1 e F2) e um fator geral que diretamente explicam a variância dos itens. Todos os fatores são ortogonais entre si. Essa estrutura permite uma análise mais detalhada da relação entre o fator geral e os fatores específicos, oferecendo uma visão mais abrangente de cada componente da inteligência em diferentes tarefas cognitivas.

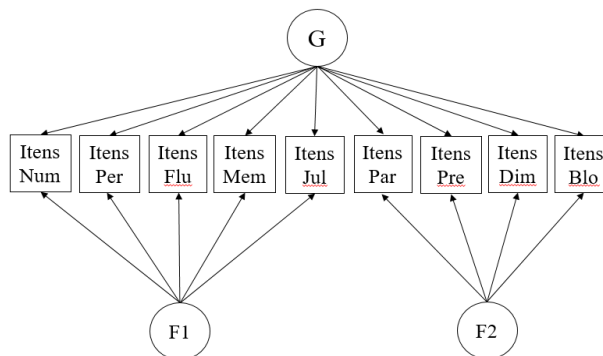


Figura 8. Estrutura do Modelo 8 hipotético

Os modelos 5, 6, 7 e 8 (Figuras 5 a 8) puderam ser testados tanto pela seleção dos itens de cada teste, quanto pelos escores somados dos testes individualmente. Embora houvesse interesse em testar um modelo semelhante à teoria CHC de inteligência, isso não foi viável devido à quantidade limitada de instrumentos disponíveis na Bateria TSP.

4.6 Estratégias analíticas para investigação da estrutura de inteligência da Bateria TSP

Em função da grande quantidade de dados faltantes no banco de dados, antes da realização das análises foi necessário um processo de imputação visando minimizar a perda de informações relevantes. Nos casos em que o indivíduo respondeu apenas a um item de um determinado teste, os demais itens foram codificadas como zero. Apenas os casos em que o respondente não respondeu a nenhum item de um teste específico foram mantidos como dados faltantes. Esta estratégia não implicou na remoção direta de linhas, permitindo que alguns respondentes tivessem dados faltantes em alguns testes, mas não em outros. Assim, a Tabela 5 mostra o número de respondentes contabilizados para cada um dos testes.

Tabela 5. Frequência de sujeitos por teste da Bateria TSP

Fluência	Números	Julgamento	Percepção	Partes	Dimensão	Precisão	Blocos	Memória
1263	1260	1305	1278	1304	1274	1243	1286	1276

As AFCs foram realizadas com um total de 1465 respondentes, especificando-se na função do R que a imputação fosse *pairwise*. Este método de imputação foi utilizado em função de aproveitar todos os dados disponíveis para cada par de variáveis, calculando correlações apenas com os casos não faltantes de cada par. Dessa forma, fica viável realizar a AFC com um número total de respondentes superior ao observado nos testes individuais, pois esse método maximiza a utilização das respostas disponíveis. Com o objetivo de escolher o estimador apropriado para as AFCs, a normalidade multivariada foi avaliada pelo teste de Mardia, por meio do pacote *Multivariate Normality Tests* (MVN), versão 5.9 (Korkmaz, Goksuluk, & Zararsiz, 2021).

Para avaliar o ajuste de cada modelo, os índices de ajuste foram inspecionados e os modelos são considerados aceitáveis se apresentarem CFI ou TLI $\geq 0,90$ (Hu & Bentler, 1998; Morin et al., 2016) e RMSEA ou SRMR $< 0,08$ (Marsh, Wen, & Hau, 2004). O ajuste dos modelos é avaliado por meio da comparação da diferença dos qui-quadrados e os graus de liberdade, utilizando o método Satorra (2000). Todas as análises foram realizadas utilizando o software R, versão 4.2.2 (R Core Team, 2022). Para realizar a análise fatorial confirmatória, utilizou-se o pacote *lavaan*, versão 0.6.7 (Rosseel et al., 2023).

Na avaliação da confiabilidade dos fatores, foi utilizado o alfa de Cronbach, o ômega de McDonald e a confiabilidade composta. No entanto, para fatores de segunda ordem, o ômega hierárquico foi examinado em detrimento do ômega de McDonald. Índices iguais ou

superiores a 0,60 são tidos como confiáveis. Especificamente para o modelo bifatorial, consideram-se aceitáveis valores de ômega e da confiabilidade composta a partir de 0,50. Tal flexibilização se justifica pelo fato desses índices considerarem as cargas fatoriais e, devido à presença de um fator geral na explicação da variância dos itens que gera uma subestimação dos valores reais desses índices (Gomes, Golino, Souza Peres, 2020; Valentini et al., 2015). Para o cálculo dos índices de confiabilidade, empregou-se o *semTools*, versão 0.5-5 (Jorgensen et al., 2021).

5. RESULTADOS E DISCUSSÃO DAS PROPRIEDADES PSICOMÉTRICAS DOS TESTES DA BATERIA TSP

Optou-se por apresentar os resultados para cada teste individualmente, pois o manual da Bateria TSP permite o uso dos instrumentos de forma separada. Uma avaliação detalhada dos parâmetros de cada teste possibilita uma análise crítica mais relevante do que examinar os resultados da Bateria em conjunto. Ressalta-se que a maioria da amostra respondeu a todos os nove testes que compõem a Bateria TSP, mas houve casos em que os participantes responderam apenas a alguns testes selecionados. Esse fato justifica algumas diferenças encontradas nas características e nos tamanhos amostrais.

Outro ponto que precisa ser observado diz respeito a uma discrepância entre o tamanho total da amostra e o número de respostas analisadas na seção de resultados da TRI. Isso ocorre em função de alguns dados de participantes com ensino fundamental, terem sido registrados apenas o escore bruto, sem lançar as respostas para cada item. Assim, resulta em uma quantidade maior de sujeitos na amostra total em comparação com a dos itens analisados pela TRI. Também é importante destacar que, para todos os testes da Bateria TSP, é possível pular um item, o que justifica uma diminuição não perfeitamente linear na quantidade de respondentes para cada item. É importante ressaltar que, devido à natureza das respostas abertas em cada seção do teste Fluência, não foi possível realizar análises pela TRI, pois não houve codificação das respostas em termos de acertos e erros. Em vez disso, foi registrada apenas a soma para os acertos em cada seção.

5.1 Teste Fluência

Participaram desse estudo 1352 sujeitos, a amostragem foi por conveniência, residentes nos estados de Minas Gerais (N=809), Distrito Federal (N=299), Espírito Santo

(N=150) e Ceará (N=94). Foram 887 mulheres (65,6%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=26,4; DP=11,45). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=703), seguido pelo Ensino Médio completo ou incompleto (N=392) e Ensino Fundamental completo ou incompleto (N=257).

A Tabela 6 mostra os dados descritivos para o teste de Fluência. Observa-se uma discrepância no tamanho da amostra entre os itens individuais do teste (Flu1 ao Flu4) e o resultado total de Fluência. Alguns protocolos de participantes com ensino fundamental foram digitados apenas com o escore bruto do sujeito, ou seja, os resultados parciais não foram registrados. Isso explica a maior quantidade de participantes para o resultado total em comparação com as seções individuais.

Tabela 6. Estatísticas descritivas do teste Fluência

Estatística Descritiva	Flu1	Flu2	Flu3	Flu4	Teste Fluência
N	1251	1251	1251	1251	1352
Média	9,10	7,06	6,70	18,6	40,4
Mediana	9	7	7	19	40,0
Moda	8,00	7,00	5,00	23,0	43,0
Desvio-padrão	3,69	3,10	3,47	5,30	13,0
Amplitude	22	17	19	30	80
Mínimo	0	0	0	0	0
Máximo	22	17	19	30	80
W de Shapiro-Wilk	0,988	0,979	0,981	0,980	0,998
p Shapiro-Wilk	< 0,001	< 0,001	< 0,001	< 0,001	0,173

Nota. Flu1: Fluência seção 1 do teste (o mesmo é válido para Flu2, 3 e 4).

Para avaliar a normalidade dos dados, foram considerados dois parâmetros: o teste Shapiro-Wilk e o diagrama quantil-quantil. O parâmetro estatístico indicou valor para W próximo de 1 e p não significativo para o valor total do teste Fluência ($W=0,998$; $p=0,173$). No diagrama Q-Q (Figura 9), os quantis estão distribuídos de forma linear sobre a reta esperada para uma distribuição normal. Com base nesses resultados, pode-se inferir uma distribuição normal dos dados para o teste.

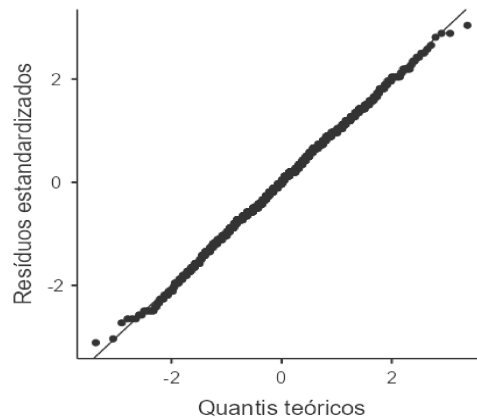


Figura 9. Diagrama quantil-quantil do teste Fluência

5.1.1 Evidências de validade baseadas na estrutura interna para o teste Fluência

Na AFC realizada para o teste de Fluência foram considerados os resultados parciais de cada seção para verificar se o modelo é unidimensional, conforme previamente estabelecido no instrumento original. Observa-se na Tabela 7 que as cargas fatoriais foram adequadas, variando de 0,64 a 0,77, e foram encontrados bons índices de ajuste ao modelo previamente estabelecido na AFC.

Tabela 7. Cargas fatoriais e índices de ajustes do teste Fluência da Bateria TSP de acordo com a Análise Fatorial Confirmatória

Item	Cargas Fatoriais
1	0,77
2	0,64
3	0,74
4	0,72
Ajustes	Fluência
RMSEA	0,020
GFI	1,000
TLI	0,998
CFI	0,999

Ao realizar análises de correlação entre as seções do teste de Fluência (Figura 10), foram identificadas correlações positivas e moderadas para todos os itens. Identifica-se, assim, uma relação importante entre os itens constituintes do teste.

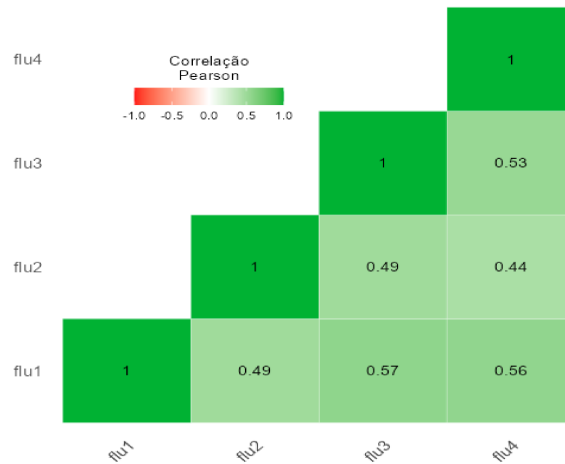


Figura 10. Mapa Térmico de Correlações

5.1.2 Evidências de validade baseadas nas relações com variáveis externas para o teste Fluência

a) Evidência de validade convergente

Foram avaliados 83 sujeitos por meio do teste de Fluência e da prova de Raciocínio Verbal da BPR-5. A amostra foi composta por 57 mulheres (68,7%), com idades variando entre 18 e 43 anos ($M=21,47$; $DP=4,79$), e ensino superior incompleto e completo. Os resultados analisados indicaram evidências de validade convergente para o teste de Fluência, com uma correlação moderada e estatisticamente significativa ($r=0,46$; $p<0,001$).

b) Evidências de validade de critério (variáveis sociodemográficas)

Com relação à variável sexo, a Tabela 8, apresenta que as diferenças entre as médias de desempenhos de homens e mulheres foram estatisticamente significativas no teste Fluência ($p=0,002$), no entanto, o tamanho do efeito foi inferior ao ponto de corte adotado ($d=0,17$). Desse modo, a grande sobreposição da distribuição dos desempenhos das mulheres e dos homens no teste apontou que a diferença estatisticamente significativa na variável sexo têm pouco impacto prático sobre o resultado. Esses achados corroboram com os estudos que fundamentaram o último manual da Bateria TSP no Brasil (Nascimento & Carmen-Flores, 2007). Além disso, estão alinhados com a pesquisa de Bilash e Zubek (1960), a qual não identificou diferença significativa entre os sexos no teste de Fluência.

Tabela 8. Desempenho dos participantes no teste Fluência de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	Teste <i>t</i>	Valor <i>p</i>	<i>d</i> de Cohen
Fluência	Feminino	887	41,22	12,39	3,05	0,002	0,17
	Masculino	465	38,95	14,01			

Os resultados das análises entre os grupos etários, conforme apresentados na Tabela 9, indicaram a presença de diferenças estatisticamente significativas. Considerando o tamanho do efeito, observou-se que as diferenças entre as médias foram moderadas ($\eta^2 = 0,05$). As comparações *post hoc* indicaram a existência de diferenças significativas ($p < 0,001$) entre os desempenhos médios apenas para o grupo etário de 15-16 anos com todos os demais grupos, com tamanho de efeito médio a forte ($d = -0,50$ a $-0,91$).

É possível verificar que nos grupos etários estudados a partir dos 17 anos de idade, os desempenhos médios no teste de Fluência foram consistentes e praticamente estáveis. Esses resultados assemelham-se ao desempenho de adultos, com idades entre 20 e 52 anos, que também demonstraram uma estabilidade de resultados na prova de Raciocínio Verbal da BPR-5 (Primi et al., 2022).

Tabela 9. Desempenho dos participantes no teste Fluência segundo a faixa etária

Teste	Faixa etária	N	Média	DP	F	Valor <i>p</i>	η^2
Fluência	15-16	88	31,79	11,41	19,59	<0,001	0,05
	17-18	273	39,38	11,42			
	19-25	524	43,36	12,23			
	26-39	250	40,50	12,62			
	≥40 anos	216	38,07	15,46			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

O raciocínio verbal, associado à inteligência cristalizada, tende a aumentar ao longo de todo o ciclo vital (Kaufman, 2001; Kaufman & Horn, 1996). Essa tendência pode ser observada na comparação dos resultados médios para os três grupos etários entre 15 e 25 anos no teste Fluência. No entanto, a partir da faixa etária de 26 anos ou mais, o desempenho médio foi gradativamente diminuindo. Uma possível explicação para esse dado é que o resultado pode ter sido influenciado pela presença predominante de indivíduos com ensino fundamental nos últimos grupos etários. Dos 257 sujeitos com ensino fundamental, aproximadamente 81% possuem 26 anos ou mais. Isso implica que há pouca presença de sujeitos com essa escolaridade nos grupos etários mais jovens e, esse contingente representa cerca de 44,6% dos sujeitos para os grupos etários com 26 anos ou mais.

A terceira variável investigada foi a escolaridade (Tabela 10), as diferenças nas médias de desempenho dos grupos foram estatisticamente significativas ($p < 0,001$) e com tamanho de efeito forte ($\eta^2 = 0,25$). As comparações *post hoc* apontaram que as médias dos desempenhos dos participantes com ensino fundamental diferiram daqueles com ensino médio e superior e, as comparações entre os grupos do ensino médio e superior também foram estatisticamente significativas, todos com tamanho de efeito de moderado à forte. Observou-se que as médias de desempenho foram inferiores para os participantes com ensino fundamental, enquanto as maiores médias foram alcançadas pelos que possuíam ensino superior completo ou incompleto.

Tabela 10. Desempenho dos participantes no teste Fluência segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor <i>p</i>	η^2
Fluência	Fundamental	257	28,56	11,24	230,27	<0,001	0,25
	Médio	392	38,53	10,79			
	Superior	703	45,84	11,49			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

Foi feita uma análise de correlação com o objetivo de aprofundar a compreensão sobre o impacto da idade e da escolaridade no desempenho de um teste de capacidade verbal. Observando a Tabela 11, nota-se que as relações entre escolaridade e idade, e entre idade e o teste de fluência, apresentaram correlações negativas de magnitude fraca, estatisticamente significativas. Esse achado é coerente com a alta proporção de participantes com ensino fundamental em idades mais avançadas, refletindo um contingente significativo de adultos nos quais o progresso educacional não foi compatível com a idade devido a interrupção no percurso escolar.

Portanto, neste estudo observou-se que o aumento esperado ao longo do ciclo vital na capacidade verbal tende a se manifestar apenas quando o percurso educacional está alinhado com a idade, evidenciando a significativa influência do sistema educacional formal, pelo menos no que diz respeito ao crescimento da habilidade verbal. Os resultados das diferenças médias em relação à escolaridade indicaram que os participantes com maior nível educacional apresentaram desempenhos médios superiores aos daqueles com menor nível escolar. Essa tendência foi corroborada pelos resultados das correlações, os quais revelaram uma associação positiva e moderada ($r = 0,50$; $p < 0,001$) entre o nível de escolaridade dos indivíduos e seus desempenhos no teste de Fluência.

Tabela 11. Matriz de Correlações entre as variáveis idade e escolaridade e o teste Fluência

Variável	Estatística	Fluência	Escolaridade
Escolaridade	R de <i>Pearson</i>	0,50***	—
	gl	1350	—
	valor de <i>p</i>	<0,001	—
Idade	R de <i>Pearson</i>	-0,09**	-0,16***
	gl	1349	1349
	valor de <i>p</i>	0,001	<0,001

Nota. ** $p < 0,01$, *** $p < 0,001$

5.1.3 Evidências de confiabilidade do teste Fluência

Os coeficientes de consistência interna alfa de Cronbach e ômega de McDonald apresentaram resultados de 0,79 e 0,81, respectivamente, indicando adequada consistência interna. Além disso, a evidência de estabilidade temporal demonstrou um coeficiente de correlação de *Pearson* de magnitude forte ($r=0,77$), estatisticamente significativo ($p<0,001$), o que indica uma estabilidade temporal satisfatória (Tabela 12).

Tabela 12. Estatísticas do estudo de estabilidade temporal do teste Fluência (n=78).

Teste	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	r_{12}	Valor <i>p</i>
Fluência	48,06	12,97	53,85	11,59	0,77	<0,001

Nota. M: Média, DP: Desvio- Padrão, r_{12} : correlação bivariada, valor significativo se $p < 0,05$.

Ao analisar os resultados gerais do estudo com o teste Fluência, observa-se que a pesquisa de validade interna por meio da AFC revelou um bom ajuste ao modelo unidimensional previamente estabelecido. Esse ajuste foi evidenciado pelos índices de ajustamento que atenderam aos valores recomendados pela literatura especializada (Hair et al., 2009; Morin; Arens; Marsh, 2016; Pilati; Laros, 2007).

Além disso, foi obtida evidência de validade convergente entre o teste Fluência e a prova de Raciocínio Verbal da BPR-5, um instrumento reconhecido nacionalmente que avalia o mesmo construto. Os resultados sugeriram que o teste Fluência é capaz de avaliar eficazmente o construto pretendido, fortalecendo sua validade como ferramenta de avaliação do vocabulário simples (Kalkbrenner, 2021). As análises dos resultados no teste em relação às variáveis sociodemográficas demonstraram a capacidade do instrumento para discriminar grupos distintos com base na idade, escolaridade e sexo, em concordância com o que é previsto na literatura especializada (Kaufman et al., 2009; Primi et al., 2022).

As análises de confiabilidade, por meio dos índices de consistência interna e estabilidade temporal, revelaram resultados empiricamente robustos, condizente com as atuais diretrizes normativas (CFP, 2022). Os achados sugerem que o teste Fluência apresenta adequação teórica e métrica, sendo apropriado para avaliar a fluência de vocabulário em indivíduos a partir de 15 anos, a partir de ensino fundamental incompleto.

5.2 Teste Números

Participaram desse estudo 1366 sujeitos, residentes nos estados de Minas Gerais (N=800), Distrito Federal (N=302), Espírito Santo (N=152) e Ceará (N=112). Foram 899 mulheres (65,8%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=26,4; DP=11,5). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=714), seguido pelo Ensino Médio completo ou incompleto (N=401) e Ensino Fundamental completo ou incompleto (N=251).

Para avaliar a normalidade dos dados (Tabela 13) utilizou-se o teste estatístico Shapiro-Wilk ($W=0,976$; $p<0,001$). O valor-p menor que o nível de significância indica que esta amostra difere de uma distribuição normal. No entanto, o valor de W próximo de 1, sugere uma distribuição aproximadamente normal. Além disso, observa-se uma pequena assimetria positiva (0,409) e uma curtose platicúrtica (-0,468), indicando uma distribuição levemente achatada. A moda foi a menor medida de tendência central, enquanto que a média foi a maior, sugerindo a presença de valores extremos na parte superior do conjunto de dados, o que pode estar influenciando para a assimetria positiva (Henderson, 2006). Dada a possibilidade de erro do tipo II, também foram analisados parâmetros de inspeção visual (Miot, 2017; Vetter, 2017).

Tabela 13. Estatística descritiva do teste Números

Estatística descritiva	Teste Números
N	1366
Média	20,5
Mediana	19,0
Moda	17,0
Desvio-padrão	8,79
Amplitude	44
Mínimo	1
Máximo	45
Assimetria	0,409

Erro-padrão da Assimetria	0,066
Curtose	-0,468
Erro-padrão da Curtose	0,132
W de Shapiro-Wilk	0,976
p Shapiro-Wilk	<0,001

A inspeção da Figura 11 aponta para a proximidade dos dados amostrais de uma distribuição normal, nos quais os quantis tendem a acompanhar a reta indicativa de uma simetria. Em função da discrepância entre os testes estatísticos e o diagrama Q-Q foi analisado, de maneira complementar, um histograma de frequência. A Figura 12 evidencia a distribuição dos dados com uma inclinação levemente para a direita (positiva).

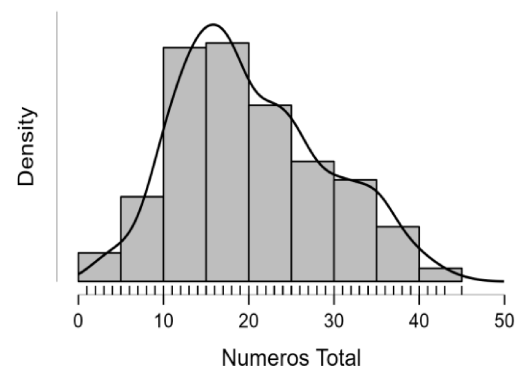
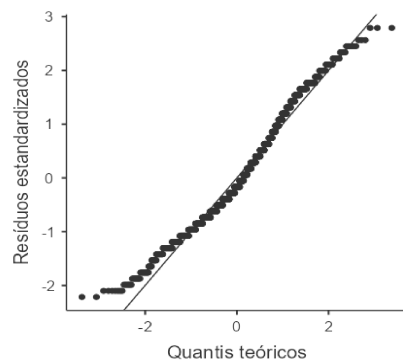


Figura 11. Diagrama Q-Q teste Números Figura 12. Curva de Gauss do teste Números

Os resultados de todos os parâmetros estudados, apontam uma pequena assimetria positiva, mas acredita-se que esta não seja suficiente para descartar a hipótese nula, de uma distribuição normal dos dados do teste Números. Desta maneira, as análises serão realizadas com testes estatísticos paramétricos em função do pressuposto de uma distribuição dos dados aproximadamente normal (Dancey & Reidy, 2019).

5.2.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Números

Inicialmente, o teste Números foi avaliado, considerando os 54 itens. Entretanto, os últimos quatro itens não apresentaram respostas válidas (50, 51, 52, 54), conforme também observado na AFC. Sendo assim, apenas 50 itens foram analisados, com percentual de casos omissos associado à ordem das questões, ou seja, quanto mais longínqua é a questão, maior é a quantidade de *missing values*. Por se tratar de um teste com tempo máximo estipulado, observa-se na Tabela 14, que a partir do item 37, o número de sujeitos respondentes é menor que 100 e, a partir do item 44, menos de 10 sujeitos por item.

O indicador de discriminação dos itens do teste Números variou entre 0,849 (moderada; Q36) e 6,561 (muito alta; Q46), com média de 2,116 (DP=1,504). Já o parâmetro de dificuldade variou entre -3,654 (Q7) e +0,166 (Q42), indicando que o teste apresenta itens fáceis e moderados, com média de dificuldade de -1,884 (DP=0,839).

Apenas seis itens apresentaram probabilidade de acerto ao acaso acima do ideal para itens dicotômicos ($c > 0,50$), são eles: Q1 ($c=0,698$), Q2 ($c=0,588$), Q10 ($c=0,635$), Q32 ($c=0,625$), Q36 ($c=0,564$) e Q43 ($c=0,667$). Apesar de os itens demonstrarem resultados satisfatórios em termos de discriminação e dificuldade, a observação de um aumento no parâmetro (c) sugere uma diminuição na qualidade de informação proporcionada por estes itens. Isso se deve ao fato de que um aumento em (c) indica uma maior probabilidade de acerto ao acaso, mesmo para sujeitos com um traço latente (θ) extremamente baixo.

Tabela 14. Parâmetros individuais dos itens do teste Números (N=1255)

Item	N	a	b	c
num1	1255	1,010	-2,002	0,698
num2	1253	1,325	-1,390	0,588
num3	1249	1,134	-3,189	0,002
num4	1239	0,853	-3,644	0,002
num5	1228	2,429	-2,046	0,003
num6	1211	1,282	-2,371	0,002
num7	1237	1,019	-3,654	0,003
num8	1235	1,287	-2,694	0,196
num9	1223	1,243	-2,313	0,248
num10	1230	1,440	-1,688	0,635
num11	1200	1,371	-2,297	0,337
num12	1086	2,514	-0,761	0,371
num13	1114	1,542	-2,091	0,001
num14	945	1,290	-0,748	0,295
num15	1070	0,951	-2,523	0,001
num16	958	2,116	-1,505	0,000
num17	988	1,248	-2,214	0,001
num18	842	2,771	-1,005	0,143
num19	851	1,151	-2,750	0,002
num20	707	4,415	-0,815	0,416
num21	661	1,131	-1,930	0,001
num22	677	1,163	-2,404	0,195
num23	623	2,681	-1,885	0,000
num24	548	1,808	-1,960	0,001
num25	539	1,080	-3,121	0,007

num26	442	1,281	-1,743	0,000
num27	434	1,219	-2,299	0,301
num28	388	1,659	-2,454	0,008
num29	369	1,636	-2,689	0,003
num30	291	1,613	-1,931	0,003
num31	260	1,902	-0,677	0,235
num32	233	2,460	-1,049	0,625
num33	197	0,962	-2,563	0,003
num34	189	2,088	-1,574	0,001
num35	162	0,870	-1,670	0,356
num36	154	0,849	-2,231	0,564
num37	91	1,584	-0,678	0,001
num38	84	2,824	-1,061	0,001
num39	59	1,015	-1,088	0,001
num40	47	2,047	-0,802	0,001
num41	30	0,850	-2,012	0,004
num42	27	4,332	0,166	0,300
num43	16	5,462	-1,211	0,667
num44	8	6,561	-1,385	0,001
num45	7	5,876	-1,087	0,001
num46	6	2,081	-2,433	0,003
num47	3	1,935	-2,413	0,003
num48	3	6,302	-3,115	0,003
num49	3	5,316	-0,755	0,333
num53	3	2,801	-2,458	0,003

As Curvas Características dos Itens (CCIs) foram analisadas (Figura 13), e ilustram que os itens Q20, Q42, Q43, Q44, Q45, Q48 e Q49, apresentaram parâmetros de discriminação irrealistas ($a > 3,0$; Barbetta et al., 2014). Por exemplo, as CCIs dos itens Q44 e Q48 revelam uma discriminação quase perfeita, sugerindo que esses itens conseguem distinguir claramente entre dois grupos de respondentes: aqueles que acertam o item e possuem um nível *theta* acima do nível de dificuldade do item, e aqueles que erram, com um *theta* abaixo do nível de dificuldade do item (Andrade, Laros & Gouveia, 2010). Destaca-se que, com exceção do item 20, os demais itens apresentam uma amostra muito reduzida (<30 sujeitos), o que sugere que os parâmetros individuais desses itens devem ser interpretados com cautela.

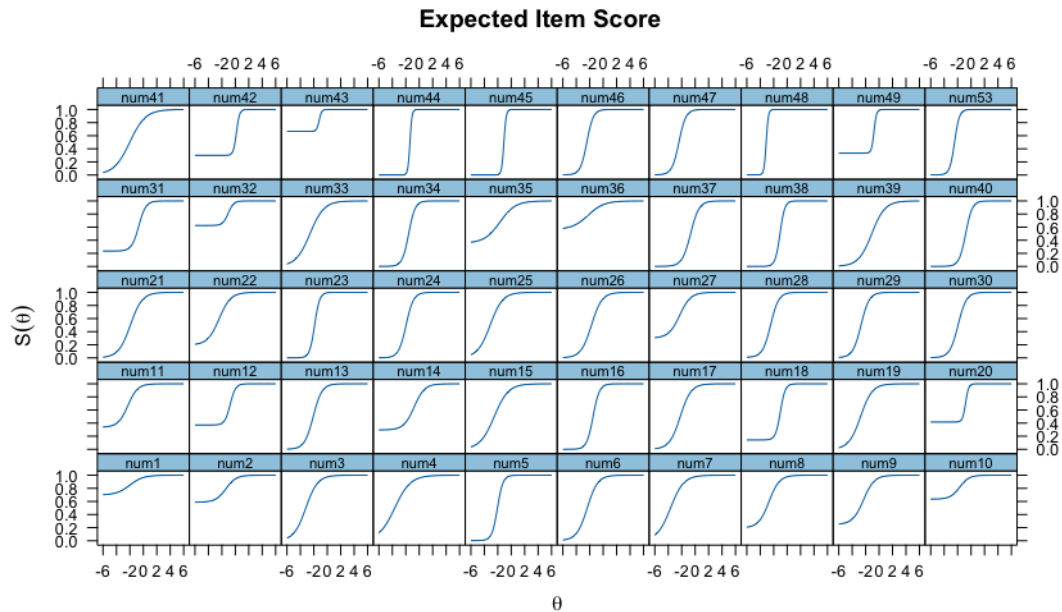


Figura 13. Curvas Características dos Itens (CCIs) do teste Números

A partir da observação da CIT (Figura 14) nota-se o teste Números teve maior poder de informação para valores de θ entre $-1,7$ e 0 , ou seja, nesta região a medida mostra-se mais precisa. Há uma melhoria na precisão do teste também para o valor de θ em torno de $-3,1$. Isso é evidenciado pela presença de dois picos distintos na curva, indicando duas tendências diferentes, ambas dentro do intervalo negativo, abaixo de zero. Pelo exame do erro padrão, para os níveis de θ inferiores a -4 e superiores a $+1$, o teste produz mais erro do que informação legítima. A CIT informa que o teste funciona melhor para sujeitos com habilidades baixas. Esses resultados são consistentes com as características dos itens considerados em sua maioria fáceis.

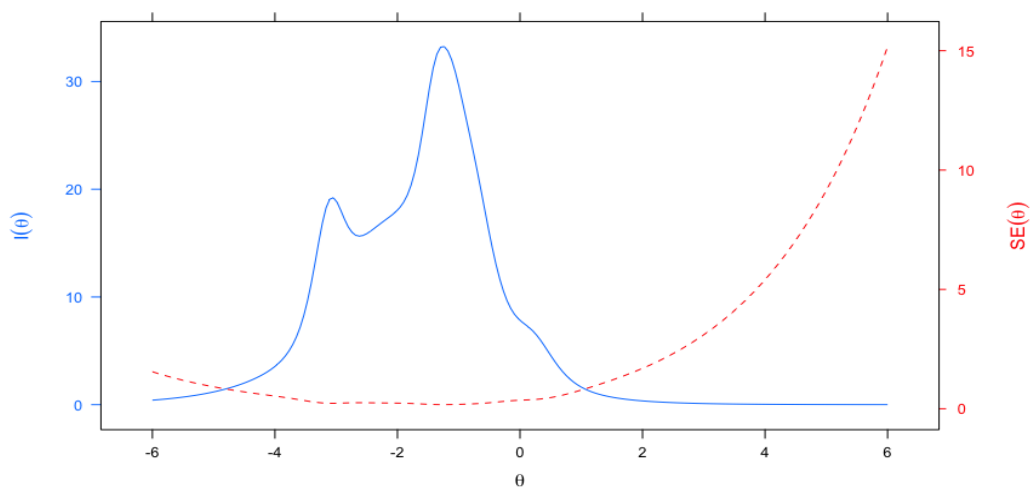


Figura 14. Curva de Informação do Teste (CIT) e erro-padrão

5.2.2 Evidências de validade baseadas na estrutura interna para o teste Números

Os resultados das cargas fatoriais obtidas pela AFC são apresentados na Tabela 15, na qual se observa que os itens com variabilidade de resposta baixa e/ou nula não foram incluídos nas análises (Q50, Q51, Q52, Q54). As cargas fatoriais até o item 44 foram em sua maioria de magnitudes fortes. No entanto, nota-se que para os itens subsequentes, as magnitudes das cargas foram gradativamente diminuindo, e a partir do item 47, exibiram valores muito baixas. Esse padrão possivelmente reflete o baixo número de respondentes e a falta de variabilidade nos dados.

Tabela 15. Cargas fatoriais e índices de ajustes do teste Números da Bateria TSP de acordo com a Análise Fatorial Confirmatória.

Item	Cargas Fatoriais
1	0,82
2	0,79
3	0,89
4	0,76
5	0,93
6	0,82
7	0,89
8	0,91
9	0,86
10	0,87
11	0,89
12	0,74
13	0,83
14	0,68
15	0,75
16	0,85
17	0,81
18	0,86
19	0,88
20	0,90
21	0,89
22	0,95
23	0,99
24	0,97
25	0,95
26	0,95
27	0,95
28	0,98
29	0,99
30	0,97
31	0,95

32	0,96
33	0,97
34	0,98
35	0,96
36	0,92
37	0,94
38	0,96
39	0,93
40	0,96
41	0,96
42	0,93
43	0,84
44	0,53
45	0,49
46	0,40
47	0,03
48	0,01
49	0,03
50	***
51	***
52	***
53	0,03
54	***
<hr/>	
Ajustes	Números
RMSEA	0,09
GFI	0,95
TLI	0,93
CFI	0,93

Nota. ***Itens retirados da análise por falta de variabilidade de resposta.

Os índices de ajuste preconizados na literatura foram inspecionados, revelando resultados satisfatórios da AFC para o modelo unidimensional investigado. Esses resultados evidenciam a validade da estrutura interna do teste Números (Hu & Bentler, 1998; Marsh, Hau, & Wen, 2004; Morin, Arens, & Marsh, 2016).

5.2.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Números

a) Evidências de validade convergente

Foram 83 sujeitos submetidos ao teste Números e a prova de Raciocínio Numérico da BPR-5. A maioria do sexo feminino (68,7%), a idade variou entre 18 e 43 anos (M=21,47; DP=4,79), com ensino superior incompleto e completo. Foi feita análise de correlação de

Pearson, os resultados apontaram evidências de validade convergente para o teste Números com correlação moderada e estatisticamente significativa ($r=0,43$; $p=0,001$).

b) Evidências de validade de critério (variáveis sociodemográficas)

Com relação à variável sexo (Tabela 16), as diferenças entre as médias de desempenhos de homens e mulheres mostraram-se estatisticamente significativas no teste Números ($p<0,001$). O menor resultado no desempenho das mulheres é comumente notado em avaliações que requerem habilidades matemáticas, como é o caso do teste Números (Flores-Mendoza, 2000; Primi et al., 2022). No entanto, o tamanho do efeito foi inferior ao ponto de corte adotado ($d=-0,23$). Ou seja, apesar da significância estatística, essa diferença é considerada desprezível em termos práticos, indicando que as discrepâncias na variável sexo têm pouco impacto nos resultados. Suspeita-se que o baixo tamanho de efeito observado possa estar relacionado ao baixo nível de dificuldade dos itens do teste Números, conforme demonstrado pelas análises da TRI. Como os itens não demandam uma habilidade alta para serem acertados, as diferenças entre grupos podem não ser claramente evidenciadas.

Tabela 16. Desempenho dos participantes no teste Números de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	Teste <i>t</i>	Valor <i>p</i>	<i>d</i> de Cohen
Números	Feminino	899	19,79	8,54	-3,998	<0,001	-0,23
	Masculino	467	21,78	9,10			

Conforme mostrado na Tabela 17, os resultados das análises para a variável idade indicaram a presença de diferenças estatisticamente significativas entre os grupos etários ($p<0,001$). O tamanho do efeito apontou que as diferenças entre as médias foram moderadas ($\eta^2=0,06$). As comparações *post hoc* revelaram diferenças significativas e tamanhos de efeito acima do ponto de corte apenas para o grupo etário de 19 a 25 anos, no qual o desempenho médio foi significativamente superior em comparação com os grupos extremos de 15-16 anos e ≥ 40 anos.

Observa-se entre os resultados um aumento médio progressivo nos grupos etários até a idade de um adulto jovem. Este padrão também foi observado com os respondentes da prova de Raciocínio Numérico da BPR-5, onde os melhores desempenhos registrados foram entre os indivíduos com idades entre 20 e 25 anos (Primi et al., 2022). Além disso, os resultados médios tenderam a diminuir no teste Números a partir dos 26 anos, um padrão também

observado no desempenho na prova RN, especialmente em sujeitos a partir da idade de 28 anos (Primi et al., 2022). É importante ressaltar que o grupo etário entre 19-25 anos apresenta o maior número de estudantes universitários, o que pode ter influenciado para o aumento considerável na pontuação do teste Números em comparação com os grupos etários extremos.

Tabela 17. Desempenho dos participantes no teste Números segundo a faixa etária

Teste	Faixa etária	N	Média	DP	F	Valor <i>p</i>	η^2
Números	15-16	88	18,45	8,14	20,10	<0,001	0,06
	17-18	275	19,29	7,74			
	19-25	533	23,04	9,03			
	26-39	248	19,01	8,20			
	≥40 anos	221	18,24	8,52			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

Ao considerar nível de ensino (Tabela 18), as diferenças nas médias de desempenho dos grupos no teste Números foram estatisticamente significativas ($p < 0,001$) e com tamanho de efeito forte ($\eta^2 = 0,16$). As comparações *post hoc* apontaram que as médias de desempenhos dos participantes entre todos os níveis de ensino foram estatisticamente significativas. No entanto, o tamanho de efeito foi acima do ponto de corte apenas entre o grupo com ensino fundamental em relação aos demais. Ou seja, as comparações entre os grupos do ensino médio e superior demonstraram uma grande sobreposição, não sendo observados efeitos práticos de diferenças significativas entre eles ($d = -0,39$).

Tabela 18. Desempenho dos participantes no teste Números segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor <i>p</i>	η^2
Números	Fundamental	251	13,62	6,46	131,30	<0,001	0,16
	Médio	401	19,97	7,92			
	Superior	714	23,16	8,61			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

5.2.4 Evidências de confiabilidade do teste Números

Os resultados para as evidências de confiabilidade obtidas pelos coeficientes de consistência interna alfa de Cronbach e ômega de McDonald foram α 0,95 e Ω 0,97, respectivamente. Foi também utilizada a análise *r* de Pearson para evidência de estabilidade temporal (Tabela 19). Observa-se o coeficiente de correlação de magnitude forte ($r = 0,79$) e

estatisticamente significativo ($p < 0,001$). Portanto, o teste Números apresenta adequada consistência interna e estabilidade temporal satisfatória.

Tabela 19. Estatísticas do estudo de estabilidade temporal do teste Números (N=81)

Teste	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	<i>r Pearson</i>	Valor <i>p</i>
Números	29,90	7,58	30,94	7,74	0,79	<0,001

Nota. M: Média, DP: Desvio-Padrão, valor significativo se $p < 0,05$.

Ao analisar os resultados gerais, é importante destacar uma diminuição significativa no número de respondentes a partir do item 37 do teste Números ($n < 100$). Diante dessa limitação amostral, os resultados tanto pela AFC quanto pela TRI obtidos para os itens finais devem ser interpretados com cautela, independentemente de apresentarem resultados psicométricos satisfatórios ou não. Um aspecto que requer atenção são os cinco itens que demonstraram índices de chute elevados, apesar de terem uma boa quantidade de respondentes (Q1, Q2, Q10, Q32, Q36). Embora os outros parâmetros estejam dentro dos limites aceitáveis, um alto valor do parâmetro c sugere que esses itens podem não ser eficazes na diferenciação entre participantes com diferentes níveis de habilidade, já que possui uma alta probabilidade de acerto ao acaso. Isso indica a necessidade de uma revisão desses itens em futuras pesquisas, visando garantir sua melhor qualidade. De maneira geral, os resultados da TRI apontaram que os itens são predominantemente fáceis e o teste tem um desempenho mais preciso para sujeitos com habilidades baixas (Peres, Lessa, Valentini, 2019).

A pesquisa sobre a validade interna por meio da AFC revelou um bom ajuste ao modelo unidimensional do teste Números (Hair et al., 2009; Morin; Arens; Marsh, 2016; Pilati; Laros, 2007). Além disso, foi obtida evidência de validade convergente entre o teste Números e a prova de Raciocínio Numérico da BPR-5, com resultado satisfatório sobre sua capacidade de avaliar eficazmente o construto pretendido (Kalkbrenner, 2021). A análise da validade de critério apresentou resultados condizentes com a literatura sobre capacidade para discriminar grupos distintos com base na idade e escolaridade (Kaufman et al., 2009; Primi et al., 2022).

As investigações de confiabilidade, através dos índices de consistência interna e estabilidade temporal, indicaram resultados robustos do ponto de vista empírico, em conformidade com as diretrizes normativas atuais (CFP, 2022). Os resultados sugerem que o teste Números possui fundamentação teórica e métrica sólidas, sendo adequado para avaliar o raciocínio quantitativo em indivíduos com 15 anos ou mais, incluindo aqueles com ensino

fundamental incompleto.

5.3 Teste Julgamento

Participaram desse estudo 1401 sujeitos, a amostragem foi por conveniência, residentes nos estados de Minas Gerais (N=822), Distrito Federal (N=298), Espírito Santo (N=151) e Ceará (N=130). Foram 934 mulheres (66,7%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=26,4; DP=11,3). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=756), seguido pelo Ensino Médio completo ou incompleto (N=396) e Ensino Fundamental completo ou incompleto (N=249).

Foram conduzidas análises descritivas do teste Julgamento para avaliação da distribuição dos dados coletados (Tabela 20). Embora o resultado do teste de Shapiro-Wilk demonstre que a amostra é diferente de uma distribuição normal ($W=0,982$; $p=0,001$), ao considerar outros parâmetros é possível contestar este resultado. As medidas de tendência central estão muito próximas, o resultado da assimetria (-0,414) indica uma pequena assimetria negativa e a curtose (0,026) indica que os dados se aproximam de uma distribuição normal (Henderson,2006; Miot, 2017; Vetter, 2017).

Tabela 20. Estatística descritiva do teste Julgamento

Estatística descritiva	Teste Julgamento
N	1401
Média	18,76
Mediana	19
Moda	24
Desvio-padrão	6,99
Amplitude	39
Mínimo	0
Máximo	39
Assimetria	-0,414
Erro-padrão da Assimetria	0,065
Curtose	0,026
Erro-padrão da Curtose	0,131
W de Shapiro-Wilk	0,982
p Shapiro-Wilk	< 0,001

A Figura 15 demonstra que os pontos referentes aos dados amostrais formam uma linha diagonal sobrepondo a maior parte dos quantis indicando uma distribuição

aproximadamente normal (Miot, 2017). De maneira complementar a inspeção visual do histograma (Figura 16) confirma os achados de uma distribuição aproximadamente normal, com uma leve inclinação da cauda para a esquerda (Dancey & Reidy, 2019).

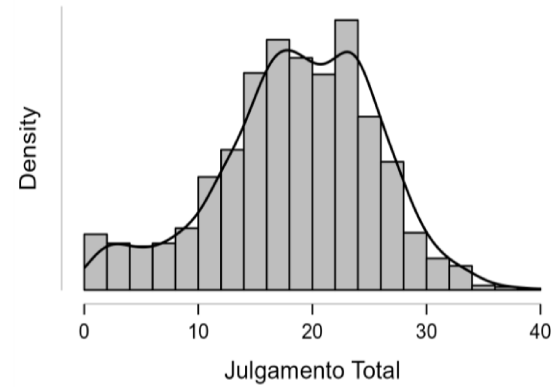
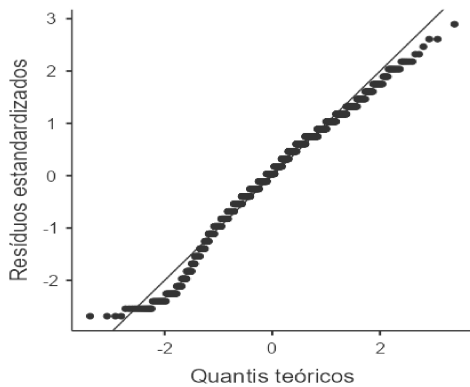


Figura 15. Diagrama Q-Q teste Julgamento Figura 16. Curva de Gauss do teste Julgamento

5.3.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Julgamento

Conforme pode ser observado na Tabela 21, a discriminação variou entre 0,111 (Q38) a 14,269 (Q43), com média de 3,447 (DP=3,061). Apenas três itens apresentaram resultados de discriminação baixos (Q6, Q38 e Q40). Foi também possível observar que 12 itens apresentaram discriminação irrealista, sendo eles: Q32, Q41, Q43, Q44, Q46, Q47, Q48, Q49, Q50, Q52, Q53 e Q54 ($a > 3,0$; Barbetta et al., 2014).

Os índices de dificuldade variaram entre -3,137 (Q18) a +6,598 (Q40), com média de -0,729 (DP=1,615), indicando que a maior parte dos itens apresentou parâmetros de dificuldade com tendências a moderados e fáceis. Por fim, o parâmetro acerto ao acaso variou entre 0,000 (Q3) a 0,600 (Q45), com média de probabilidade de chute de 0,101 (DP=0,133). É importante ressaltar que apenas o item Q45 apresentou probabilidade de chute acima do ideal ($c=0,600$). Embora este item tenha apresentado resultados satisfatórios para o grau de discriminação e dificuldade, o aumento percebido do parâmetro (c) diminuiu a sua qualidade de informação, pois há uma alta probabilidade de acerto casual, ou seja, maior possibilidade do sujeito mesmo com traço latente (θ) extremamente baixo acertar o item.

Tabela 21. Parâmetros individuais dos itens do teste Julgamento (N=1290)

Item	N	a	b	c
ju1	1290	1,453	-2,958	0,005
ju2	1280	2,292	-2,842	0,001

ju3	1280	2,509	-1,817	0,000
ju4	1286	2,591	-1,989	0,000
ju5	1224	1,844	-2,242	0,000
ju6	1184	0,403	-2,174	0,002
ju7	1274	2,091	-2,203	0,001
ju8	1181	1,140	-2,047	0,003
ju9	1212	2,528	-1,098	0,350
ju10	1254	2,156	-2,224	0,179
ju11	1200	1,795	-1,325	0,104
ju12	1233	1,383	-2,495	0,001
ju13	1224	2,243	-2,043	0,001
ju14	1057	0,972	-1,945	0,001
ju15	1141	1,544	0,209	0,082
ju16	1148	2,104	-2,068	0,000
ju17	1005	2,291	-1,854	0,001
ju18	1023	0,774	-3,137	0,001
ju19	866	1,284	-0,937	0,001
ju20	871	2,033	-1,242	0,001
ju21	769	3,293	-0,645	0,090
ju22	810	2,172	-1,805	0,218
ju23	705	3,277	-1,002	0,140
ju24	676	3,807	-0,934	0,255
ju25	566	1,730	-1,102	0,160
ju26	446	0,994	-1,457	0,001
ju27	374	2,485	-0,847	0,000
ju28	289	1,562	-1,442	0,001
ju29	201	1,967	1,109	0,154
ju30	179	3,071	-0,594	0,122
ju31	143	1,706	-1,544	0,002
ju32	74	4,124	-0,121	0,250
ju33	89	3,331	0,376	0,176
ju34	81	2,512	-0,865	0,008
ju35	47	2,734	0,616	0,126
ju36	45	2,830	-0,215	0,000
ju37	40	3,761	-0,848	0,182
ju38	22	0,111	3,178	0,036
ju39	22	2,369	-0,596	0,199
ju40	18	0,272	6,598	0,003
ju41	16	9,215	-0,498	0,000
ju42	12	1,609	-0,521	0,001
ju43	14	14,269	-0,718	0,399
ju44	11	12,644	0,147	0,311
ju45	10	3,281	1,456	0,600

ju46	9	7,491	0,640	0,111
ju47	8	6,776	0,881	0,125
ju48	7	6,875	-0,214	0,429
ju49	8	7,264	0,482	0,125
ju50	8	6,338	0,937	0,125
ju51	6	1,693	-1,997	0,004
ju52	7	11,094	-0,028	0,000
ju53	6	6,790	0,880	0,167
ju54	5	7,252	-0,255	0,200

Os indicadores de discriminação podem ser interpretados através da inclinação da reta observadas nas Curvas Características dos Itens (CCIs). Sendo assim, foi possível observar pela Figura 17 que os itens Q6, Q38 e Q40 apresentaram discriminação muito baixa, com padrão de CCI muito diferente das demais. Por outro lado, alguns itens apresentaram discriminação irrealista, próxima da perfeição, e são ilustrados a partir de angulações próximas a 90° (Andrade, Laros & Gouveia, 2010). Ressalta-se que a maioria dos itens com esse padrão apresentam uma amostra muito pequena ($n < 16$), dessa maneira, esses resultados devem ser avaliados com ressalvas, inclusive àqueles que apresentaram índices adequados. Por fim, as curvas ilustram parâmetro (c) através da altura do início da curva, no qual apenas o item Q45 apresentou resultados acima do ideal para itens dicotômicos ($c > 0,5$).

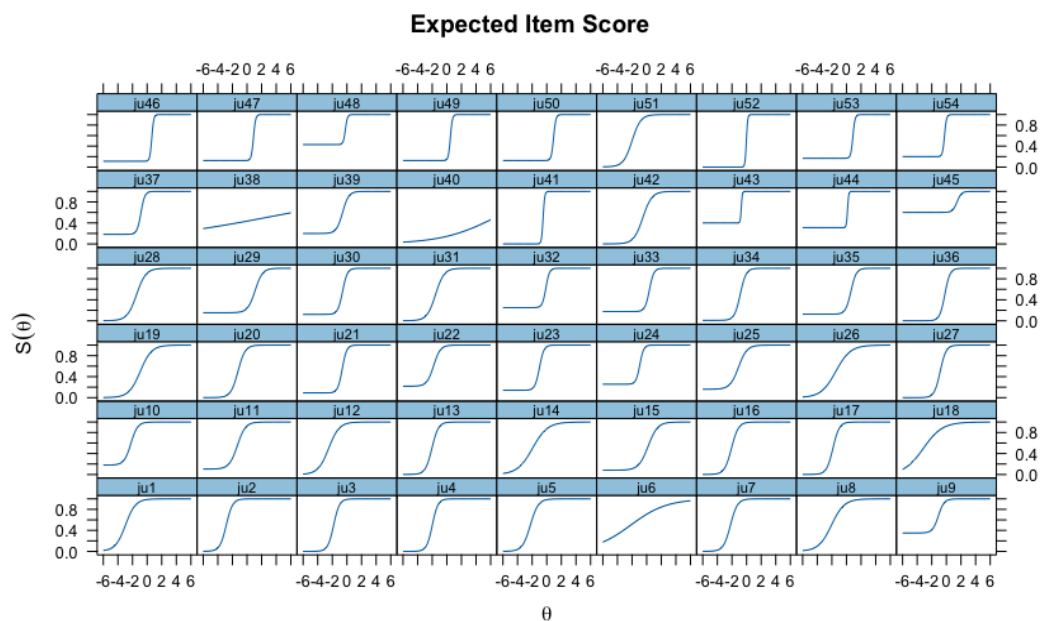


Figura 17. Curvas Características dos Itens (CCI's) do teste Julgamento

É importante destacar que o tamanho amostral varia ao longo dos itens devido ao tempo predeterminado para responder ao teste e à possibilidade de o sujeito pular algum item.

Dessa forma, os itens com um tamanho amostral inferior a 100 devem ser interpretados com cautela, independentemente da adequação ou não dos resultados dos parâmetros, ou seja, a partir do item 32. No entanto, o item 6 foi o único que, mesmo com um número considerável de respondentes, apresentou resultado ruim para o parâmetro de discriminação, sugerindo que o item tem pouco poder para distinguir entre sujeitos com níveis de θ semelhantes.

A observação da CIT (Figura 18) sinaliza que o teste Julgamento teve maior poder de informação para, principalmente, dois valores de θ : 0 e -0,8, aproximadamente. Nestas regiões a medida é mais precisa demonstrando um menor erro. Pelo exame do erro padrão, o teste tende a maior precisão para pessoas que variam o θ entre -3 e +1,7, aproximadamente, indicando que o teste funciona melhor para sujeitos que têm um nível de θ médio. Resultado condizente com as características dos itens considerados em sua maioria médios.

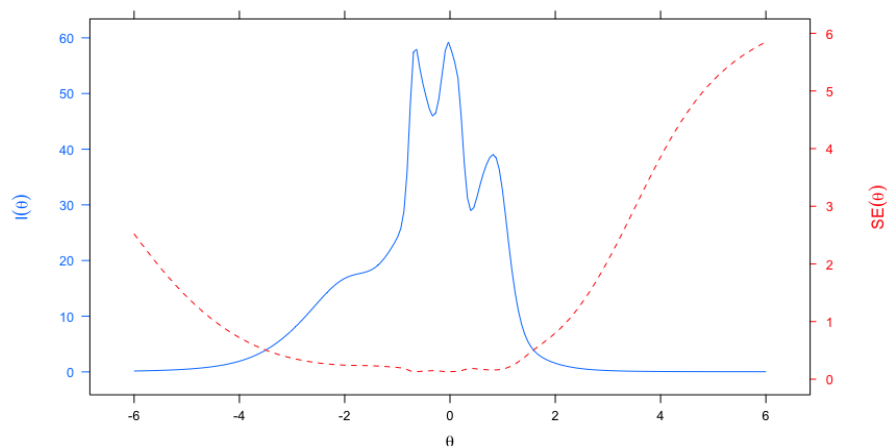


Figura 18. Curva de Informação do Teste (CIT) e erro-padrão

5.3.2 Evidências de validade baseadas na estrutura interna para o teste Julgamento

A investigação da estrutura interna foi conduzida por uma AFC, no qual foi testado o modelo unidimensional do teste Julgamento (Tabela 22). Os índices de ajuste preconizados na literatura foram inspecionados e os resultados da AFC foram considerados satisfatórios para o modelo unidimensional investigado, o que evidencia a validade da estrutura interna (Hu & Bentler, 1998; Marsh, Hau, & Wen, 2004; Morin, Arens, & Marsh, 2016).

Ao analisar os resultados das cargas fatoriais para os itens, nota-se que aqueles com variabilidade de resposta baixa e/ou nula não foram incluídos nas análises (item 34 ao 54). Apenas os itens 6 e 29 demonstraram cargas fatoriais fracas, as demais cargas fatoriais foram

em sua maioria de magnitude forte (Dancey & Reidy, 2019). O resultado para a maior parte dos itens acima de 0,40 reforçou que estes representam adequadamente o construto.

Tabela 22. Cargas fatoriais e índices de ajustes do teste Julgamento da Bateria TSP de acordo com a Análise Fatorial Confirmatória

Item	Carga Fatorial
1	0,60
2	0,83
3	0,80
4	0,83
5	0,72
6	0,22
7	0,77
8	0,53
9	0,70
10	0,73
11	0,70
12	0,60
13	0,80
14	0,53
15	0,54
16	0,76
17	0,73
18	0,45
19	0,57
20	0,75
21	0,81
22	0,72
23	0,84
24	0,84
25	0,66
26	0,56
27	0,87
28	0,75
29	0,31
30	0,90
31	0,82
32	0,71
33	0,64
34 a 54	***
Ajustes	Julgamento
RMSEA	0,01
GFI	0,98
TLI	0,98
CFI	0,98

Nota. ***Itens retirados da análise por falta de variabilidade de resposta.

5.3.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Julgamento

a) Evidência de validade convergente

Foram submetidos 27 sujeitos ao teste Julgamento e ao teste G-38. Participaram 21 mulheres (77,78%), a idade variou entre 19 e 62 anos ($M=27,70$; $DP=11,15$), com ensino superior incompleto e completo. Os resultados analisados apontaram evidências de validade convergente para o teste Julgamento com correlação moderada e estatisticamente significativa ($r=0,58$; $p=0,002$).

No teste Julgamento, os itens apresentam uma alternância de estímulos fixa que o sujeito irá responder, a saber: avaliar o número seguinte de uma sequência estabelecida, averiguar o grupo que diverge de outros que compartilham características comuns e identificar a letra seguinte de uma sequência alfabética. Em função disso, o objetivo do estudo foi obter evidência de validade para o teste Julgamento, por meio da correlação com o teste de Atenção Alternada da BPA. Participaram do estudo 17 sujeitos, a maioria mulheres ($n=15$), a idade variou entre 19 e 48 anos ($M=29,65$; $DP=10,71$), com ensino superior incompleto e completo. Os resultados analisados também evidenciam validade convergente para o teste Julgamento com correlação moderada e estatisticamente significativa para o teste de atenção alternada ($r=0,53$; $p=0,03$).

b) Evidências de validade de critério (variáveis sociodemográficas)

Conforme pode ser observado na Tabela 23, as diferenças entre as médias de desempenhos de homens e mulheres não foram estatisticamente significativas no teste Julgamento ($p=0,912$). O tamanho do efeito praticamente nulo ($d=0,006$) confirma que não houve uma diferença no desempenho no teste em função do sexo dos respondentes.

O teste Julgamento aborda a capacidade de resolução de problemas lógicos e é considerado mais relacionado à inteligência fluída. Estudos ao longo do tempo têm consistentemente mostrado que não há diferenças significativas nos desempenhos em função do sexo dos respondentes nos testes de raciocínio abstrato (Burgaleta et al, 2012; Flores-Mendoza, 2000; Kaufman & Horn, 1996). Dessa forma, este resultado constitui uma evidência de validade para o teste Julgamento, uma vez que os resultados estão em conformidade com a literatura.

Tabela 23. Desempenho dos participantes no teste Julgamento de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	Valor <i>p</i>	<i>d</i> de Cohen
Julgamento	Feminino	934	18,78	6,76	0,912	0,006
	Masculino	467	18,77	7,43		

Os resultados das análises para a variável idade (Tabela 24), indicaram a presença de diferenças estatisticamente significativas entre os grupos etários ($p < 0,001$). Considerando o tamanho do efeito, as diferenças entre as médias foram fortes ($\eta^2 = 0,19$). De modo geral, as comparações *post hoc* indicaram haver diferenças significativas entre os desempenhos médios do grupo etário ≥ 40 anos com todos os demais, com tamanho de efeito variando entre $d = 0,54$ e $d = 1,302$. Alguns resultados não estatisticamente significativo entre os grupos etários 15 e 16 anos com o de 26 a 39 anos e, com os grupos com idade entre 17 e 25 anos.

Entende-se que há um aumento médio gradual entre os primeiros três grupos etários, porém, a partir do grupo de 26 a 39 anos, nota-se um declínio no desempenho médio no teste Julgamento. Esse padrão de resultados é semelhante ao encontrado em uma pesquisa anterior conduzida por Fernandes, Rueda e Sisto (2009), que também observou um aumento nos escores do teste até os 23 anos de idade, seguido por uma diminuição no desempenho dos participantes.

Tabela 24. Desempenho dos participantes no teste Julgamento segundo a faixa etária.

Teste	Faixa etária	N	Média	DP	F	Valor <i>p</i>	η^2
Julgamento	15-16	83	17,20	7,33	82,87	<0,001	0,19
	17-18	272	20,46	6,03			
	19-25	565	21,39	5,89			
	26-39	251	16,62	6,18			
	≥ 40 anos	229	13,19	7,21			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

Conforme a Tabela 25, as diferenças nas médias de desempenho dos grupos em função do ensino foram estatisticamente significativas e com tamanho de efeito forte ($\eta^2 = 0,31$). As comparações *post hoc* apontaram que as médias dos desempenhos dos sujeitos com ensino fundamental diferiram significativamente daqueles com ensino médio e superior. As comparações entre os grupos do ensino médio e superior também foram estatisticamente significativas ($p < 0,001$), no entanto o tamanho do efeito foi pequeno $d = -0,26$, ou seja, existe um risco do Erro Tipo I, pois há uma sobreposição substancial entre os grupos estudados.

Tabela 25. Desempenho da amostra no teste Julgamento segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor p	η^2
Julgamento	Fundamental	249	10,54	5,91	312,01	<0,001	0,31
	Médio	396	19,53	6,30			
	Superior	756	21,07	5,51			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor p = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

5.3.4 Evidências de confiabilidade do teste Julgamento

Evidências de confiabilidade foram obtidas pelos coeficientes de consistência interna alfa de Cronbach (α 0,93) e ômega de McDonald (Ω 0,94). Concluiu-se, portanto, que o teste apresenta adequada consistência interna. A evidência de estabilidade temporal pelo coeficiente de correlação de *Pearson* (Tabela 26) apresentou magnitude forte ($r=0,73$) e estatisticamente significativo, demonstrando também uma estabilidade temporal satisfatória para o teste Julgamento.

Tabela 26. Estatísticas do estudo de estabilidade temporal do teste Julgamento (N=87)

Testes	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	r_{12}	Valor p
Julgamento	24,31	5,06	26,78	5,01	0,73	<0,001

Nota. M: Média, DP: Desvio-Padrão, r_{12} : correlação bivariada, valor significativo se $p < 0,05$.

Em suma, a pesquisa sobre a validade interna por meio da AFC revelou um bom ajuste ao modelo unidimensional do teste Julgamento (Hair et al., 2009; Morin; Arens; Marsh, 2016). No entanto, foi observado que o item 6 necessita de revisão, podendo ser reformulado ou removido, devido aos seus resultados insatisfatórios nos métodos de AFC e dos parâmetros individuais. No geral, os resultados da TRI sugerem que os itens do teste Julgamento são predominantemente de dificuldade moderada, indicando que o teste pode ser mais adequado para indivíduos com habilidades cognitivas médias (Peres, Lessa, Valentini, 2019).

Além disso, foi obtida evidência de validade convergente entre o teste Julgamento e o teste G-38 com resultado satisfatório sobre sua capacidade de avaliar eficazmente o construto pretendido (Kalkbrenner, 2021). A análise da validade de critério pelo estudo das diferenças das médias entre grupos e as variáveis sexo, faixa etária e escolaridade constituem uma evidência de validade do teste, pois os resultados estão consoantes com a literatura (Burgaleta et al., 2012; Kaufman et al., 2009; Primi et al., 2022).

As análises de confiabilidade, utilizando índices de consistência interna e estabilidade

temporal, revelaram resultados satisfatórios, em linha com as diretrizes normativas atuais (CFP, 2022). Isso sugere que o teste Julgamento possui uma base teórica e métrica sólida, sendo apropriado para avaliar o raciocínio quantitativo em indivíduos com 15 anos ou mais e com escolaridade a partir do ensino fundamental incompleto.

5.4 Teste Percepção

Participaram desse estudo 1374 sujeitos, a amostragem foi por conveniência, residentes nos estados de Minas Gerais (N=792), Distrito Federal (N=299), Espírito Santo (N=152) e Ceará (N=131). Foram 906 mulheres (65,9%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=26,6; DP=11,47). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=725), seguido pelo Ensino Médio completo ou incompleto (N=398) e Ensino Fundamental completo ou incompleto (N=251).

Para avaliação da normalidade dos dados observa-se (Tabela 27) que o teste estatístico Shapiro-Wilk ($W=0,993$; $p<0,001$) demonstra pelo parâmetro de significância que esta amostra é diferente de uma distribuição normal, no entanto, nota-se um valor W muito próximo a 1, o qual pode indicar uma distribuição aproximadamente normal. Vale, então, observar os resultados das medidas de tendência central que estão muito próximas e a assimetria (-0,131) e curtose (-0,360) com valores baixos, o que é característico de uma distribuição normal. Em função da possibilidade do erro tipo II, dois parâmetros de inspeção visual também foram analisados (Henderson,2006; Miot, 2017; Vetter, 2017).

Tabela 27. Estatísticas descritivas do teste Percepção

Estatística descritiva	Teste Percepção
N	1374
Média	17,86
Mediana	18,0
Moda	18,0/16,0*
Desvio-padrão	6,84
Amplitude	33
Mínimo	0
Máximo	37
Assimetria	-0,131
Erro-padrão da Assimetria	0,066
Curtose	-0,360
Erro-padrão da Curtose	0,132
W de Shapiro-Wilk	0,993
p Shapiro-Wilk	<0,001

*Distribuição bimodal

Conforme observado no diagrama Q-Q (Figura 19), a maior parte da distribuição dos quantis da amostra de dados está distribuída de maneira linear sobre a reta esperada para a distribuição normal (Miot, 2017). De maneira complementar foi gerado um histograma de frequência (Figura 20), a inspeção visual dos dados confirma os achados de uma distribuição aproximadamente normal (Dancey & Reidy, 2019).

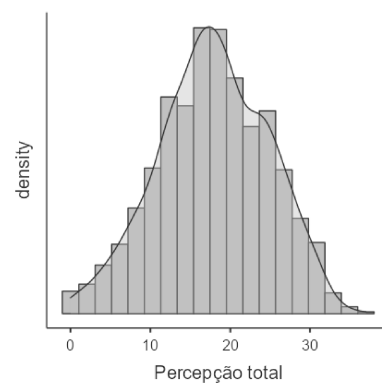
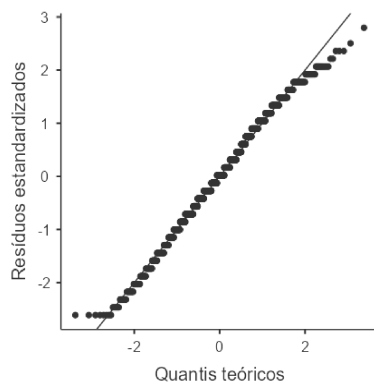


Figura 19. Diagrama Q-Q teste Percepção

Figura 20. Curva de Gauss do teste Percepção

5.4.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Percepção

Os resultados dos parâmetros individuais dos itens do teste Percepção estão apresentados na Tabela 28. Nota-se que os itens Q24B, Q25, Q26 e Q27 (formas A e B) foram retirados da análise por não apresentarem casos válidos, em função de nenhum participante ter respondido a esses itens dentro do limite de tempo estabelecido. Além disso, os itens Q19A, Q21B e Q23B também foram retirados, devido a apresentarem apenas uma categoria de resposta (0), ou seja, todos os participantes que responderam a esses itens, assinalaram uma mesma categoria de resposta, não apresentando variabilidade de respostas. Destaca-se que os itens com um tamanho amostral inferior a 100 devem ser interpretados com cautela, a partir do item 16B, independentemente da adequação dos resultados dos parâmetros.

Após a limpeza do banco de dados, o cálculo dos parâmetros individuais dos itens foi realizado. Os resultados apontam que a discriminação variou entre 0,083 (muito baixa; Q20B) e 10,774 (praticamente perfeita; Q19B), com média de 2,452 (DP= 2,87). É importante ressaltar que os itens Q20B e Q16B apresentaram discriminações insuficientes ($\alpha < 0,65$), enquanto os itens Q15B, Q19B, Q22A, Q22B, Q23A e Q24A, apresentaram parâmetros de

discriminação quase perfeitos ($a > 3,0$; Barbetta et al., 2014). Os parâmetros de dificuldade variaram entre $-4,029$ (Q16B) e $+7,679$ (Q20B), com média de $-1,138$ ($DP = 2,058$). É possível perceber que a maior parte dos itens apresentaram índices de dificuldade entre fáceis e médios.

A análise do parâmetro de chute (c) indicou que apenas três itens apresentaram probabilidade de acerto ao acaso acima do ideal para itens dicotômicos ($c > 0,50$), são eles: Q12B, Q13B, e Q15B. Isso sugere que uma pessoa mesmo com uma habilidade muito baixa possui cerca de 59% (13B) a 84% (15B) de chances de responder corretamente aos itens devido ao acaso. Embora os itens 12B e 13B tenham apresentado resultados satisfatórios para o grau de discriminação e dificuldade, o aumento percebido do parâmetro (c) diminui a sua qualidade de informação, pois há uma alta probabilidade de acerto casual.

Tabela 28. Parâmetros individuais dos itens do teste Percepção (N=1270)

Item	N	a	b	c
per1A	1268	1,251	-2,343	0,001
per1B	1270	1,252	-2,459	0,003
per2A	1262	1,448	-2,640	0,000
per2B	1262	0,891	-2,569	0,000
per3A	1257	0,704	-3,167	0,002
per3B	1252	1,368	-2,483	0,002
per4A	1207	1,988	-0,705	0,315
per4B	1231	1,607	-2,624	0,000
per5A	1219	1,604	-2,570	0,001
per5B	1187	1,874	-1,688	0,166
per6A	1175	2,453	-2,322	0,000
per6B	1169	1,768	-2,288	0,002
per7A	1119	1,866	-2,735	0,000
per7B	1095	1,256	0,952	0,205
per8A	1024	2,480	-0,824	0,443
per8B	1001	1,644	-2,361	0,001
per9A	926	2,059	-1,667	0,001
per9B	888	0,904	-2,611	0,001
per10A	781	1,190	-0,587	0,201
per10B	732	1,250	-2,157	0,000
per11A	639	1,522	-2,860	0,171
per11B	600	0,766	-3,396	0,093
per12A	510	1,234	-1,736	0,212
per12B	471	2,282	-1,328	0,633
per13A	384	1,004	-3,859	0,007

per13B	353	0,845	0,882	0,594
per14A	283	1,440	-1,494	0,001
per14B	249	0,921	-0,894	0,002
per15A	178	0,705	0,002	0,015
per15B	158	8,021	-0,040	0,845
per16A	104	1,152	-1,329	0,001
per16B	81	0,647	-4,029	0,008
per17A	55	1,240	-1,077	0,002
per17B	37	0,977	3,200	0,000
per18A	26	1,854	-0,426	0,001
per18B	23	1,162	2,107	0,001
per19B	12	10,774	-1,267	0,001
per20A	11	0,954	-0,973	0,002
per20B	9	0,083	7,679	0,160
per21A	9	1,589	-0,536	0,002
per22A	7	9,822	0,719	0,000
per22B	5	9,523	0,654	0,000
per23A	3	9,067	-0,866	0,002
per24A	3	9,468	0,650	0,001

Os indicadores de discriminação podem ser interpretados através da inclinação da reta observadas nas Curvas Características dos Itens (CCIs). É possível observar (Figura 21) que os itens com curvas mais próximas ao ângulo reto (90°) apresentam os maiores índices de discriminação, enquanto os itens com início das curvas mais altas indicam maior parâmetro c (chute), associado à maiores probabilidades de acerto ao acaso (Andrade, Laros & Gouveia, 2010). É possível perceber que os itens Q13B, Q15B, 16B e Q20B apresentaram as curvas mais discrepantes quando considerados os pontos de corte da TRI, indicando serem itens com ressalvas empíricas.

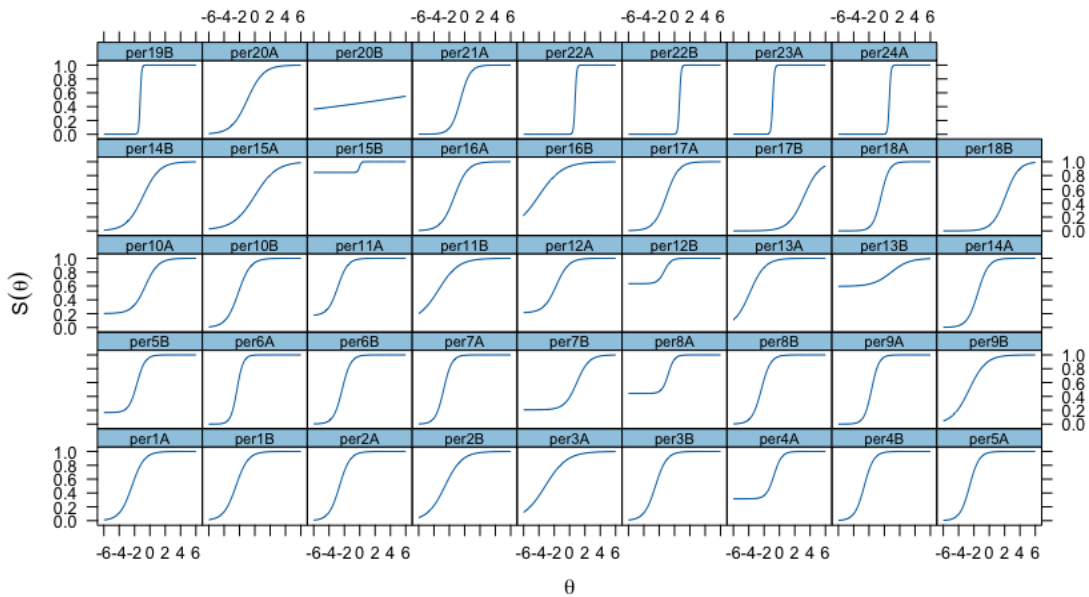


Figura 21. Curvas Características dos Itens (CCI's) do teste Percepção

A partir da observação da CIT (Figura 22), nota-se que o teste Percepção teve maior poder de informação para, principalmente, o valor de $\theta + 0,8$, seguido pelo valor de $\theta - 1,4$. Nestas regiões, a medida é mais precisa demonstrando um menor erro. Pelo exame do erro padrão, o teste demonstra maior precisão para pessoas que variam o θ entre $-3,5$ e $+1,2$, aproximadamente, indicando que o teste funciona melhor para sujeitos que têm um nível de θ baixo e médio.

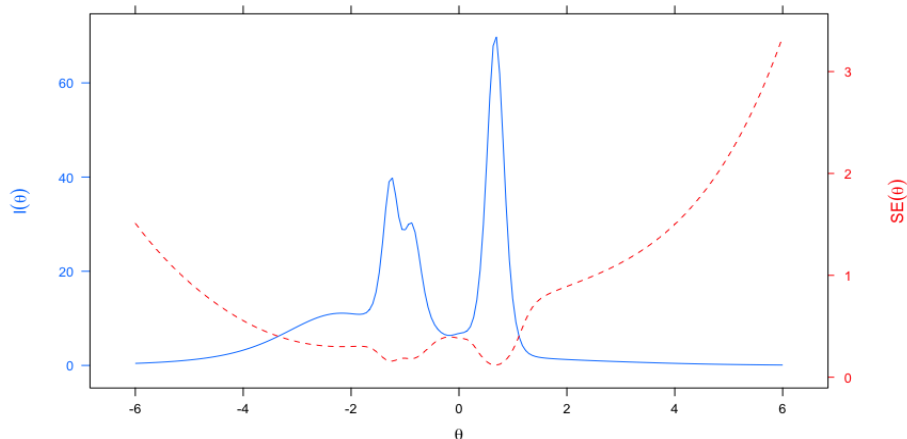


Figura 22. Curva de Informação do Teste (CIT) e erro-padrão

5.4.2 Evidências de validade baseadas na estrutura interna para o teste Percepção

A Tabela 29 apresenta os resultados das cargas fatoriais obtidas pela AFC. Nota-se que os itens com variabilidade de resposta baixa e/ou nula não participaram das análises

(>17B). As cargas fatoriais até o item 17A foram de magnitudes moderadas a fortes (exceção 13B e 15A, fraco). Itens posteriores apresentaram cargas muito baixas, acredita-se que este resultado reflita o baixo número de respondentes. O resultado para a maior parte das cargas fatoriais acima de 0,40 reforçou que os itens representam adequadamente o respectivo construto. Os índices de ajuste preconizados na literatura foram inspecionados, os resultados da AFC foram satisfatórios para o modelo unidimensional investigado, o que evidencia a validade da estrutura interna (Marsh, Hau, & Wen, 2004; Morin, Arens, & Marsh, 2016).

Tabela 29. Cargas fatoriais e índices de ajustes do teste Percepção da Bateria TSP de acordo com a Análise Fatorial Confirmatória

Item	Cargas Fatoriais
1A	0,60
1B	0,56
2A	0,70
2B	0,50
3A	0,37
3B	0,62
4A	0,55
4B	0,74
5A	0,71
5B	0,67
6A	0,88
6B	0,71
7A	0,85
7B	0,31
8A	0,60
8B	0,69
9A	0,71
9B	0,45
10A	0,45
10B	0,65
11A	0,79
11B	0,47
12A	0,54
12B	0,63
13A	0,63
13B	0,13
14A	0,64
14B	0,49
15A	0,33
15B	0,57
16A	0,62
16B	0,55
17A	0,70
17B a 27B	***

Ajustes	Percepção
RMSEA	0,05
GFI	1,00
TLI	1,00
CFI	1,00

Nota. ***Itens retirados da análise por falta de variabilidade de resposta.

5.4.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Percepção

a) Evidências de validade de critério (variáveis sociodemográficas)

Com relação à variável sexo, a Tabela 30, apresenta que as diferenças entre as médias de desempenhos de homens e mulheres foram estatisticamente significativas no teste Percepção ($p < 0,001$), no entanto, o tamanho do efeito foi inferior ao ponto de corte adotado ($d = 0,26$). Deste modo, a grande sobreposição da distribuição dos desempenhos das mulheres e dos homens no teste apontou que a diferença estatisticamente significativas na variável sexo têm pouco impacto prático sobre os resultados nos testes.

Tabela 30. Desempenho dos participantes no teste Percepção de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	Teste t	Valor p	d de Cohen
Percepção	Feminino	906	18,47	6,88	4,59	<0,001	0,26
	Masculino	468	16,69	6,61			

Os resultados apresentados na Tabela 31 revelam diferenças estatisticamente significativas ($p < 0,001$) no desempenho no teste Percepção em relação à variável idade. Além disso, o tamanho do efeito ($\eta^2 = 0,13$) foi considerado forte, o que sugere que as diferenças entre as médias de desempenho são substanciais.

Análises adicionais, com comparações *post hoc*, indicaram diferenças significativas e tamanho de efeito moderado a grande entre os desempenhos médios do grupo etário de 19 a 25 anos com todos os demais, com exceção do grupo 17-18 anos. O grupo etário ≥ 40 anos também teve tamanho de efeito acima do corte estabelecido ($d = 0,70$), com o grupo de idade 17-18 anos. Portanto, com base nessas análises, pode-se concluir que o desempenho no teste Percepção varia significativamente em relação à idade dos participantes, com diferenças particularmente notáveis entre os grupos de 19 a 25 anos e aqueles com 40 anos ou mais.

Tabela 31. Desempenho dos participantes no teste Percepção segundo a faixa etária

Teste	Faixa etária	N	Média	DP	F	Valor <i>p</i>	η^2
Percepção	15-16	83	16,31	5,50	49,19	<0,001	0,13
	17-18	272	18,11	5,91			
	19-25	529	20,39	6,17			
	26-39	258	16,74	6,95			
	≥40 anos	231	13,62	7,13			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

Os resultados para a escolaridade, conforme apresentados na Tabela 32, revelam diferenças estatisticamente significativas ($p < 0,001$) nas médias de desempenho dos grupos, levando em consideração os níveis de ensino (fundamental, médio e superior). Além disso, o tamanho do efeito ($\eta^2 = 0,27$) foi classificado como forte, sugerindo que as diferenças entre as médias de desempenho são substanciais. As comparações *post hoc* indicaram que as médias dos desempenhos dos participantes com ensino fundamental diferiram daqueles com ensino médio e superior e, as comparações entre os grupos do ensino médio e superior também foram estatisticamente significativas, com magnitudes de efeito classificadas como moderadas a fortes.

Tabela 32. Desempenho dos participantes no teste Percepção segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor <i>p</i>	η^2
Percepção	Fundamental	251	10,74	5,34	260,38	<0,001	0,27
	Médio	398	17,61	6,04			
	Superior	725	20,47	5,87			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

5.4.4 Evidências de confiabilidade do teste Percepção

Evidências de confiabilidade foram obtidas pelos coeficientes de consistência interna alfa de Cronbach e ômega de McDonald, os resultados foram α 0,81 e Ω 0,87, respectivamente. A evidência de estabilidade temporal (Tabela 33) demonstrou coeficiente de correlação de *Pearson* de magnitude moderada ($r=0,66$) e estatisticamente significativo ($p < 0,001$). Concluiu-se, portanto, que o teste Percepção apresenta adequada consistência interna e demonstra também uma estabilidade temporal satisfatória.

Tabela 33. Estatísticas do estudo de estabilidade temporal do teste Percepção (n=78)

Teste	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	r ₁₂	Valor p
Percepção	21,94	5,10	25,74	5,30	0,66	<0,001

Nota. M: Média, DP: Desvio-Padrão, r₁₂: correlação bivariada, valor significativo se $p < 0,05$.

Em suma, a pesquisa sobre a validade interna por meio da AFC revelou um bom ajuste ao modelo unidimensional do teste Percepção (Hair et al., 2009; Morin; Arens; Marsh, 2016). No entanto, o item 13B destacou-se como necessitando de revisão devido aos seus resultados insatisfatórios na AFC e nos parâmetros individuais. Além disso, outros itens, como o 12B e o 15B, também apontam para a necessidade de reavaliação devido aos seus resultados inadequados nos parâmetros de discriminação e de chute. De maneira geral, os resultados da TRI sugerem que os itens do teste Percepção são predominantemente de dificuldade fácil e média, indicando também que o teste não é indicado para sujeitos com habilidades elevadas (Peres, Lessa, Valentini, 2019).

A verificação da validade de critério por meio da análise das diferenças de médias entre grupos em relação às variáveis sociodemográficas forneceu uma evidência substancial para a validade do teste, uma vez que os resultados estão em consonância com a literatura existente (Burgaleta et al., 2012; Kaufman et al., 2009; Primi et al., 2022). As análises de confiabilidade, utilizando índices de consistência interna e estabilidade temporal, apresentaram resultados satisfatórios, em conformidade com as normas atuais (CFP, 2022). Isso sugere que o teste Percepção possui uma base teórica e métrica robusta, sendo adequado para avaliar habilidade perceptual em indivíduos com 15 anos ou mais, iniciando com aqueles com níveis de escolaridade a partir do ensino fundamental incompleto.

5.5 Teste Precisão

Participaram desse estudo 1342 sujeitos, a amostragem foi por conveniência, residentes nos estados de Minas Gerais (N=794), Distrito Federal (N=294), Espírito Santo (N=152) e Ceará (N=102). Foram 890 mulheres (66,3%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=27,0; DP=11,87). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=707), seguido pelo Ensino Médio completo ou incompleto (N=391) e Ensino Fundamental completo ou incompleto (N=244).

Para avaliação da normalidade dos dados observa-se (Tabela 34) que o teste estatístico Shapiro-Wilk ($W=0,986$; $p<0,001$) demonstra pelo parâmetro de significância que esta

amostra é diferente de uma distribuição normal, no entanto, nota-se um valor W muito próximo a 1, o qual pode indicar uma distribuição aproximadamente normal. Em função da possibilidade do erro tipo II no teste Shapiro-Wilk, outros dados estatísticos e dois parâmetros de inspeção visual também foram analisados. Nota-se as medidas de tendência central muito próximas, o resultado de uma pequena assimetria positiva (0,322) e a curtose (-0,066) indica que os dados se aproximam de uma distribuição normal (mesocúrtica) (Henderson,2006; Miot, 2017; Vetter, 2017).

Tabela 34. Estatísticas descritivas do teste Precisão

Estatística Descritiva	Teste Precisão
N	1.342
Média	24,71
Mediana	23,0
Moda	23,0
Desvio-padrão	8,91
Amplitude	47
Mínimo	1
Máximo	48
Assimetria	0,322
Erro-padrão da Assimetria	0,067
Curtose	-0,066
Erro-padrão da Curtose	0,133
W de Shapiro-Wilk	0,986
p Shapiro-Wilk	< 0,001

O diagrama Q-Q (Figura 23) aponta que os dados do teste Precisão se aproximam de uma distribuição normal (Miot, 2017). De maneira complementar foi gerado um histograma de frequência (Figura 24), a inspeção visual dos dados confirma os achados de uma distribuição aproximadamente normal com a curva da distribuição apresentando uma pequena inclinação para a esquerda (Dancey & Reidy, 2019).

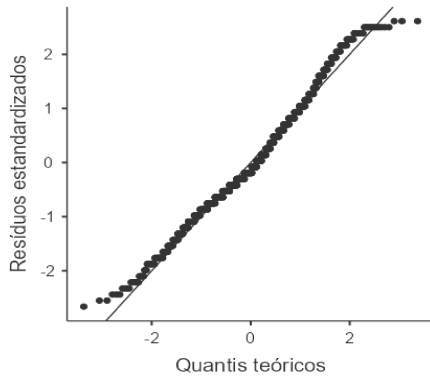


Figura 23. Diagrama Q-Q teste Precisão

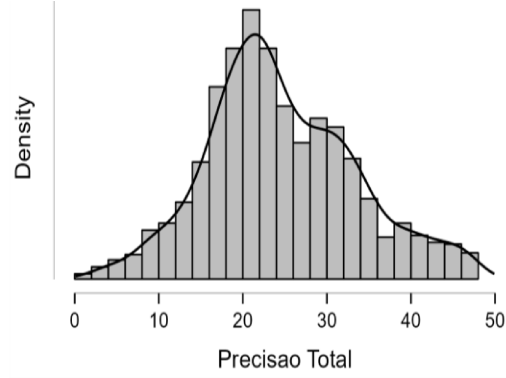


Figura 24. Curva de Gauss do teste Precisão

5.5.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Precisão

Os resultados dos parâmetros individuais dos itens do teste Precisão podem ser conferidos na Tabela 35. Os indicadores de discriminação variaram entre 0,368 (Q45) a 5,344 (Q46), com média de discriminação de 1,365 (DP=0,799). Ressalta-se que alguns itens apresentaram discriminação baixa, são eles: Q29 ($a=0,471$), Q42 ($a=0,450$), Q45 ($a=0,368$) e Q47 ($a=0,444$). Por fim, o item Q46 ($a=5,344$) apresenta índice irrealista de discriminação ($a > 3,0$; Barbetta et al., 2014). Observa-se que todos os itens apresentaram nível de dificuldade menor que o ponto médio, indicando que o teste Precisão apresenta itens fáceis, com valores variando entre -8,425 e -0,068 ($M = -2,867$, $DP=1,298$).

Os parâmetros (c) variaram entre 0,000 e 0,849 ($M = 0,137$, $DP=0,230$), quatro dos 48 itens apresentaram probabilidade de acerto ao acaso acima do ideal ($c > 0,50$), são eles: Q3 ($c=0,728$), Q14 ($c= 0,564$), Q36 ($c=0,849$) e Q38 ($c=0,586$). Dessa maneira, embora estes itens tenham apresentado resultados satisfatórios para o grau de discriminação e dificuldade, o aumento percebido do parâmetro (c) diminui a sua qualidade de informação, pois há uma alta probabilidade de acerto casual (Andrade, Laros & Gouveia, 2010).

Tabela 35. Parâmetros individuais dos itens do teste Precisão (N=1202)

Item	N	a	b	c
pre1	1193	0,843	-4,523	0,006
pre2	1202	1,055	-2,753	0,002
pre3	1182	1,768	-1,163	0,728
pre4	1187	0,828	-4,326	0,005
pre5	1169	1,300	-2,539	0,001
pre6	1179	2,245	-2,168	0,467
pre7	1177	1,193	-3,619	0,002
pre8	1170	3,138	-2,843	0,001

pre9	1160	1,970	-3,216	0,002
pre10	1155	1,827	-2,910	0,370
pre11	1141	0,745	-3,183	0,003
pre12	1143	1,302	-1,318	0,375
pre13	1130	1,028	-2,914	0,004
pre14	1124	1,350	-1,397	0,564
pre15	1118	1,060	-3,331	0,003
pre16	1087	0,740	-2,337	0,002
pre17	1074	1,375	-3,254	0,002
pre18	1055	1,076	-2,606	0,004
pre19	989	2,073	-1,009	0,412
pre20	942	1,849	-1,535	0,000
pre21	893	1,277	-3,185	0,002
pre22	845	0,751	-3,767	0,004
pre23	790	1,059	-1,183	0,200
pre24	718	0,895	-0,068	0,414
pre25	720	1,002	-5,076	0,009
pre26	690	1,818	-3,048	0,002
pre27	662	1,360	-3,687	0,002
pre28	595	1,388	-2,629	0,001
pre29	568	0,471	-4,390	0,005
pre30	511	1,104	-2,601	0,001
pre31	463	1,622	-1,787	0,471
pre32	432	1,834	-2,621	0,001
pre33	399	0,955	-2,485	0,003
pre34	360	1,508	-2,922	0,002
pre35	305	0,791	-2,792	0,002
pre36	290	1,221	-2,261	0,849
pre37	244	0,848	-2,266	0,031
pre38	260	0,909	-3,080	0,586
pre39	230	1,600	-1,946	0,409
pre40	227	0,882	-4,272	0,008
pre41	198	1,362	-2,595	0,002
pre42	174	0,450	-2,581	0,007
pre43	158	1,726	-1,933	0,410
pre44	160	1,559	-3,239	0,003
pre45	140	0,368	-4,207	0,009
pre46	133	5,344	-3,548	0,003
pre47	123	0,444	-8,425	0,186
pre48	106	2,206	-2,063	0,004

Os parâmetros individuais dos itens podem ser interpretados através das Curvas Características dos Itens (CCIs). É possível observar (Figura 25) itens com discriminação

baixa, através da pouca inclinação da curva (Q29, Q42, Q45 e Q47.), bem como aqueles com discriminação alta (Q8 e Q46, por exemplo), com curvas alcançando quase o ângulo reto (90°). Além disso, o parâmetro de chute (c) é facilmente percebido, uma vez que está relacionado a altura do início da curva, indicando a alta probabilidade de acerto ao acaso observada, por exemplo, nos itens Q3, Q36 e Q47.

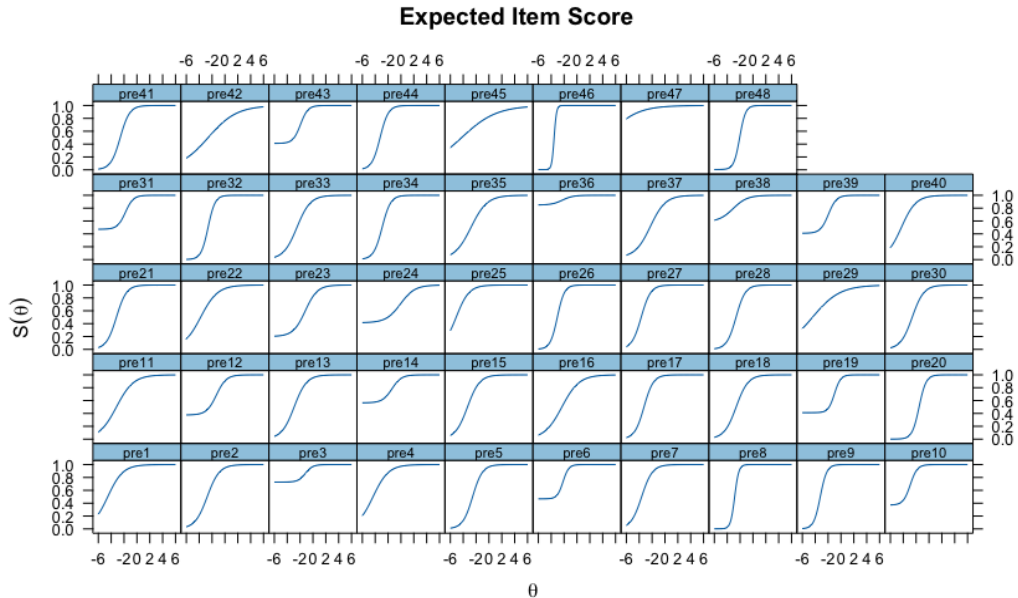


Figura 25. Curvas Características dos Itens (CCI's) do teste Precisão

A partir da observação da CIT (Figura 26) é possível perceber que o teste Precisão teve maior poder de informação para, principalmente, o valor de θ em torno de -3,6, indicando que nesta região a medida é mais precisa. A análise do erro padrão revela que o teste apresenta maior precisão para indivíduos com θ s variando entre -6 e 0, aproximadamente. Isso sugere que o teste é mais eficaz para sujeitos com níveis de θ mais baixos, o que está alinhado com os resultados de todos os itens com parâmetros de dificuldade negativos.



Figura 26. Curva de Informação do Teste (CIT) e erro-padrão

5.5.2 Evidências de validade baseadas na estrutura interna do teste Precisão

Os resultados das análises de cargas fatoriais (Tabela 36) revelam que os itens com baixa ou nula variabilidade de resposta no teste Precisão não foram incluídos nas análises, a partir do item 32. Observa-se que sete itens exibiram cargas fatoriais baixas: Q1, Q4, Q11, Q16, Q22, Q24 e Q29. Ao examinar os parâmetros individuais obtidos pela TRI para esses itens, nota-se que, exceto pelo item 24, todos os outros apresentaram níveis muito baixos de dificuldade (b), variando entre -4,523 e -2,337. Isso indica que é necessário ter um θ baixo para obter uma probabilidade de acerto de 50% nos itens analisados. Essa condição observada pode ter contribuído para a baixa representatividade desses itens no que diz respeito ao construto avaliado. Além disso, o item 29 demonstra baixo poder discriminativo. Esses resultados sugerem a necessidade de realizar estudos para revisão e possíveis ajustes ou dos itens mencionados.

Os demais resultados das cargas fatoriais foram acima de 0,40 o que reforça a qualidade geral dos itens representam o respectivo construto. Os índices de ajuste preconizados na literatura foram inspecionados, os resultados da AFC foram satisfatórios para o modelo unidimensional investigado, o que evidencia a validade da estrutura interna do teste Precisão (Marsh, Hau, & Wen, 2004; Morin, Arens, & Marsh, 2016).

Tabela 36. Cargas fatoriais e índices de ajustes do teste Precisão da Bateria TSP de acordo com a Análise Fatorial Confirmatória

Item	Cargas Fatoriais
1	0,33
2	0,46
3	0,44
4	0,32
5	0,57
6	0,68
7	0,49
8	0,98
9	0,73
10	0,64
11	0,35
12	0,47
13	0,45
14	0,44
15	0,44
16	0,38
17	0,54
18	0,47

19	0,57
20	0,69
21	0,53
22	0,33
23	0,47
24	0,32
25	0,53
26	0,80
27	0,81
28	0,63
29	0,23
30	0,50
31	0,61
32	0,70
33 a 48	***
Ajustes	Precisão
RMSEA	0,01
GFI	1,00
TLI	1,00
CFI	1,00

Nota. ***Itens retirados da análise por falta de variabilidade de resposta.

5.5.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Precisão

a) Evidências de validade de critério (variáveis sociodemográficas)

A Tabela 37 apresenta que as diferenças entre as médias de desempenhos de homens e mulheres não foram estatisticamente significativas no teste Precisão ($p=0,86$), com o tamanho do efeito inferior ao ponto de corte adotado ($d= -0,01$). Sendo assim, não foi observada uma diferença significativa no desempenho no teste Precisão em função do sexo.

Tabela 37. Desempenho dos participantes no teste Precisão de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	Teste t	Valor p	d de Cohen
Precisão	Feminino	890	24,68	8,94	-0,18	0,86	-0,01
	Masculino	452	24,77	8,86			

Os resultados das Tabela 38, indicaram a presença de diferenças estatisticamente significativas entre os grupos etários. Considerando o tamanho do efeito, as diferenças entre as médias foram fortes ($\eta^2=0,19$). As comparações *post hoc* revelaram diferenças significativas e tamanhos de efeito acima do ponto de corte estabelecido entre os

desempenhos médios do grupo com idade ≥ 40 anos em comparação com todos os outros. E, entre o grupo etário de 26 a 39 anos com os dois grupos entre 17 a 25 anos ($d=0,57$ e $d=0,67$, respectivamente).

Tabela 38. Desempenho dos participantes no teste Precisão segundo a faixa etária

Teste	Faixa etária	N	Média	DP	F	Valor p	η^2
Precisão	15-16	65	26,41	8,17	78,83	<0,001	0,19
	17-18	264	27,96	8,36			
	19-25	524	27,19	8,16			
	26-39	246	22,62	7,80			
	≥ 40 anos	242	17,38	7,49			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor p = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

As diferenças nas médias de desempenho dos grupos observadas na Tabela 39 foram estatisticamente significativas e com tamanho de efeito forte ($\eta^2=0,14$). As comparações *post hoc* apontaram que as médias dos desempenhos dos participantes com ensino fundamental diferiram significativamente das médias dos grupos com ensino médio e superior, ambos com grande tamanho de efeito. As comparações entre os grupos do ensino médio e superior não foram estatisticamente significativas ($p=0,70$).

Tabela 39. Desempenho dos participantes no teste Precisão segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor p	η^2
Precisão	Fundamental	244	17,73	8,04	106,19	<0,001	0,14
	Médio	391	25,99	8,97			
	Superior	707	26,41	7,97			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor p = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

5.5.4 Evidências de confiabilidade do teste Precisão

Evidências de confiabilidade foram obtidas pelos coeficientes de consistência interna alfa de Cronbach e ômega de McDonald, os resultados foram α 0,71 e Ω 0,71, respectivamente. Portanto, observa-se que o teste apresenta adequada consistência interna. A evidência de estabilidade temporal (Tabela 40) demonstrou coeficiente de correlação de *Pearson* de magnitude moderada ($r=0,68$) e estatisticamente significativo ($p<0,001$), demonstrando também uma estabilidade temporal satisfatória.

Tabela 40. Estatísticas do estudo de estabilidade temporal do teste Precisão (n=82)

Teste	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	r_{12}	Valor p
Precisão	30,08	7,76	31,16	7,79	0,68	<0,001

Nota. M: Média, DP: Desvio-Padrão, r_{12} : correlação bivariada, valor significativo se $p < 0,05$.

De maneira geral, a pesquisa sobre a validade interna por meio da AFC revelou um bom ajuste ao modelo unidimensional do teste Precisão (Hair et al., 2009; Morin; Arens; Marsh, 2016). No entanto, alguns itens requerem revisão devido aos baixos valores de discriminação, indicando uma capacidade limitada de distinguir entre sujeitos com habilidades semelhantes, e também devido aos altos valores de acerto ao acaso. De forma geral, os resultados da TRI sugerem que os itens do teste de Precisão tendem a ser predominantemente fáceis, o que sugere que o teste pode ser mais apropriado para indivíduos com habilidades cognitivas mais baixas (Peres, Lessa, Valentini, 2019).

A análise das diferenças nas médias entre grupos, considerando as variáveis sexo, faixa etária e escolaridade, fornece uma evidência da validade de critério para o teste, uma vez que os resultados estão alinhados com o que é encontrado na literatura (Burgaleta et al., 2012; Kaufman et al., 2009; Primi et al., 2022). As análises de confiabilidade, com base nos índices de consistência interna e estabilidade temporal, demonstraram resultados satisfatórios, em concordância com as normas vigentes (CFP, 2022).

5.6 Teste Partes

Participaram desse estudo 1424 sujeitos, a amostragem foi por conveniência, residentes nos estados de Minas Gerais (N=837), Distrito Federal (N=301), Espírito Santo (N=152) e Ceará (N=134). Foram 948 mulheres (66,6%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=26,8; DP=11,62). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=771), seguido pelo Ensino Médio completo ou incompleto (N=396) e Ensino Fundamental completo ou incompleto (N=257).

Para avaliação da normalidade dos dados observa-se (Tabela 41) que o teste estatístico Shapiro-Wilk ($W=0,986$; $p<0,001$) demonstra pelo parâmetro de significância que esta amostra é diferente de uma distribuição normal, no entanto, nota-se um valor W muito próximo a 1, o qual pode indicar uma distribuição aproximadamente normal. É também

observada uma pequena assimetria negativa (-0,398) e a curtose (-0,126) indica uma distribuição levemente achatada (Henderson,2006; Miot, 2017; Vetter, 2017).

Tabela 41. Estatísticas descritivas do teste Partes

Estatística descritiva	Teste Partes
N	1.424
Média	26,83
Mediana	27,0
Moda	31,0
Desvio-padrão	8,21
Amplitude	44
Mínimo	2
Máximo	46
Assimetria	-0,398
Erro-padrão da Assimetria	0,065
Curtose	-0,126
Erro-padrão da Curtose	0,130
W de Shapiro-Wilk	0,986
p Shapiro-Wilk	<0,001

Conforme observado no diagrama Q-Q (Figura 27), os dados do teste Partes se aproximam de uma distribuição normal (Miot, 2017). De maneira complementar foi gerado um histograma de frequência (Figura 28), a inspeção visual dos dados confirma os achados de uma distribuição aproximadamente normal, com uma leve inclinação da cauda para a esquerda (Dancey & Reidy, 2019).

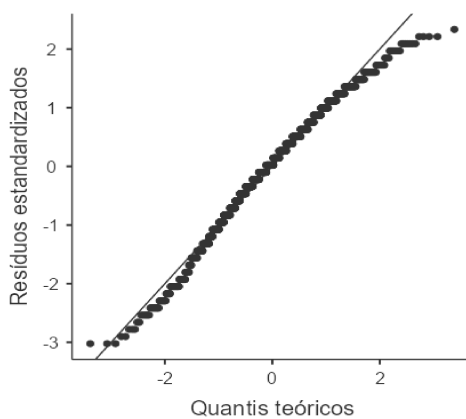


Figura 27. Diagrama Q-Q teste Partes

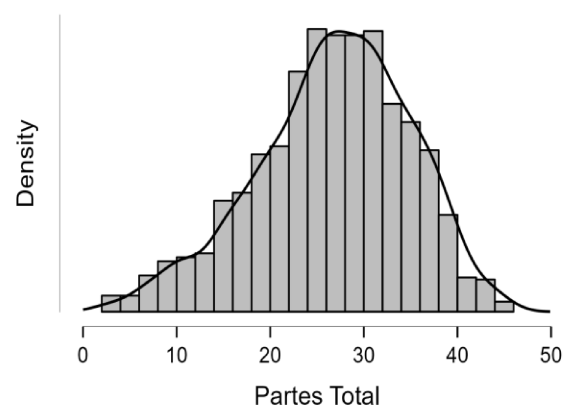


Figura 28. Curva de Gauss do teste Partes

5.6.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Partes

Os resultados dos parâmetros individuais foram calculados com todos os itens do teste Partes e podem ser conferidos na Tabela 42. Os indicadores de discriminação (a) variaram entre 0,512 (Q29) e 7,764 (Q45), com média de 1,800 (DP=1,16). Os itens Q43 e Q45 apresentaram índices muito agudos para discriminação, quase perfeitos ($a > 3,0$; Barbetta et al., 2014). Para os parâmetros de dificuldade (b), os resultados variaram entre -4,274 (Q1) e +0,891 (Q38), com média de -1,372 (DP= 1,10). Os resultados indicam que a maioria dos itens apresentou um nível de dificuldade negativo, sugerindo que são itens predominantemente fáceis. No entanto, é importante notar que muitos desses itens estão próximos de zero, o que sugere a presença também de itens com dificuldade média.

Os parâmetros do chute (c), variaram entre 0,002 (Q8) a 0,806 (Q4), no qual apenas um item apresentou resultados acima do ideal ($c > 0,50$). Embora o item Q4 tenha apresentado resultados satisfatórios para o grau de discriminação e dificuldade, o aumento percebido do parâmetro (c) diminui a sua qualidade de informação, pois há uma alta probabilidade de acerto casual (Andrade, Laros & Gouveia, 2010).

Tabela 42. Parâmetros individuais dos itens do teste Partes (N = 1268)

Item	N	a	b	c
par1	1262	1,133	-4,274	0,020
par2	1268	1,718	-3,030	0,004
par3	1259	1,335	-0,737	0,321
par4	1264	3,623	-0,676	0,806
par5	1261	2,397	-2,275	0,227
par6	1252	2,103	-2,214	0,253
par7	1264	3,223	-2,636	0,004
par8	1257	1,888	-2,541	0,002
par9	1245	1,866	-2,335	0,081
par10	1247	1,748	-2,116	0,201
par11	1252	1,630	-1,253	0,374
par12	1254	1,168	-3,126	0,005
par13	1250	1,515	-2,698	0,002
par14	1221	1,099	-2,656	0,023
par15	1243	1,083	-1,622	0,006
par16	1204	1,476	-2,205	0,136
par17	1201	0,984	-1,832	0,003
par18	1215	1,752	-1,458	0,361
par19	1187	0,769	-1,834	0,086

par20	1133	1,727	-0,757	0,339
par21	1162	2,365	-2,185	0,077
par22	1100	1,571	-1,806	0,211
par23	1098	0,966	-0,994	0,013
par24	1097	2,654	-1,922	0,373
par25	1066	2,522	-1,382	0,320
par26	1041	0,732	-1,365	0,002
par27	1028	0,830	-1,873	0,017
par28	876	1,258	0,380	0,307
par29	916	0,512	-3,057	0,007
par30	833	2,449	-0,552	0,454
par31	707	1,705	-0,563	0,124
par32	701	1,934	-1,719	0,041
par33	689	0,919	-1,382	0,004
par34	582	1,810	-0,901	0,279
par35	521	2,374	-0,949	0,320
par36	447	1,216	-0,580	0,203
par37	377	1,630	-0,085	0,047
par38	325	1,376	0,891	0,227
par39	360	2,409	-1,683	0,096
par40	256	1,420	-0,009	0,117
par41	247	1,783	-0,264	0,238
par42	165	1,574	0,054	0,135
par43	172	4,296	-0,332	0,231
par44	139	1,600	-0,967	0,004
par45	111	7,764	0,846	0,413
pat46	101	0,731	-0,165	0,005
par47	94	1,010	-0,191	0,002
par48	103	0,747	-0,804	0,003

As CCIs do teste de Partes foi analisada, possibilitando a análise gráfica dos parâmetros individuais dos itens (Figura 29). O item Q4, que apresentou parâmetro c muito alto ($c=0,806$) demonstrou o início da curva mais alto do que os demais, os itens Q43 e Q45, com índices de discriminação muito altos, geraram CCIs agudas ($a=4,296$ e $7,764$, respectivamente). Por outro lado, percebe-se que na medida que a discriminação diminui, a curva fica menos perceptível, como no caso do item Q29 ($a=0,512$).

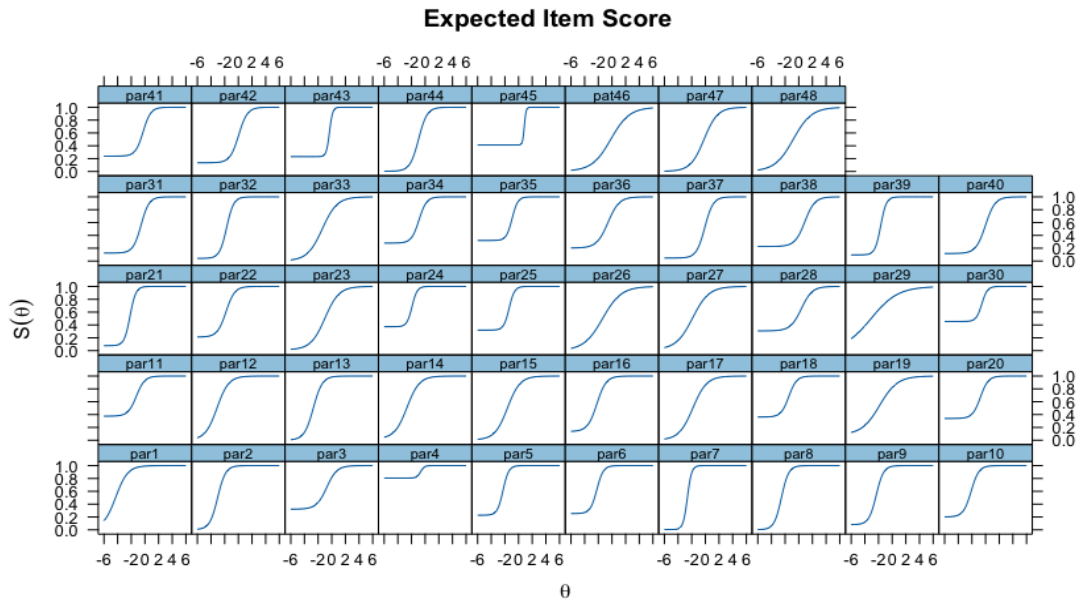


Figura 29. Curvas Características dos Itens (CCI's) do teste Partes

Observa-se pela Figura 30, que o teste Partes teve maior poder de informação para, principalmente, dois valores de θ : -2 e -0,2, aproximadamente (ambas dentro do intervalo negativo, abaixo de zero). Existe uma certa melhoria na precisão do teste também para o θ +1. Nessas faixas, a medida é mais precisa, demonstrando um menor erro. A análise do erro padrão revela que o teste apresenta maior precisão para pessoas com um intervalo de θ entre -4 e +2, aproximadamente, indicando que o teste é mais eficaz para sujeitos com um nível de habilidade baixo a médio.

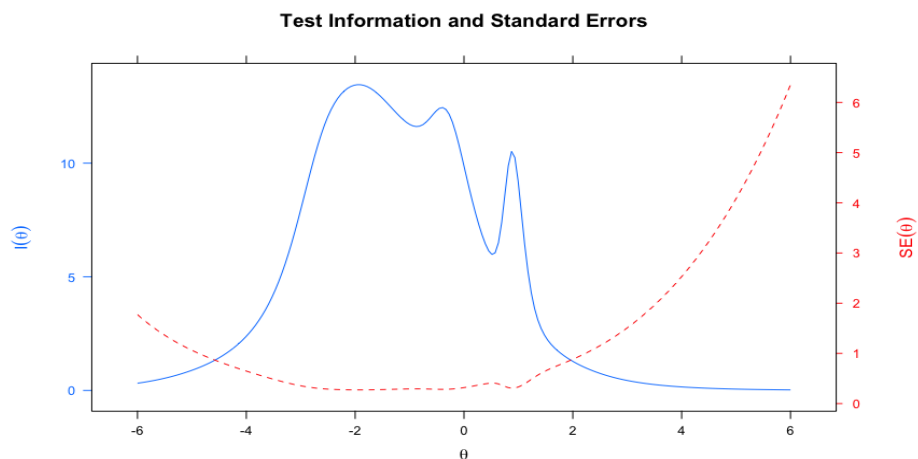


Figura 30. Curva de Informação do Teste (CIT) e erro-padrão

5.6.2 Evidências de validade baseadas na estrutura interna do teste Partes

Na Tabela 43, os resultados da AFC são apresentados, onde os itens com baixa ou nula variabilidade de resposta foram excluídos das análises (Q1, Q44 a Q48). Ao examinar os parâmetros individuais obtidos pela TRI para esses itens, nota-se que, em geral, os itens 44 a 48 apresentaram parâmetros individuais adequados. Por outro lado, o item Q1 demonstrou um nível de dificuldade muito fácil (-4,274), sugerindo que um *theta* baixo já possui uma alta probabilidade de acerto no item. A baixa variabilidade de resposta observada pode ter contribuído para a representação limitada desse item em relação ao construto avaliado.

As cargas fatoriais dos outros itens foram, em sua maioria, superiores a 0,40 sugerindo que estes itens representam de forma adequada o construto pretendido. Os índices de ajuste recomendados na literatura foram verificados, e os resultados da AFC foram considerados satisfatórios para validar a estrutura interna do modelo unidimensional avaliado (Marsh, Hau, & Wen, 2004; Morin, Arens, & Marsh, 2016).

Tabela 43. Cargas fatoriais e índices de ajustes do teste Partes da Bateria TSP de acordo com a Análise Fatorial Confirmatória

Item	Cargas Fatoriais
1	***
2	0,60
3	0,64
4	0,41
5	0,71
6	0,99
7	0,81
8	0,78
9	0,97
10	0,53
11	0,86
12	0,73
13	0,76
14	0,36
15	0,78
16	0,57
17	0,65
18	0,71
19	0,47
20	0,83
21	0,92
22	0,83
23	0,46
24	0,88
25	0,90
26	0,55
27	0,75

28	0,32
29	0,42
30	0,77
31	0,67
32	0,82
33	0,68
34	0,66
35	0,89
36	0,58
37	0,61
38	0,37
39	0,81
40	0,59
41	0,67
42	0,57
43	0,69
44 a 48	***
Ajustes	Partes
RMSEA	0,03
GFI	0,92
TLI	0,96
CFI	0,96

Nota. ***Itens retirados da análise por falta de variabilidade de resposta.

5.6.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Partes

a) Evidência de validade convergente

Foram 83 sujeitos submetidos às provas de raciocínio da BPR-5. A maioria do sexo feminino (68,7%), a idade variou entre 18 e 43 anos ($M=21,47$; $DP=4,79$), com ensino superior incompleto e completo. A análise de correlação de *Pearson* apontou evidências de validade convergente para o teste Partes com correlação moderada e estatisticamente significativa para as provas de RA ($r=0,46$; $p<0,001$) e RM ($r=0,44$; $p<0,001$), a magnitude fraca, mas estatisticamente significativa foi observada para a RE ($r=0,33$; $p=0,002$) e para o escore geral da BPR-5, correlação moderada ($r=0,49$; $p<0,001$).

Foram submetidos 27 sujeitos ao teste Partes e ao G-38. Participaram 21 mulheres (77,78%), a idade variou entre 19 e 62 anos ($M=27,7$; $DP=11,15$), com ensino superior incompleto e completo. Os resultados analisados apontaram evidências de validade convergente para o teste Partes com correlação moderada e estatisticamente significativa ($r=0,54$; $p=0,004$).

b) Evidências de validade de critério (variáveis sociodemográficas)

Com relação à variável sexo, a Tabela 44, mostra que o resultado do desempenho médio não foi estatisticamente significativo no teste Partes ($p=0,05$) e o tamanho do efeito foi pequeno ($d=-0,11$).

Tabela 44. Desempenho dos participantes no teste Partes de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	Teste <i>t</i>	Valor <i>p</i>	<i>d</i> de Cohen
Partes	Feminino	948	26,52	8,17	-1,96	0,05	-0,11
	Masculino	476	27,43	8,26			

A Tabela 45 indicou a presença de diferenças estatisticamente significativas entre os grupos etários ($p<0,001$). Considerando o tamanho do efeito, as diferenças entre as médias foram fortes ($\eta^2 =0,12$). De forma geral, as comparações *post hoc* revelaram diferenças significativas e tamanhos de efeito acima de $d=0,50$ entre o grupo com idade ≥ 40 anos e todos os outros grupos etários. O mesmo foi observado para o grupo de 17-18 anos em comparação com o grupo com idade 26 a 39 anos.

Tabela 45. Desempenho dos participantes no teste Partes segundo a faixa etária

Teste	Faixa etária	N	Média	DP	F	Valor <i>p</i>	η^2
Partes	15-16	86	28,27	6,94	48,89	<0,001	0,12
	17-18	268	29,20	7,74			
	19-25	562	28,60	7,46			
	26-39	260	25,29	7,77			
	≥ 40 anos	247	21,30	8,39			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

As diferenças nas médias de desempenho dos grupos em função da escolaridade foram estatisticamente significativas e com tamanho de efeito forte ($\eta^2 =0,13$) (Tabela 46). As comparações *post hoc* apontaram que as médias dos desempenhos dos participantes com ensino fundamental diferiram significativamente daqueles com ensino médio e superior ($p<0,001$), com grandes tamanhos de efeito, *d* de Cohen acima 0,80. As comparações entre os grupos do ensino médio e superior não foram estatisticamente significativas ($p=0,45$; $d=0,07$).

Tabela 46. Desempenho dos participantes no teste Partes segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor p	η^2
Partes	Fundamental	257	20,58	7,90	104,68	<0,001	0,13
	Médio	396	28,58	7,83			
	Superior	771	28,00	7,51			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor p = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

5.6.4 Evidências de confiabilidade do teste Partes

Evidências de confiabilidade foram obtidas pelos coeficientes de consistência interna alfa de Cronbach e ômega de McDonald, os resultados foram α 0,91 e Ω 0,92, respectivamente. Concluiu-se, portanto, que o teste apresenta adequada consistência interna. A evidência de estabilidade temporal (Tabela 47) demonstrou coeficiente de correlação de *Pearson* de magnitude moderada ($r=0,64$) e estatisticamente significativo ($p<0,001$), demonstrando também uma estabilidade temporal satisfatória.

Tabela 47. Estatísticas do estudo de estabilidade temporal do teste Partes (n=90)

Teste	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	r_{12}	Valor p
Partes	31,23	6,48	35,49	4,94	0,64	<0,001

Nota. M: Média, DP: Desvio-Padrão, r_{12} : correlação bivariada, valor significativo se $p < 0,05$.

Os resultados do teste Partes revelaram uma boa consistência interna, com índices de ajustes adequados ao modelo unidimensional observado pela AFC. Tanto pelas cargas fatoriais quanto pelos resultados da TRI, os itens Q1, Q4 e Q29 demonstram necessária revisão em futuras pesquisas. De modo geral, os resultados da TRI indicaram que, apesar da presença de alguns itens de dificuldade média, a maioria deles tende a ser fácil. Além disso, o teste demonstrou ser mais preciso para avaliar sujeitos com habilidades cognitivas médias e baixas (Peres, Lessa, Valentini, 2019).

Foram encontradas evidências de validade convergente entre o teste Partes e as provas de RA, RM e RE. As correlações foram mais elevadas com as provas de habilidade abstrata e mecânica. Esperava-se inicialmente uma correlação mais alta com o teste que mede raciocínio espacial da BPR-5. No entanto, a correlação de magnitude fraca pode ser atribuída às características dos itens da prova RE, que envolvem a manipulação de representações visuais complexas, especialmente com cubos tridimensionais (Primi, Correia, & Almeida, 2018). Em

contraste, os itens do teste Partes não exigem tanta complexidade na rotação mental ou na manipulação de imagens.

De maneira geral, os resultados dos estudos das propriedades psicométricas do teste Partes indicaram validade convergente satisfatórios para demonstrar a capacidade do teste avaliar com eficiência o construto pretendido e a comum relação observada com a capacidade de inteligência fluída (Gomes & Borges, 2009b). Além disso, a análise da validade de critério também revelou resultados consistentes com a literatura, indicando a capacidade do teste de discriminar grupos distintos com base na idade, escolaridade e sexo (Kaufman et al., 2009; Primi et al., 2022). Os estudos de confiabilidade do teste mostraram resultados robustos, em linha com as diretrizes atuais (CFP, 2022), indicando uma base teórica e métrica sólida para avaliar o raciocínio espacial em indivíduos com 15 anos ou mais, com diferentes níveis de escolaridade.

5.7 Teste Dimensão

Participaram desse estudo 1385 sujeitos, a amostragem foi por conveniência, residentes nos estados de Minas Gerais (N=803), Distrito Federal (N=295), Espírito Santo (N=152) e Ceará (N=135). Foram 920 mulheres (66,4%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=26,7; DP=11,59). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=736), seguido pelo Ensino Médio completo ou incompleto (N=391) e Ensino Fundamental completo ou incompleto (N=258).

Para avaliação da normalidade dos dados observa-se (Tabela 48) que o teste estatístico Shapiro-Wilk ($W=0,986$; $p<0,001$) demonstra pelo parâmetro de significância que esta amostra é diferente de uma distribuição normal, no entanto, nota-se um valor W muito próximo a 1, o qual pode indicar uma distribuição aproximadamente normal. Nota-se também uma pequena assimetria negativa (Henderson,2006; Miot, 2017; Vetter, 2017).

Tabela 48. Estatísticas descritivas do teste Dimensão

Estatística descritiva	Teste Dimensão
N	1385
Média	24,40
Mediana	25,0
Moda	28,0
Desvio-padrão	10,42
Amplitude	48

Mínimo	0
Máximo	48
Assimetria	-0,212
Erro-padrão da Assimetria	0,066
Curtose	-0,555
Erro-padrão da Curtose	0,131
W de Shapiro-Wilk	0,986
p Shapiro-Wilk	<0,001

Conforme observado no diagrama Q-Q (Figura 31), a maior parte da distribuição dos quantis da amostra de dados está distribuída de maneira linear sobre a reta esperada para a distribuição normal (Miot, 2017). O diagrama Q-Q aponta que os dados do teste Dimensão se aproximam de uma distribuição normal. De maneira complementar foi gerado um histograma de frequência (Figura 32), a inspeção visual dos dados confirma os achados de uma distribuição aproximadamente normal, com uma leve inclinação da cauda para a esquerda (Dancey & Reidy, 2019).

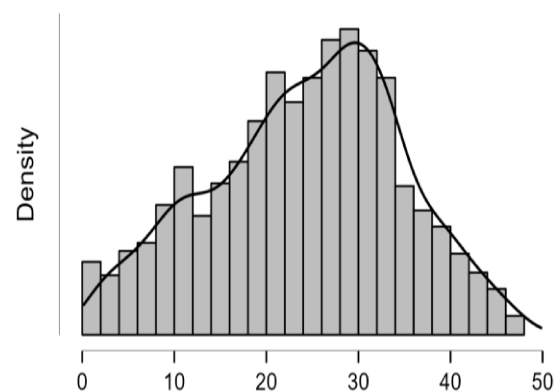
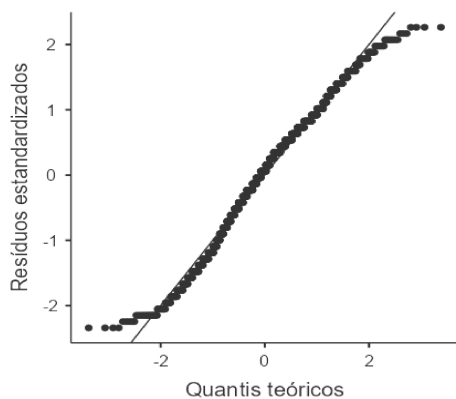


Figura 31. Diagrama Q-Q teste Dimensão

Figura 32. Curva de Gauss do teste Dimensão

5.7.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Dimensão

Os resultados dos parâmetros individuais dos itens do teste Dimensão podem ser conferidos na Tabela 49. O índice de discriminação variou entre 1,416 (Q42) e 9,735 (Q33), com média de 3,021 (DP=1,78). É importante ressaltar que os itens Q16 (a=4,935), Q21 (a=4,280), Q26 (a=5,151), Q33 (a=9,735), Q35 (a=6,150), Q37 (a=6,343), Q39 (a=5,647), Q40 (a=7,505), Q41 (a=4,071), Q47 (a=4,218) e Q48 (a=5,168), apresentaram discriminação irrealista (a >3,0; Barbetta et al., 2014).

Os índices de dificuldade (b) do teste Dimensão variaram entre -2,015 (Q1) a 1,711 (Q47), com média de dificuldade de -0,572 (DP=1,14), indicando que a maior parte dos itens apresentaram dificuldade baixa e moderada por estar próximo a zero. Por fim, o índice de chute (c) apresentou resultados entre 0,001 (Q25) a 0,492 (Q46), com média de 0,143 (DP=0,11), ou seja, nenhum item apresentou probabilidade de acerto ao acaso acima do ideal ($c > 0,5$) (Andrade, Laros & Gouveia, 2010).

Tabela 49. Parâmetros individuais dos itens do teste Dimensão (N=1236)

Item	N	a	b	c
dim1	1236	1,496	-2,015	0,127
dim2	1216	2,094	-1,635	0,088
dim3	1229	1,542	-2,001	0,007
dim4	1215	2,100	-1,736	0,002
dim5	1211	2,562	-1,430	0,228
dim6	1231	2,652	-1,840	0,001
dim7	1211	2,139	-0,798	0,180
dim8	1213	2,481	-1,538	0,070
dim9	1212	2,710	-1,231	0,218
dim10	1215	1,641	-1,701	0,001
dim11	1202	1,947	-1,305	0,219
dim12	1197	2,429	-1,884	0,001
dim13	1165	2,318	-0,887	0,153
dim14	1179	2,470	-1,726	0,003
dim15	1157	1,929	-1,167	0,170
dim16	1151	4,935	-0,499	0,401
dim17	1128	2,054	-1,401	0,040
dim18	1080	1,602	-0,816	0,224
dim19	1108	2,674	-1,723	0,001
dim20	1070	1,657	-1,925	0,008
dim21	1021	4,280	0,635	0,288
dim22	1022	1,745	-0,549	0,201
dim23	1017	2,582	-1,606	0,001
dim24	952	1,555	-0,099	0,303
dim25	911	1,581	-1,031	0,001
dim26	905	5,151	1,332	0,121
dim27	893	1,657	-1,993	0,027
dim28	839	1,933	-1,577	0,098
dim29	788	2,111	0,088	0,216
dim30	776	2,288	-1,313	0,083
dim31	738	1,805	0,767	0,251
dim32	736	2,406	-1,508	0,226

dim33	663	9,735	0,887	0,276
dim34	654	1,978	-1,584	0,004
dim35	579	6,150	0,931	0,246
dim36	575	1,755	-0,383	0,170
dim37	528	6,343	0,713	0,147
dim38	444	1,960	0,437	0,122
dim39	428	5,647	0,607	0,132
dim40	367	7,505	0,904	0,106
dim41	324	4,071	0,880	0,177
dim42	294	1,416	-0,171	0,040
dim43	293	3,487	0,428	0,110
dim44	249	3,941	1,144	0,215
dim45	239	3,776	0,694	0,211
dim46	231	3,323	0,458	0,492
dim47	180	4,218	1,711	0,226
dim48	171	5,168	0,980	0,229

As Curvas Características dos Itens (CCIs) do teste Dimensão (Figura 33), ilustram itens com parâmetro de discriminação altos, principalmente os itens Q33, Q35, Q37 e Q40, que apresentaram resultados muito acima de 6,0, indicando discriminação praticamente perfeita, avaliada a partir dos ângulos próximos a 90°. Além disso, os itens com maior probabilidade de acerto ao acaso também são facilmente identificados, os itens Q16 e Q46, mesmo não apresentando índices alarmantes ($c=0,401$ e $0,492$, respectivamente).

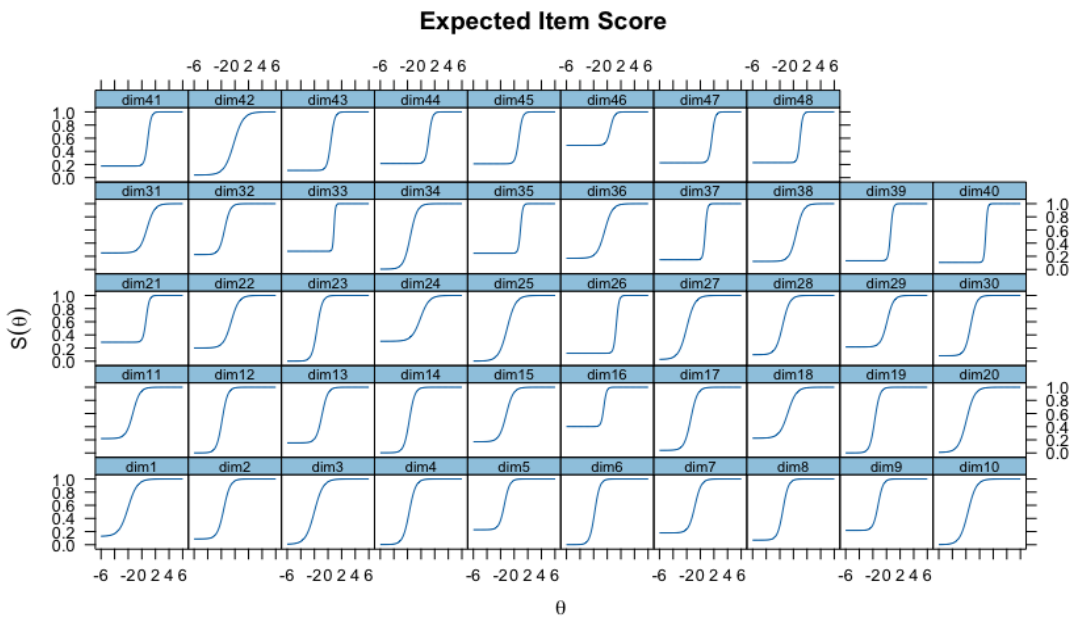


Figura 33. Curvas Características dos Itens (CCI's) do teste Dimensão

A partir da CIT (Figura 34) é possível observar que o teste Dimensão teve maior poder de informação para o valor de θ +0,9. Existe uma pequena melhoria na precisão do teste também para o θ -1,7. Nestas regiões a medida é mais precisa demonstrando um menor erro. Pelo exame do erro padrão, o teste demonstra maior precisão para pessoas que variam o θ entre -3 e +2, aproximadamente, indicando que o teste funciona melhor para sujeitos que têm um nível de θ baixo à médio.



Figura 34. Curva de Informação do Teste (CIT) e erro-padrão

5.7.2 Evidências de validade baseadas na estrutura interna do teste Dimensão

A Tabela 50 apresenta a investigação da estrutura interna conduzida por uma AFC. Observa-se que apenas os itens 26 e 47 apresentaram cargas fatoriais fracas. Esses mesmos itens apresentaram índice de discriminação irrealista na averiguação pela TRI. O resultado para os demais itens apontaram cargas fatoriais moderadas à fortes indicando que estes representam adequadamente o respectivo construto. Os índices de adequação recomendados foram examinados, revelando resultados satisfatórios na AFC para o modelo unidimensional avaliado, reforçando a validade da estrutura interna. (Hu & Bentler, 1998; Marsh, Hau, & Wen, 2004; Morin, Arens, & Marsh, 2016).

Tabela 50. Cargas fatoriais e índices de ajustes do teste Dimensão da Bateria TSP de acordo com a Análise Fatorial Confirmatória

Item	Cargas Fatoriais
1	0,66
2	0,78
3	0,67

4	0,79
5	0,78
6	0,89
7	0,69
8	0,83
9	0,80
10	0,71
11	0,71
12	0,85
13	0,74
14	0,84
15	0,68
16	0,71
17	0,76
18	0,61
19	0,87
20	0,71
21	0,52
22	0,61
23	0,88
24	0,51
25	0,70
26	0,30
27	0,72
28	0,78
29	0,59
30	0,81
31	0,49
32	0,81
33	0,54
34	0,82
35	0,53
36	0,65
37	0,75
38	0,69
39	0,82
40	0,67
41	0,57
42	0,71
43	0,76
44	0,40
45	0,60
46	0,45
47	0,22
48	0,47
Ajustes	Dimensão
RMSEA	0,03
GFI	0,95
TLI	0,96
CFI	0,96

5.7.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Dimensão

a) Evidência de validade convergente

Foram 79 sujeitos submetidos às provas de raciocínio da BPR-5. A maioria do sexo feminino (70,9%), a idade variou entre 18 e 43 anos ($M=21,39$; $DP = 4,82$), com ensino superior incompleto e completo. A análise de correlação de *Pearson* foi realizada, revelando evidências de validade convergente para o teste Dimensão. Observou-se uma correlação moderada e estatisticamente significativa com as provas de raciocínio abstrato, mecânico, espacial e o escore geral da BPR-5 (ver Tabela 51).

Foram submetidos 24 sujeitos ao teste Dimensão e ao G-38. Participaram 18 mulheres (75%), a idade variou entre 19 e 62 anos ($M=28,21$; $DP=11,69$), com ensino superior incompleto e completo. A correlação entre os resultados dos dois testes foi de magnitude moderada ($r=0,57$; $p=0,004$), o que sugere uma relação significativa entre eles e evidências de validade convergente para o teste Dimensão.

Tabela 51. Matriz de correlações entre o teste Dimensão, G-38 e provas do BPR-5

Testes	<i>Pearson's r</i>	p
Dimensão Total - BPR-5: RA	0,45	< 0,001
Dimensão Total - BPR-5:RM	0,47	< 0,001
Dimensão Total - BPR-5:RE	0,50	< 0,001
Dimensão Total - BPR-5:EG	0,58	< 0,001
Dimensão Total - Teste G-38	0,57	0,004

b) Evidências de validade de critério (variáveis sociodemográficas)

Com relação à variável sexo, a Tabela 52, apresenta o resultado do desempenho da amostra no teste Dimensão, com resultado da diferença estatisticamente significativo, no entanto, pequeno tamanho do efeito ($d=-0,30$). Deste modo, a grande sobreposição da distribuição dos desempenhos das mulheres e dos homens no teste apontou que não há um efeito prático para o resultado no teste. O resultado maior no desempenho dos homens, frequentemente observado neste teste, é comumente notado em avaliações que requerem habilidades espaciais, como no caso do teste Dimensão (Flores-Mendoza, 2000; Primi et al.,

2022). No entanto, o tamanho do efeito foi menor ao ponto de corte adotado. Acredita-se que o baixo tamanho de efeito observado possa estar relacionado ao nível de dificuldade predominantemente baixo dos itens no teste Dimensão, conforme evidenciado pelas análises da TRI. Como os itens não requerem habilidades muito avançadas, o instrumento pode não ser sensível as diferenças entre grupos.

Tabela 52. Desempenho dos participantes no teste Dimensão de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	Teste <i>t</i>	Valor <i>p</i>	<i>d</i> de Cohen
Dimensão	Feminino	920	23,34	10,17	-5,36	<0,001	-0,30
	Masculino	465	26,49	10,61			

Os resultados das Tabela 53, indicaram a presença de diferenças estatisticamente significativas entre os grupos etários, com tamanho do efeito forte ($\eta^2=0,17$). As comparações *post hoc* revelaram diferenças estatisticamente significativas, com tamanhos de efeito de médio a grande, entre os desempenhos médios do grupo com idade ≥ 40 anos e todas as outras faixas etárias avaliadas. Similarmente, o grupo etário de 19-25 anos exibiu diferenças significativas e tamanho de efeito moderado ($d=0,59$) em comparação com as médias do grupo de 26 a 39 anos.

Tabela 53. Desempenho dos participantes no teste Dimensão segundo a faixa etária

Teste	Faixa etária	N	Média	DP	F	Valor <i>p</i>	η^2
Dimensão	15-16	88	24,98	9,83	70,33	<0,001	0,17
	17-18	265	26,85	9,33			
	19-25	544	27,76	9,34			
	26-39	252	22,17	10,29			
	≥ 40 anos	235	16,03	9,14			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

A variável escolaridade (Tabela 54) apresentou diferenças nas médias de desempenho dos grupos estatisticamente significativas e com tamanho de efeito forte ($\eta^2=0,12$). As comparações *post hoc* apontaram que as médias dos desempenhos dos participantes com ensino fundamental diferiram daqueles com ensino médio e superior, com grandes resultados para os tamanhos de efeito. As comparações entre os grupos do ensino médio e superior não foram estatisticamente significativas ($p=0,33$).

Tabela 54. Desempenho dos participantes no teste Dimensão segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor <i>p</i>	η^2
Dimensão	Fundamental	258	16,89	9,64	94,36	<0,001	0,12
	Médio	391	25,54	9,71			
	Superior	736	26,42	9,87			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

5.7.4 Evidências de confiabilidade do teste Dimensão

Os resultados dos coeficientes alfa de Cronbach ($\alpha=0,93$) e ômega de McDonald ($\Omega=0,94$) evidenciaram adequada consistência interna para o teste Dimensão. A evidência de estabilidade temporal demonstrou coeficiente de correlação de *Pearson* de magnitude forte ($r=0,82$) e estatisticamente significativo, demonstrando também uma estabilidade temporal satisfatória (Tabela 55).

Tabela 55. Estatísticas do estudo de estabilidade temporal do teste Dimensão (n=95)

Teste	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	r_{12}	Valor <i>p</i>
Dimensão	32,66	8,96	35,59	7,86	0,82	<0,001

Nota. M: Média, DP: Desvio-Padrão, r_{12} : correlação bivariada, valor significativo se $p < 0,05$.

De modo geral, os resultados dos estudos das propriedades psicométricas do teste Dimensão indicaram uma consistência interna adequada, conforme observado na AFC. Os resultados da TRI revelaram que todos os itens do teste estavam dentro dos padrões ideais para o índice de chute. Em relação à dificuldade, os níveis variaram entre fáceis e médios, o que é promissor. Isso contrasta com os demais testes, que são predominantemente compostos por itens fáceis. No entanto, a análise de discriminação apontou a necessidade de novos estudos para compreender os índices irrealistas observados, especialmente para os dois itens que apresentaram baixa carga fatorial na AFC (Q26 e Q47).

As evidências de validade convergente foram satisfatórias, revelando correlações de magnitude moderada entre o teste e as provas de raciocínio da BPR-5, bem como com os fatores gerais. Esses resultados indicam a eficácia do teste em avaliar habilidades espaciais e sua forte relação com a inteligência fluida (Gomes & Borges, 2009b). Além disso, a análise da validade de critério também corroborou com os achados da literatura, demonstrando a habilidade do teste em diferenciar grupos com base nas características sociodemográficas,

idade e escolaridade (Kaufman et al., 2009; Primi et al., 2022). Os estudos de confiabilidade do teste mostraram resultados robustos, em conformidade com as diretrizes atuais (CFP, 2022), indicando uma sólida fundamentação teórica e métrica para avaliar o raciocínio espacial em indivíduos com 15 anos ou mais, independentemente de seu nível educacional.

5.8 Teste Blocos

Participaram desse estudo 1398 sujeitos, a amostragem foi por conveniência, residentes nos estados de Minas Gerais (N=808), Distrito Federal (N=301), Espírito Santo (N=152) e Ceará (N=137). Foram 920 mulheres (66,4%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=26,66; DP=11,55). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=740), seguido pelo Ensino Médio completo ou incompleto (N=403) e Ensino Fundamental completo ou incompleto (N=255).

Para avaliação da normalidade dos dados observa-se (Tabela 56) que o teste estatístico Shapiro-Wilk ($W=0,994$; $p<0,001$) demonstra pelo parâmetro de significância que esta amostra é diferente de uma distribuição normal, no entanto, nota-se um valor W muito próximo a 1, o qual pode indicar uma distribuição aproximadamente normal. Verifica-se que o resultado apresenta uma assimetria negativa muito pequena (-0,016), o índice próximo a zero é um indicativo da normalidade dos dados (Henderson,2006; Miot, 2017; Vetter, 2017).

Tabela 56. Estatísticas descritivas do teste Blocos

Estatística descritiva	Teste Blocos
N	1.398
Média	13,26
Mediana	13,0
Moda	15,0
Desvio-padrão	5,13
Amplitude	29
Mínimo	0
Máximo	29
Assimetria	-0,016
Erro-padrão da Assimetria	0,065
Curtose	-0,348
Erro-padrão da Curtose	0,131
W de Shapiro-Wilk	0,994
p Shapiro-Wilk	< 0,001

O diagrama Q-Q (Figura 35) aponta que os dados do teste Blocos se aproximam de uma distribuição normal. De maneira complementar foi gerado um histograma de frequência (Figura 36), a inspeção visual dos dados confirma os achados de uma distribuição aproximadamente normal, com uma inclinação da cauda muito leve para a esquerda (Dancey & Reidy, 2019).

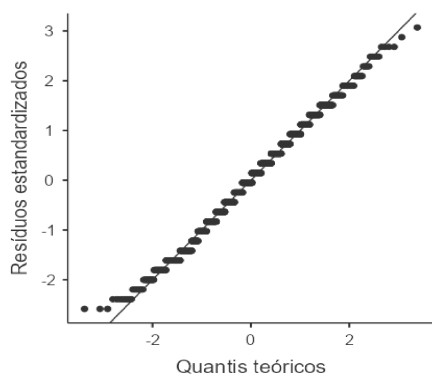


Figura 35. Diagrama Q-Q teste Blocos

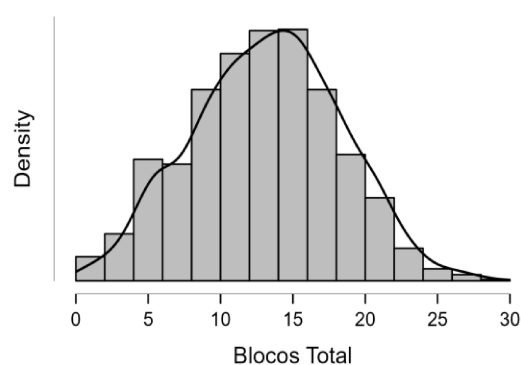


Figura 36. Curva de Gauss do teste Blocos

5.8.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Blocos

A Tabela 57 apresenta os resultados para os parâmetros individuais dos itens. Verifica-se que os itens apresentaram discriminação entre 0,347 (Q31) e 4,013 (Q25), com média de 1,548 (DP= 0,761). Observou-se que os itens Q5 ($a=0,480$), Q31 ($a=0,347$) e Q32 ($a=0,371$) apresentaram parâmetros de discriminação insuficientes ($a < 0,65$). O item Q25 ($a=4,013$) apresentou discriminação irrealista ($a > 3,0$; Barbetta et al., 2014).

Quando avaliada a dificuldade do teste, os resultados variaram entre -4,366 (Q5) a 4,242 (Q31), com dificuldade média de -0,173 (DP= 1,910). Nota-se que os itens com índices de dificuldade altos (Q5 e Q31), demonstraram também índices de discriminação baixos. Já os resultados do parâmetro de chute apontaram variabilidade entre 0,000 a 0,480 (Q1), indicando que não houve itens com probabilidade de resposta ao acaso acima do ideal para itens dicotômicos ($c > 0,5$; Andrade, Laros & Gouveia, 2010).

Tabela 57. Parâmetros individuais dos itens do teste Blocos (N=1276)

Item	N	a	b	c
bl1	1275	2,885	-2,788	0,480
bl2	1273	1,448	-1,758	0,000
bl3	1276	0,951	-3,162	0,001
bl4	1273	1,188	-3,110	0,002

bl5	1254	0,480	-4,366	0,005
bl6	1246	0,805	-2,350	0,001
bl7	1263	2,459	-0,952	0,000
bl8	1264	2,091	-1,028	0,017
bl9	1225	1,163	-0,962	0,000
bl10	1220	1,455	-1,061	0,000
bl11	1236	2,406	-1,317	0,000
bl12	1170	1,784	-0,206	0,000
bl13	1136	2,051	-0,272	0,000
bl14	1153	0,988	0,291	0,000
bl15	1040	1,826	-0,405	0,000
bl16	1010	1,330	0,505	0,106
bl17	957	1,916	-0,027	0,000
bl18	623	1,774	0,524	0,000
bl19	696	1,661	0,061	0,000
bl20	571	0,942	-0,271	0,013
bl21	543	1,351	-0,118	0,000
bl22	450	1,570	0,616	0,000
bl23	374	1,547	-0,225	0,000
bl24	306	0,889	1,353	0,005
bl25	212	4,013	0,870	0,364
bl26	200	1,347	0,909	0,277
bl27	183	1,401	-0,279	0,001
bl28	132	1,865	0,742	0,000
bl29	90	2,358	1,829	0,068
bl30	76	0,881	3,059	0,000
bl31	73	0,347	4,424	0,002
bl32	72	0,371	3,935	0,001

As CCI's do teste Blocos (Figura 37), ilustram a baixa capacidade preditiva dos itens Q5, Q31 e Q32, nota-se que a medida que a discriminação diminui a curva fica menos perceptível. Enquanto o item com curva mais próxima do ângulo reto foi o item Q25 ($a=4,013$), com discriminação muito alta. O item com maior probabilidade de acerto ao acaso foi o item Q1 ($c=0,480$), que apresentou resultado no limiar do ideal para itens dicotômicos. Embora o item Q1 possa ter sido mais suscetível a respostas aleatórias, ele ainda contribuiu efetivamente para a avaliação das habilidades dos participantes devido aos seus bons índices de discriminação e dificuldade adequada.

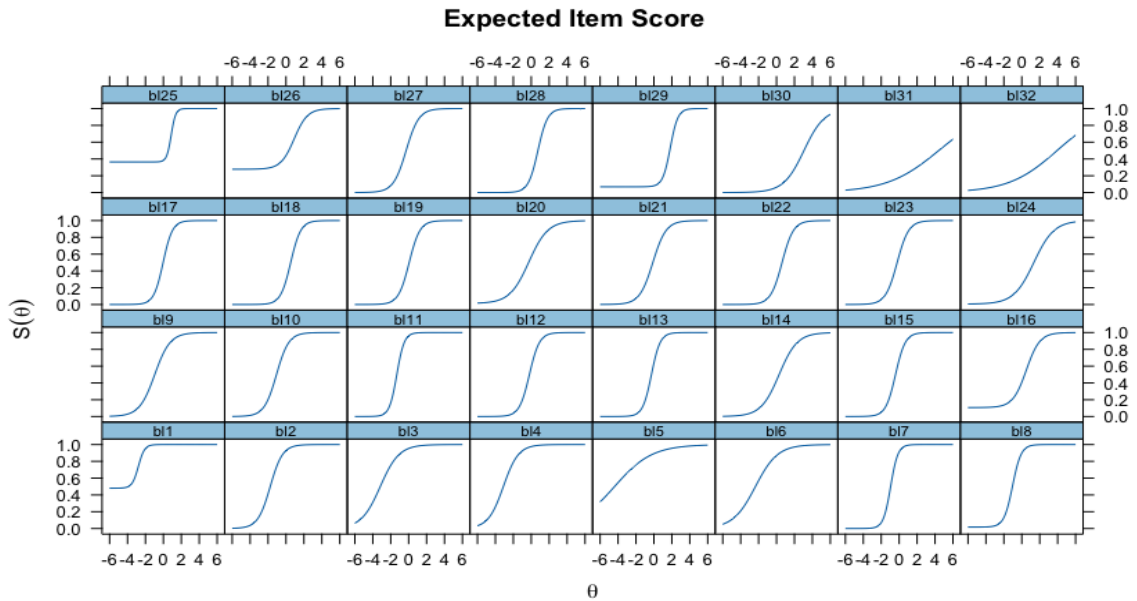


Figura 37. Curvas Características dos Itens (CCI's) do teste Blocos

A partir da observação da CIT (Figura 38), o teste Blocos teve maior poder de informação para o valor de θ em torno de -0,5, aproximadamente. Nesta região a medida é mais precisa demonstrando um menor erro. Pelo exame do erro padrão, o teste demonstra maior precisão para pessoas que variam o θ entre -3 e +2,7, aproximadamente, indicando que o teste funciona de maneira adequada para sujeitos com diferentes níveis de θ .

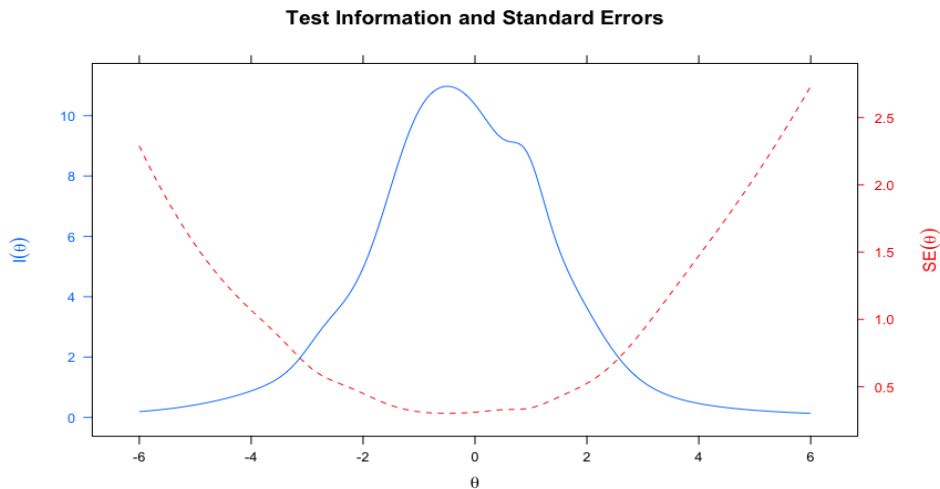


Figura 38. Curva de Informação do Teste (CIT) e erro-padrão

5.8.2 Evidências de validade baseadas na estrutura interna do teste Blocos

A estrutura interna do teste Blocos foi examinada por meio de uma AFC, cujos resultados estão apresentados na Tabela 58. Identificou-se que seis itens apresentaram cargas

fatoriais consideradas fracas, sendo o item Q5, Q26 e os itens entre 29 e 32. Ao comparar os resultados da AFC com os obtidos pela TRI, observou-se que, embora os itens Q26, Q29 e Q30, tenham apresentado baixas cargas fatoriais (entre 0,33 e 0,38), eles apresentaram bons índices pela TRI. Por outro lado, os itens Q5, Q31 e Q32 também revelaram resultados insatisfatórios em relação ao índice de discriminação (a) e ao parâmetro dificuldade (b), com o item Q5 indicando um alto grau de facilidade, enquanto os itens Q31 e Q32 apresentaram um alto índice de dificuldade. É importante ressaltar que a partir do item Q29, o tamanho amostral foi inferior a 100, o que requer uma avaliação cuidadosa dos resultados sendo adequados ou não.

O resultado para a maior parte das cargas fatoriais acima de 0,40 reforçou que os itens representam adequadamente o respectivo construto. Os índices de ajuste preconizados na literatura foram inspecionados, os resultados da AFC foram satisfatórios para o modelo unidimensional investigado, o que evidencia a validade da estrutura interna (Hu & Bentler, 1998; Marsh, Hau, & Wen, 2004; Morin, Arens, & Marsh, 2016).

Tabela 58. Cargas fatoriais e índices de ajustes do teste Blocos da Bateria TSP de acordo com a Análise Fatorial Confirmatória

Item	Cargas Fatoriais
1	0,79
2	0,68
3	0,52
4	0,58
5	0,26
6	0,45
7	0,85
8	0,79
9	0,60
10	0,68
11	0,85
12	0,71
13	0,74
14	0,51
15	0,73
16	0,50
17	0,73
18	0,72
19	0,69
20	0,52
21	0,64
22	0,66
23	0,71
24	0,44

25	0,42
26	0,38
27	0,69
28	0,75
29	0,37
30	0,33
31	0,16
32	0,20
Ajustes	Blocos
RMSEA	0,01
GFI	0,98
TLI	0,98
CFI	0,98

5.8.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Blocos

a) Evidência de validade convergente

Foram 79 sujeitos submetidos às provas de raciocínio da BPR-5. A maioria do sexo feminino (70,9%), a idade variou entre 18 e 43 anos ($M=21,39$; $DP=4,82$), com ensino superior incompleto e completo. O estudo foi feito com as provas da BPR-5 que incluem processamento de imagem (RA, RE, RM), e uma análise adicional foi realizada com a prova RN. A capacidade de estimar quantidades pode ser importante para o teste Blocos, já que os respondentes precisam avaliar o número de blocos dentro de um conjunto para resolver o problema posto. Essa abordagem é interessante porque permite explorar como diferentes aspectos cognitivos podem estar relacionados entre si e como eles contribuem para habilidades específicas, como processamento de imagem e resolução de problemas espaciais.

Os resultados para a correlação de *Pearson* apontaram evidências de validade convergente para o teste Blocos com correlação moderada e estatisticamente significativa para a prova de raciocínio abstrato, mecânico, espacial, numérico e escore geral da BPR-5 (Tabela 59). Foram submetidos 24 sujeitos ao teste Blocos e ao G-38. Participaram 18 mulheres (75%), a idade variou entre 19 e 62 anos ($M=28,21$; $DP=11,69$), com ensino superior incompleto e completo. Os resultados analisados apontaram evidências de validade convergente para o teste Blocos com correlação moderada e estatisticamente significativa ($r=0,64$; $p<0,001$).

Tabela 59. Matriz de correlações entre o teste Blocos, G-38 e provas de raciocínio (BPR-5)

Testes	<i>Pearson's r</i>	<i>p</i>
Blocos Total - BPR-5:RA	0,49	< 0,001
Blocos Total - BPR-5:RM	0,46	< 0,001
Blocos Total - BPR-5:RE	0,45	< 0,001
Blocos Total - BPR-5:RN	0,42	< 0,001
Blocos Total - BPR-5:EG	0,56	< 0,001
Blocos Total - Teste G-38	0,64	< 0,001

b) Evidências de validade de critério (variáveis sociodemográficas)

A análise da variável sexo, conforme mostrado na Tabela 60, revelou diferenças estatisticamente significativas nas médias de desempenho entre homens e mulheres no teste Blocos ($p < 0,001$). No entanto, o tamanho do efeito foi ligeiramente menor do que o limiar estabelecido ($d = -0,47$). Nesse contexto, sugere-se a consideração de pesquisas anteriores para contextualizar mais amplamente o significado desse resultado (Durlak, 2009). Deste modo, de acordo com os dados da pesquisa que embasou o último manual da Bateria TSP, foi observado um tamanho de efeito de $d = 0,55$, indicando um forte efeito. Na pesquisa em questão, os participantes do sexo masculino apresentaram desempenho superior aos do sexo feminino (Nascimento & Carmen-Flores, 2007). Portanto, torna-se possível, na presente pesquisa, validar os efeitos práticos das diferenças de desempenho por sexo no teste Blocos.

Tabela 60. Desempenho dos participantes no teste Blocos de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	Teste <i>t</i>	Valor <i>p</i>	<i>d</i> de Cohen
Blocos	Feminino	927	12,47	5,04	-8,27	<0,001	-0,47
	Masculino	471	14,81	4,93			

A Tabela 61 indica que houve uma diferenças estatisticamente significativas entre os grupos etários. Considerando o tamanho do efeito, as diferenças entre as médias foram fortes ($\eta^2 = 0,11$). As comparações *post hoc* revelaram diferenças estatisticamente significativas, com tamanhos de efeito superiores ao ponto de corte adotado, apenas entre os desempenhos médios do grupo etário ≥ 40 anos e todos os outros grupos avaliados.

Tabela 61. Desempenho dos participantes no teste Blocos segundo a faixa etária

Teste	Faixa etária	N	Média	DP	F	Valor p	η^2
Blocos	15-16	85	13,07	4,49	41,99	<0,001	0,11
	17-18	273	14,77	4,95			
	19-25	548	14,23	5,07			
	26-39	256	12,78	4,69			
	≥40 anos	235	9,84	4,48			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor p = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

As diferenças nas médias de desempenho dos grupos em função da escolaridade foram estatisticamente significativas e com tamanho de efeito moderado ($\eta^2=0,09$) (Tabela 62). As comparações *post hoc* indicaram que as médias de desempenho dos participantes com ensino fundamental diferiram significativamente daqueles com ensino médio e superior, apresentando grandes tamanhos de efeito. As comparações entre os grupos do ensino médio e superior não foram estatisticamente significativas ($p=0,55$).

Tabela 62. Desempenho dos participantes no teste Blocos segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor p	η^2
Blocos	Fundamental	255	9,96	4,41	71,94	<0,001	0,09
	Médio	403	14,20	4,98			
	Superior	740	13,89	4,99			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA; Valor p = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

5.8.4 Evidências de confiabilidade do teste Blocos

Evidências de confiabilidade foram obtidas pelos coeficientes de consistência interna alfa de Cronbach ($\alpha=0,91$) e ômega de McDonald ($\Omega=0,941$), os resultados apresentam adequada consistência interna. A evidência de estabilidade temporal demonstrou coeficiente de correlação de *Pearson* de magnitude forte ($r=0,74$) e estatisticamente significativo ($p<0,001$), demonstrando também uma estabilidade temporal satisfatória (Tabela 63).

Tabela 63. Estatísticas do estudo de estabilidade temporal do teste Blocos (N=92)

Teste	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	r_{12}	Valor p
Blocos	16,85	4,13	19,24	4,24	0,74	<0,001

Nota. M: Média, DP: Desvio-Padrão, r_{12} : correlação bivariada, valor significativo se $p < 0,05$.

De maneira geral, os resultados das análises psicométricas do teste Blocos demonstraram uma consistência interna satisfatória, conforme evidenciado pela AFC. Os resultados da TRI indicaram que todos os itens do teste estavam em conformidade com os padrões ideais para o índice de chute. Quanto à dificuldade dos itens, observou-se uma heterogeneidade, variando de fácil a difícil, o que é considerado benéfico para testes que visam avaliar as capacidades cognitivas. No entanto, foram identificados problemas significativos, especialmente relacionados à capacidade de discriminação e às cargas fatoriais dos itens Q5, Q31 e Q32, apontando a necessidade de revisão desses itens.

As evidências de validade convergente foram satisfatórias, revelando correlações de magnitude moderada entre o teste e as provas de raciocínio da BPR-5, bem como com os fatores gerais. Esses resultados indicam a eficácia do teste em avaliar habilidades espaciais e sua forte relação com a inteligência fluida (Gomes & Borges, 2009b). Além disso, a análise da validade de critério também corroborou com os achados da literatura, demonstrando a habilidade do teste em diferenciar grupos com base nas características sociodemográficas (Kaufman et al., 2009; Primi et al., 2022). Os estudos de confiabilidade do teste mostraram resultados robustos, em conformidade com as diretrizes atuais (CFP, 2022), indicando uma sólida fundamentação teórica e métrica para avaliar o raciocínio espacial em indivíduos com 15 anos ou mais, independentemente de seu nível educacional.

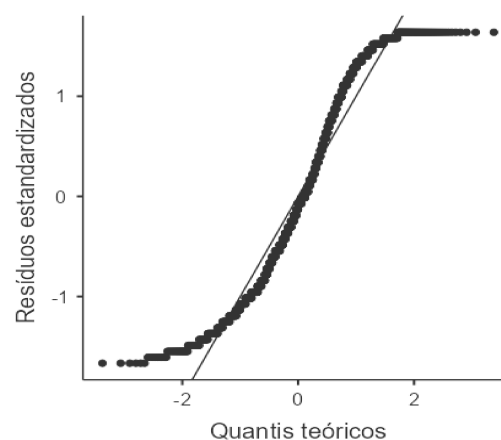
5.9 Teste Memória

Participaram desse estudo 1368 sujeitos, a amostragem foi por conveniência, residentes nos estados de Minas Gerais (N=801), Distrito Federal (N=292), Espírito Santo (N=152) e Ceará (N=123). Foram 901 mulheres (65,9%), a idade mínima foi a estabelecida de 15 anos e a máxima 76 anos (M=26,62; DP=11,52). Sobre a escolaridade a maioria foram sujeitos com Ensino superior completo ou incompleto (N=726), seguido pelo Ensino Médio completo ou incompleto (N=395) e Ensino Fundamental completo ou incompleto (N=247).

Para avaliação da distribuição dos dados observa-se na Tabela 64 que o teste estatístico Shapiro-Wilk demonstra pelo valor de W e parâmetro de significância que esta amostra é diferente de uma distribuição normal ($W=0,937$; $p<0,001$). Verifica-se também que o resultado apresenta uma pequena assimetria positiva (0,185) e a curtose sinaliza distribuição platicúrtica, muito achatada (-1,255). Os dados avaliados no diagrama Q-Q (Figura 39) confirmam que a amostra não segue uma distribuição normal pois foi observada uma distribuição em curva S (Henderson, 2006; Miot, 2017; Vetter, 2017).

Tabela 64. Estatísticas descritivas do teste Memória

Estatística descritiva	Teste Memória
N	1.368
Média	28,21
Mediana	27,0
Moda	56,0
Desvio-padrão	16,95
Amplitude	56
Mínimo	0
Máximo	56
Assimetria	0,185
Erro-padrão da Assimetria	0,066
Curtose	-1,255
Erro-padrão da Curtose	0,132
W de Shapiro-Wilk	0,937
p Shapiro-Wilk	< 0,001

**Figura 39.** Diagrama Q-Q teste Memória

5.9.1 Análise pela Teoria de Resposta ao Item (TRI) do teste Memória

Utilizou-se da TRI para avaliar os parâmetros individuais dos itens do teste de Memória (Tabela 65). Os resultados apontaram que a discriminação dos itens variou entre 0,940 (Q6) e 4,175 (Q36), entre moderado e muito alto, com média de discriminação de 2,374 (DP=0,754). Já os parâmetros de dificuldade variaram entre -2,352 (Q3) e +0,668 (Q25), com média de -0,983 (DP=0,963) indicando que a maior parte dos itens apresentou dificuldade baixa ou moderada. Quanto a probabilidade de acerto ao acaso, os resultados apontaram que apenas os itens Q5 (c= 0,720), Q25 (c= 0,590), Q32 (c= 0,512), Q33 (c= 0,506) e Q36 (c= 0,716) apresentaram resultados alarmantes, com coeficientes acima de 0,50.

Tabela 65. Parâmetros individuais dos itens do teste de Memória (N=1144)

Item	N	a	b	c
mem1	1038	1,723	-1,920	0,001
mem2	1027	3,441	-1,542	0,080
mem3	1144	1,010	-2,352	0,072
mem4	1035	2,599	-1,938	0,001
mem5	1115	1,841	-0,314	0,720
mem6	872	0,940	-0,726	0,004
mem7	972	2,128	-0,143	0,439
mem8	1014	1,556	-0,751	0,428
mem9	920	1,307	-1,178	0,116

mem10	979	2,911	-1,820	0,192
mem11	1041	2,587	-0,107	0,494
mem12	897	2,980	-1,672	0,000
mem13	985	3,711	-0,080	0,428
mem14	804	1,797	-1,394	0,000
mem15	799	2,895	0,471	0,389
mem16	796	2,201	-1,843	0,001
mem17	750	2,555	-1,583	0,001
mem18	845	2,024	0,027	0,253
mem19	714	1,733	0,121	0,195
mem20	744	2,749	-2,085	0,001
mem21	672	3,875	-1,444	0,001
mem22	790	2,318	-0,018	0,312
mem23	621	2,605	-1,586	0,191
mem24	560	1,849	-1,269	0,000
mem25	616	2,837	0,668	0,590
mem26	648	2,387	0,048	0,442
mem27	564	2,111	-1,838	0,001
mem28	589	1,798	-2,350	0,091
mem29	564	2,525	0,335	0,361
mem30	494	2,099	-1,902	0,001
mem31	455	1,362	-2,176	0,012
mem32	574	2,431	0,065	0,512
mem33	575	2,804	0,010	0,506
mem34	472	2,431	-1,844	0,002
mem35	460	3,160	-1,751	0,002
mem36	574	4,175	0,495	0,716

A análise das CCIs ilustrou os parâmetros individuais dos itens (Figura 40). Os itens com maiores índices de discriminação apresentam angulação da curva próximas ao ângulo reto (90°), enquanto itens menos discriminação apresentam curvas mais suaves. Já os itens com maior probabilidade de acerto ao acaso apresentam o início das curvas mais altos, sendo aqueles com valores do eixo y acima de 0,50 indicando alta probabilidade de acerto ao acaso (Q5, Q25, Q32, Q33 e Q36).

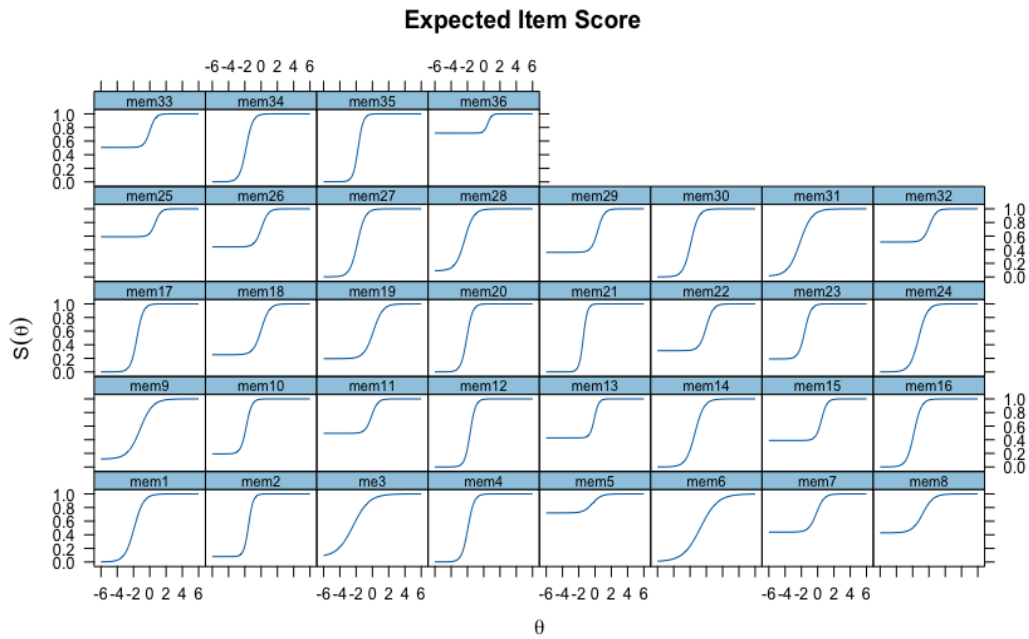


Figura 40. Curvas Características dos Itens (CCI's) do teste Memória

A partir da observação da CIT (Figura 41) nota-se que os itens do teste Memória tiveram maior poder de informação para valores de θ em torno de -1,3, ou seja, nesta região a medida demonstra um menor erro. Existe uma certa melhoria na precisão do teste também para o θ de -3,2, pois há dois picos, um mais acentuado que o outra, mas ambos dentro do intervalo negativo. Pelo exame do erro padrão, o teste demonstra maior precisão para pessoas que variam o θ entre -4 e +1, aproximadamente, indicando que o teste funciona melhor para sujeitos que têm um nível de θ baixo, principalmente.

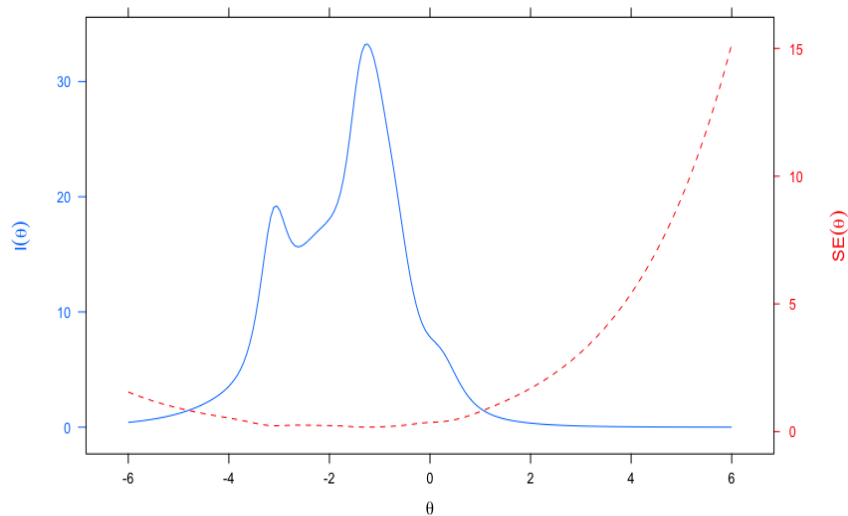


Figura 41. Curva de Informação do Teste (CIT) e erro-padrão

5.9.1 Evidências de validade baseadas na estrutura interna do teste Memória

A análise da estrutura interna foi realizada por meio de uma AFC, na qual todas as cargas fatoriais foram superiores a 0,40, variando de moderadas a fortes (Tabela 66). Esse resultado ressalta a capacidade dos itens em representar de forma apropriada o construto em questão. Além disso, os índices de ajuste obtidos pela AFC foram considerados adequados, confirmando o modelo unidimensional investigado e fortalecendo a validade da estrutura interna (Hu & Bentler, 1998; Marsh, Hau, & Wen, 2004; Morin, Arens, & Marsh, 2016).

Tabela 66. Cargas fatoriais e índices de ajustes do teste Memória da Bateria TSP de acordo com a Análise Fatorial Confirmatória

Item	Cargas Fatoriais
1	0,69
2	0,88
3	0,49
4	0,84
5	0,40
6	0,46
7	0,51
8	0,55
9	0,51
10	0,83
11	0,52
12	0,83
13	0,60
14	0,71
15	0,49
16	0,78
17	0,79
18	0,56
19	0,54
20	0,92
21	0,89
22	0,60
23	0,79
24	0,73
25	0,42
26	0,61
27	0,74
28	0,82

29	0,63
30	0,94
31	0,73
32	0,54
33	0,56
34	0,91
35	0,92
36	0,44
Ajustes	Memória
RMSEA	0,02
GFI	0,97
TLI	0,91
CFI	0,92

5.9.3 Evidências de validade baseadas nas relações com variáveis externas para o teste Memória

a) Evidências de validade de critério (variáveis sociodemográficas)

Com relação à variável sexo (Tabela 67), as diferenças entre as médias de desempenho de homens e mulheres não foram estatisticamente significativas no teste de Memória ($U=199373.000$; $p=0,112$). O tamanho do efeito, inferior ao ponto de corte adotado ($d=0,05$), confirma a grande sobreposição da distribuição dos desempenhos no teste.

Tabela 67. Desempenho dos participantes no teste Memória de acordo com o sexo

Teste	Sexo	N	Média	Desvio-padrão	U	Valor p	d de Cohen
Memória	Feminino	901	28,79	17,14	199.373.000	<0,112	0,05
	Masculino	467	27,09	16,53			

Nota. O tamanho do efeito é dado pela correlação bisserial de classificação

Os resultados das análises indicaram a presença de diferenças estatisticamente significativas entre os grupos etários (Tabela 68). Considerando o tamanho do efeito, as diferenças entre as médias foram fortes ($\eta^2 = 0,14$). As comparações *post hoc* indicaram diferenças significativas e com tamanhos de efeito superiores ao ponto de corte, entre os desempenhos médios do grupo etário de 19-25 anos com todos os demais. Dados semelhantes de diferença estatística e de tamanho de efeito foram encontrados para o grupo ≥ 40 anos em comparação com os grupos de 17-18 anos e 26-39 anos.

Tabela 68. Desempenho dos participantes no teste Memória segundo a faixa etária

Teste	Faixa etária	N	Média	DP	F	Valor <i>p</i>	η^2
Memória	15-16	85	24,83	14,15	59,55	<0,001	0,14
	17-18	267	26,02	14,84			
	19-25	535	35,41	16,28			
	26-39	253	25,81	17,08			
	≥40 anos	227	17,82	14,23			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA com *Bootstrapping*; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

As diferenças nas médias de desempenho em função da escolaridade foram estatisticamente significativas e com tamanho de efeito forte ($\eta^2 = 0,15$) (Tabela 69). As comparações *post hoc* indicaram que as médias dos desempenhos dos participantes foram estatisticamente significativas ($p < 0,001$) para os três níveis de ensino. No entanto, para os grupos fundamental e médio, a magnitude de efeito entre eles foi baixa ($d = -0,47$), indicando uma grande sobreposição dos desempenhos no teste de Memória entre eles e pouco impacto prático. Os participantes do ensino superior apresentaram tamanhos de efeito moderado e grande na comparação das médias com os grupos de ensino fundamental e médio, respectivamente.

Tabela 69. Desempenho dos participantes no teste Memória segundo o nível de escolaridade

Teste	Escolaridade	N	Média	DP	F	Valor <i>p</i>	η^2
Memória	Fundamental	247	17,10	13,26	139,06	<0,001	0,15
	Médio	395	24,44	14,86			
	Superior	726	34,04	16,68			

Nota. DP: Desvio padrão; F = resultado das diferenças entre grupos ANOVA com *Bootstrapping*; Valor *p* = significativas caso $p < 0,05$; η^2 = tamanho do efeito.

5.9.4 Evidências de confiabilidade do teste Memória

Evidências de confiabilidade foram obtidas pelos coeficientes de consistência interna alfa de Cronbach ($\alpha = 0,89$) e ômega de McDonald ($\Omega = 0,92$), os resultados apresentam adequada consistência interna. A evidência de estabilidade temporal (Tabela 70) demonstrou coeficiente de correlação Rô de Spearman de magnitude moderada ($r = 0,60$) e estatisticamente significativo, demonstrando também uma estabilidade temporal satisfatória.

Tabela 70. Estatísticas do estudo de estabilidade temporal do teste Memória (N=86)

Teste	1ª aplicação		2ª aplicação		Correlação	
	M	DP	M	DP	Rô	Valor <i>p</i>
Memória	48,36	9,51	51,69	7,40	0,60	<0,001

Nota. M: Média, DP: Desvio-Padrão, Rô de Spearman: correlação; valor significativo se $p < 0,05$.

Em suma, a pesquisa sobre a validade interna por meio da AFC revelou um bom ajuste ao modelo unidimensional e cargas fatoriais com magnitudes adequadas para todos os itens do teste Memória. No entanto, pela análise da TRI, os itens Q5, Q25, Q32, Q33 e Q 36 apresentaram valores inadequados para o parâmetro de acerto ao acaso, o que indica a necessidade de novos estudos para revisão. Em geral, os achados da TRI sugerem que os itens do teste de Memória são majoritariamente classificados como de baixa dificuldade, o que também indica que o teste pode não ser adequado para avaliar sujeitos com habilidades mais elevadas (Peres, Lessa, Valentini, 2019).

A análise das diferenças nas médias entre grupos com base em variáveis sociodemográficas proporcionou evidências robustas para a validade do teste, uma vez que os resultados estão alinhados com o que é observado na literatura (Burgaleta et al., 2012; Kaufman et al., 2009; Primi et al., 2022). As análises de confiabilidade, empregando índices de consistência interna e estabilidade temporal, revelaram resultados satisfatórios, em concordância com as diretrizes normativas (CFP, 2022).

5.10 Discussão geral dos resultados da Bateria TSP

A Tabela 71 apresenta a matriz de correlações entre os testes. Conforme observado em estudos anteriores, a maioria dos testes exibiu associações mais fortes com o teste de Julgamento (Johnson, te Nijenhuis, & Bouchard Jr, 2008; Nascimento & Carmen-flores, 2007). Os estudos com a BPR-5 têm constantemente revelado correlações entre as diferentes provas de raciocínio (Almeida et al., 2010; Primi & Almeida, 2000). Então, apesar das diferenças nas habilidades avaliadas, há características compartilhadas entre as provas. Esse padrão também foi observado nas correlações entre os testes da Bateria TSP, as quais variaram entre 0,35 e 0,69. Os coeficientes de correlação indicaram que a validade divergente não foi fortemente estabelecida, sugerindo que a natureza dos estímulos dos testes desempenha um papel significativo.

A hipótese relacionada à influência da natureza do estímulo (verbal e não verbal) sobre as correlações entre os testes foi confirmada. Os testes com predominância de estímulo não verbal, as correlações foram mais altas entre si, variando de 0,46 e 0,61. Nota-se que as correlações de menores magnitudes foram observadas entre os testes verbal e não verbal. Resultado semelhante ao estudo original da Bateria (King, 1957) e à pesquisa Holandesa realizada com a Marinha Real (Johnson, te Nijenhuis, & Bouchard Jr, 2008).

Os testes de habilidade espacial (Partes, Dimensão e Blocos) mostraram correlações moderadas entre si, variando de 0,56 a 0,61. Da mesma forma, os testes de habilidade perceptual (Percepção e Precisão) apresentaram uma correlação moderada ($r=0,52$).

Tabela 71. Matriz de correlações entre os testes da Bateria TSP.

Testes	Num	Per	Flu	Mem	Par	Prec	Dim	Jul	Blo
Números	1,00								
Percepção	0,56	1,00							
Fluência	0,50	0,52	1,00						
Memória	0,47	0,50	0,42	1,00					
Partes	0,43	0,45	0,37	0,39	1,00				
Precisão	0,38	0,52	0,37	0,41	0,60	1,00			
Dimensão	0,48	0,50	0,41	0,48	0,61	0,58	1,00		
Julgamento	0,64	0,69	0,57	0,57	0,57	0,60	0,65	1,00	
Blocos	0,51	0,42	0,35	0,35	0,56	0,46	0,59	0,58	1,00

Nota. Num: Números; Per: Percepção; Par: Partes; Jul: Julgamento; Dim: Dimensão; Blo: Blocos; Mem: Memória; Prec: Precisão e Flu: Fluência. Todas as correlações foram significativas ($p < 0,05$).

Estudos indicam que não há diferenças significativas de desempenho cognitivo entre homens e mulheres em testes de inteligência geral. As diferenças significativas geralmente estão associadas às habilidades cognitivas específicas (Flores-Mendoza, 2000; Rueda & Castro, 2013). Em um estudo realizado na Holanda sobre a diferença de desempenho no teste WAIS-III em função do sexo, constatou-se que, nos fatores de Memória de Trabalho e Organização Perceptual, o sexo masculino apresentou médias significativamente superiores às das mulheres. Enquanto que, no fator de Velocidade Perceptual, as médias foram mais altas para o sexo feminino em comparação com o sexo masculino (van der Sluis et al., 2006). Outra pesquisa envolvendo participantes com idades entre 22 e 90 anos, foi constatado que ao longo da vida, as mulheres apresentaram um desempenho significativamente melhor do que os homens em habilidades de escrita. Por outro lado, os homens demonstraram superioridade em habilidades matemáticas desde a idade adulta jovem até a velhice (Kaufman et al., 2009). No

estudo conduzido por Primi et al. (2022), verificou-se um efeito significativo da variável sexo nas provas de RM, RE e RN, revelando desempenhos inferiores para as mulheres em comparação com os homens.

Em cinco dos testes da Bateria TSP, os resultados dos desempenhos dos participantes, analisados com base no sexo, revelaram diferenças significativas. As médias foram estatisticamente maiores entre as mulheres para os testes Fluência e Percepção. Da mesma forma, nos testes Números, Dimensão e Blocos, foram observados resultados médios estatisticamente superiores para os homens. No entanto, os tamanhos de efeito não foram substanciais (com exceção do teste Blocos), indicando que, apesar da significância estatística, as diferenças encontradas não possuem relevância prática significativa.

Ao analisar os resultados das características dos itens e dos testes pela TRI, observa-se um padrão recorrente nos testes da Bateria TSP. Os itens consistentemente apresentam níveis de dificuldade baixos, o que significa que não é necessário possuir habilidades muito avançadas (*theta* alto) dos fatores cognitivos avaliados para respondê-los corretamente. Adicionalmente, a análise da curva de informação do teste sugere que a maioria dos instrumentos da Bateria TSP é mais precisa para sujeitos com níveis de habilidade baixos e, no máximo, médios para a maioria dos testes. O teste Blocos destacou-se por apresentar uma distribuição mais abrangente da dificuldade dos itens, conforme evidenciado pela análise da TRI. Ele foi o único entre os testes que avaliam habilidades espaciais a revelar um tamanho de efeito médio, o que reforçou a interpretação da significância estatística das diferenças de desempenho entre homens e mulheres nesse teste.

Dessa maneira, os resultados para os tamanhos de efeito baixos podem ter sido influenciados pelas tarefas com baixa exigência do *theta* para que o respondente tenha 50% de chance de acertar o item. Mesmo a esses pontos observados, as médias comparadas para em função da variável sexo, foram consistentes com as tendências de melhores desempenhos em habilidades específicas. Os homens geralmente apresentam maior facilidade em habilidades visuoespaciais, raciocínio matemático (testes Números, Dimensão e Blocos), enquanto que as mulheres tenderam a se destacar em tarefas envolvendo habilidades verbais, velocidade perceptual e memória de curto prazo, observadas nos testes Fluência e Percepção (Burgaleta et al., 2012; Flores-Mendoza, 2000; Keith et al., 2008).

Entre as variáveis associadas à inteligência e que podem afetar o desenvolvimento cognitivo, a idade se destaca como uma das mais estudadas. Isso ocorre porque a idade desempenha um papel crucial na compreensão das mudanças na estrutura intelectual dos indivíduos, bem como das diferenças observadas ao longo do tempo (Kaufman, 2001; Neisser

et al., 1996; Salthouse, 2010). Ao examinar as diferenças médias de desempenho nos testes da Bateria TSP conforme as faixas etárias estabelecidas, observa-se que para os testes de Julgamento, Precisão, Partes, Dimensão e Blocos, os indivíduos com 40 anos ou mais apresentaram resultados médios estatisticamente inferiores em comparação com todas as outras faixas etárias. Esses testes, mais associados a habilidades abstratas e processamento visual, apresentam uma tendência de diminuição de desempenho ao longo da vida adulta (Primi et al., 2022; Tucker-Drob, Brandmaier, & Lindenberger, 2019). Dessa forma, os resultados dos testes da Bateria TSP refletem os achados da literatura sobre a inteligência ao longo do ciclo vital, destacando que o declínio geral no desempenho em tarefas que envolvem esse tipo de habilidade se torna mais evidente na faixa etária de 40 anos ou mais.

6. RESULTADOS E DISCUSSÃO DOS MODELOS TEÓRICOS DE INTELIGÊNCIA TESTADOS PELA BATERIA TSP

6.1 Seleção dos itens a serem utilizados nas AFCs

Inicialmente, foram excluídos da análise itens com zero variância ou com variância insuficiente (menos de 1% das respostas corretas ou incorretas), incluindo os itens 50, 51, 52 e 54 do Teste Números, bem como os itens do teste Percepção: 19A, 21B, 23B, 24B, 25A, 25B, 26A, 26B, 27A e 27B, que não apresentaram variância. Itens com pouca variância identificados foram: do item 44 em diante no Teste Números; do item 17B em diante no Teste Percepção; e do item 33 em diante no Teste Precisão.

Após a remoção dos itens, ao testar os modelos previamente estabelecidos não houve convergência. Em função disso, adotou-se a estratégia de reduzir o número de itens por uma avaliação individual dos modelos de cada teste para identificar os itens mais adequados de cada fator. Cada teste foi avaliado para selecionar os itens mais representativos, mantendo-se aqueles com cargas fatoriais $\geq 0,9$ para os Testes de Números, Percepção e Memória; $\geq 0,8$ para os Testes Partes, Precisão, Dimensão e Julgamento; e $> 0,70$ para o Teste Blocos. Essa seleção se justificou pela escassez ou ausência de itens com carga acima de 0,90 nos demais fatores.

A remoção dos itens seguiu um processo iterativo: eliminava-se o item de menor carga fatorial, realizava-se novamente a análise e repetia-se o procedimento até que restassem apenas os itens com as maiores cargas fatoriais. Ressalta-se que foram diversas tentativas para excluir o menor número possível de itens, buscando minimizar a perda de informações.

Porém, essas tentativas não resultaram em modelos convergentes. A Tabela 72, demonstra quais itens permaneceram para a realização das AFCs.

Tabela 72. Itens selecionados para AFC

Teste	Itens que permaneceram para AFC	Total	Itens excluídos
Números	22,23,24,25,26,27,28,30,31,32,33,34,35,36,37,38,39,41,42	19	35
Percepção	12A, 12B, 13A, 13B, 14A, 14B, 15A, 15B, 16A, 16B	10	44
Fluência	1, 2, 3, 4	4	0
Memória	16, 17, 20, 21, 23, 24, 27, 28, 30, 31, 34, 35	12	24
Partes	30, 31, 32, 33, 34, 35, 36, 37	8	40
Precisão	6, 7, 10, 17, 18, 20, 21, 22	8	40
Dimensão	19, 20, 23, 25, 27, 28, 29, 30, 32, 36, 37, 38, 39, 40	14	34
Julgamento	20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30	10	44
Blocos	1, 7, 8, 10, 11, 12, 13, 15, 17, 18	10	22

6.2 Resultados da normalidade multivariada

Os resultados do teste *Mardia* indicaram uma violação da normalidade multivariada dos itens, com assimetria 321.490,6 ($p < 0,001$) e curtose 194,1 ($p < 0,001$). Em função do resultado, optou-se pelo estimador *Weighted Least Squares Means and Variance Adjusted* (WLSMV), recomendado para situações nas quais os dados não exibem normalidade multivariada (Li, 2016). Ressalta-se que essa análise foi realizada apenas com os itens que foram utilizados para as AFCs de itens.

6.3 Resultados das AFCs de itens: ajuste dos modelos, confiabilidade e cargas fatoriais

O ajuste dos modelos é apresentado na Tabela 73. O modelo 1 e 3 foram àqueles que obtiveram ajustes aceitáveis em todos os índices. O modelo 1 teve uma pequena superioridade na qualidade do índice SRMR quando comparado ao modelo 3. No entanto, os dois modelos podem ser considerados os melhores para os índices de ajuste apresentados. Os modelos 2 e 4 seriam rejeitados com base nos valores do SRMR. Já os modelos 5, 6 e 7 seriam rejeitados conforme todos os índices de ajuste. O modelo 8 apresentou diversos itens com cargas negativas no F1, por essa razão, esses itens tiveram suas cargas estrangidas a zero. Esse foi o caso de todos os itens do Teste Números, todos os itens do Teste Julgamento bem como os itens 12A e 12B do Teste de Percepção, o item 2 do Teste Fluência. O modelo 8 seria rejeitado segundo o SRMR e apresentou valores levemente acima do considerado aceitável

para o RMSEA. Em termos de comparação entre modelos semelhantes, o modelo 1 demonstrou ser superior ao modelo 2 (χ^2 [gl]= 742,83 [21]; $p < 0,001$), enquanto o modelo 3 mostrou-se superior ao modelo 4 (χ^2 [gl]= 359,22 [3]; $p < 0,001$). Como são modelos da inteligência semelhantes, as demais análises não irão considerar os modelos 2 e 4 por terem demonstrado ajuste inferior aos modelos 1 e 3, respectivamente.

Tabela 73. Ajuste dos modelos hipotéticos testados (AFC de itens)

Modelo	χ^2 [gl]	CFI	TLI	RMSEA [IC 90%]	SRMR
Modelo 1	11656,4 [4334]	0,996	0,995	0,034 [0,033; 0,035]	0,075
Modelo 2	26576,2 [4355]	0,986	0,986	0,059 [0,058; 0,060]	0,115
Modelo 3	15161,8 [4361]	0,993	0,993	0,041 [0,040; 0,042]	0,087
Modelo 4	26923,9 [4364]	0,986	0,986	0,059 [0,059; 0,060]	0,115
Modelo 5	67630,0 [4367]	0,885	0,883	0,099 [0,099; 0,100]	0,101
Modelo 6	74108,6 [4370]	0,873	0,871	0,104 [0,104; 0,105]	0,105
Modelo 7	69010,5 [4369]	0,883	0,880	0,101 [0,100; 0,101]	0,103
Modelo 8	46571,4 [4307]	0,923	0,921	0,082 [0,081; 0,083]	0,092

Nota. χ^2 = qui-quadrado; gl = graus de liberdade; IC = intervalo de confiança

A confiabilidade dos fatores detalhada na Tabela 74 mostra que o modelo 6, apesar de os fatores de primeira ordem exibirem confiabilidade aceitável, o fator geral não se mostrou confiável (valores menores que 0,60 para alfa e confiabilidade composta). Por outro lado, no modelo 8, a situação se inverte, observa-se que o fator geral possui confiabilidade aceitável em todos os índices, mas os fatores específicos os valores de ômega não. Essa discrepância pode ser atribuída à ortogonalização dos fatores, que cria uma competição entre eles pela explicação da variância dos itens, resultando geralmente em valores mais baixos de ômega, conforme observado no modelo 8 (Araujo, Gomes, & Jelihovschi, 2023). Há uma discrepância relevante entre os valores dos outros índices de confiabilidade e o ômega de McDonald porque o ômega tende a ser mais restritivo, de modo que valores menores para este índice no modelo 8 (que é mais restritivo) já era esperado. Além disso, a confiabilidade não aceitável do fator geral no modelo 6, em contraste com sua confiabilidade aceitável no modelo 8, sugere que o fator geral é mais confiável quando concebido como diretamente responsável pela variância dos itens, do que quando concebido como responsável pela variância dos fatores específicos (para os modelos teóricos em que os testes foram agrupados em função do tipo de estímulo).

Nota-se que o ω de McDonald ultrapassou o valor de 1 em alguns casos (F1 e F4 nos modelos 1 e 3). Inicialmente, isso poderia sugerir problemas de convergência, indicando a necessidade de reespecificações no modelo. No entanto, como não foram detectados problemas de convergência, como variâncias negativas ou casos *Heywood*, não se procedeu com uma nova especificação do modelo. Adicionalmente, é importante destacar que o ω é fortemente influenciado pela magnitude das cargas fatoriais (Flora, 2020). Considerando que estas foram elevadas, é possível que isso justifique seus valores ligeiramente maiores que 1. Com exceção dos modelos 6 e 8, todos os demais modelos apresentaram confiabilidade aceitável para os fatores.

Tabela 74. Confiabilidade dos fatores para os modelos da inteligência testados

	Modelo 1	Modelo 3	Modelo 5	Modelo 6	Modelo 7	Modelo 8
F1	$\alpha = 0,99$	$\alpha = 0,99$	$\alpha = 0,89$	$\alpha = 0,88$	$\alpha = 0,88$	$\alpha = 0,81$
	$\Omega = 1,01$	$\Omega = 1,01$	$\Omega = 0,70$	$\Omega = 0,86$	$\Omega = 0,83$	$\Omega = 0,23$
	c.c. = 0,99	c.c. = 0,99	c.c. = 0,89	c.c. = 0,88	c.c. = 0,88	c.c. = 0,81
F2	$\alpha = 0,98$	$\alpha = 0,98$	$\alpha = 0,92$	$\alpha = 0,93$	$\alpha = 0,93$	$\alpha = 0,93$
	$\Omega = 0,98$	$\Omega = 0,98$	$\Omega = 0,92$	$\Omega = 0,99$	$\Omega = 0,91$	$\Omega = 0,48$
	c.c. = 0,98	c.c. = 0,98	c.c. = 0,92	c.c. = 0,93	c.c. = 0,93	c.c. = 0,93
F3	$\alpha = 0,79$	$\alpha = 0,79$	$\alpha = 0,87$	—	—	—
	$\Omega = 0,85$	$\Omega = 0,85$	$\Omega = 0,82$	—	—	—
	c.c. = 0,79	c.c. = 0,79	c.c. = 0,87	—	—	—
F4	$\alpha = 0,98$	$\alpha = 0,98$	—	—	—	—
	$\Omega = 1,08$	$\Omega = 1,08$	—	—	—	—
	c.c. = 0,98	c.c. = 0,98	—	—	—	—
F5	$\alpha = 0,96$	$\alpha = 0,96$	—	—	—	—
	$\Omega = 0,94$	$\Omega = 0,94$	—	—	—	—
	c.c. = 0,96	c.c. = 0,96	—	—	—	—
F6	$\alpha = 0,97$	$\alpha = 0,97$	—	—	—	—
	$\Omega = 0,96$	$\Omega = 0,96$	—	—	—	—
	c.c. = 0,97	c.c. = 0,97	—	—	—	—
F7	$\alpha = 0,97$	$\alpha = 0,97$	—	—	—	—
	$\Omega = 0,98$	$\Omega = 0,98$	—	—	—	—
	c.c. = 0,97	c.c. = 0,97	—	—	—	—
F8	$\alpha = 0,97$	$\alpha = 0,97$	—	—	—	—
	$\Omega = 0,95$	$\Omega = 0,95$	—	—	—	—
	c.c. = 0,97	c.c. = 0,97	—	—	—	—
F9	$\alpha = 0,93$	$\alpha = 0,90$	—	—	—	—
	$\Omega = 0,84$	$\Omega = 0,89$	—	—	—	—
	c.c. = 0,93	c.c. = 0,90	—	—	—	—
G	—	$\alpha = 0,90$	$\alpha = 0,70$	$\alpha = 0,54$	—	$\alpha = 0,92$
	—	$\Omega^h = 0,89$	$\Omega^h = 0,84$	$\Omega^h = 0,67$	—	$\Omega^h = 0,75$
	—	c.c. = 0,90	c.c. = 0,70	c.c. = 0,54	—	c.c. = 0,92

Nota. α =alfa de Cronbach; Ω = ω de McDonald; Ω^h = ω hierárquico; c.c.=confiabilidade composta

A Tabela 75 apresenta que, mesmo com a inclusão de um fator geral no modelo 3, suas cargas fatoriais permaneceram similares às do modelo 1 para os fatores de primeira ordem (com cargas fatoriais altas), e as cargas fatoriais para o fator geral foram também altas (>0,60). Os outros modelos exibiram cargas fatoriais muito mais baixas. Nota-se que os modelos 5 e 6 demonstram resultados mais favoráveis para o fator geral do que para os fatores específicos. Esses resultados apontam para a relevância do fator geral na estrutura de inteligência testada.

Tabela 75. Estatísticas descritivas das cargas fatoriais

	Modelo 1	Modelo 3	Modelo 5	Modelo 6	Modelo 7	Modelo 8
F1	M = 0,96 (DP=0,02) Min = 0,92 Max = 0,99	M = 0,96 (DP=0,02) Min = 0,92 Max = 0,99	M = 0,48 (DP=0,16) Min = 0,23 Max = 0,72	M = 0,54 (DP=0,14) Min = 0,18 Max = 0,71	M = 0,53 (DP=0,14) Min = 0,17 Max = 0,71	M = 0,15 (DP=0,27) Min = 0,00 Max = 0,76
F2	M = 0,95 (DP=0,03) Min = 0,90 Max = 1,00	M = 0,95 (DP=0,03) Min = 0,90 Max = 1,00	M = 0,51 (DP=0,14) Min = 0,11 Max = 0,79	M = 0,51 (DP=0,15) Min = 0,12 Max = 0,81	M = 0,48 (DP=0,14) Min = 0,11 Max = 0,78	M = 0,35 (DP=0,15) Min = 0,07 Max = 0,65
F3	M = 0,71 (DP=0,15) Min = 0,56 Max = 0,92	M = 0,71 (DP=0,15) Min = 0,56 Max = 0,92	M = 0,55 (DP=0,14) Min = 0,17 Max = 0,72	—	—	—
F4	M = 0,95 (DP=0,02) Min = 0,91 Max = 0,98	M = 0,95 (DP=0,02) Min = 0,91 Max = 0,98	—	—	—	—
F5	M = 0,90 (DP=0,04) Min = 0,83 Max = 0,96	M = 0,89 (DP=0,04) Min = 0,83 Max = 0,96	—	—	—	—
F6	M = 0,92 (DP=0,05) Min = 0,86 Max = 0,99	M = 0,92 (DP=0,05) Min = 0,86 Max = 0,99	—	—	—	—
F7	M = 0,88 (DP=0,06) Min = 0,80 Max = 0,98	M = 0,88 (DP=0,06) Min = 0,80 Max = 0,98	—	—	—	—
F8	M = 0,90 (DP=0,05) Min = 0,83 Max = 0,96	M = 0,90 (DP=0,05) Min = 0,83 Max = 0,96	—	—	—	—
F9	M = 0,77 (DP=0,07) Min = 0,63 Max = 0,86	M = 0,77 (DP=0,08) Min = 0,62 Max = 0,86	—	—	—	—
G	—	M = 0,67 (DP = 0,08) Min = 0,60 Max = 0,86	M = 0,80 (DP=0,04) Min = 0,75 Max = 0,83	M = 0,71 (DP=0,00) Min = 0,71 Max = 0,71	—	M = 0,43 (DP=0,15) Min = 0,07 Max = 0,73

Nota: M=média; DP=Desvio padrão; Min=mínimo; Max=máximo.

As correlações entre os fatores, bem como as cargas dos fatores de primeira ordem nos fatores de segunda ordem, são apresentadas na Tabela 76. Analisando as correlações entre os fatores nos modelos 1 e 7, nota-se que todas são estatisticamente significativas, variando de fracas a moderadas, com a menor sendo 0,27. Essas correlações indicam uma relação relevante entre as habilidades avaliadas pelos testes. Adicionalmente, as cargas dos fatores de primeira ordem revelam a importância do fator geral, que exibe cargas de moderadas a fortes nos três modelos examinados (3, 5 e 6). Esses resultados, portanto, apoiam a concepção de um fator geral de inteligência.

Tabela 76. Correlações entre os fatores e cargas dos fatores de primeira ordem nos fatores de segunda ordem

Correlações entre os fatores									
Modelo 1								Modelo 7	
	Num	Per	Flu	Mem	Par	Pre	Dim	Jul	F2
Per	0,47*								F1 0,67*
Flu	0,51*	0,43*							
Mem	0,44*	0,43*	0,40*						
Par	0,37*	0,35*	0,35*	0,30*					
Pre	0,35*	0,46*	0,42*	0,34*	0,47*				
Dim	0,43*	0,40*	0,38*	0,45*	0,59*	0,55*			
Jul	0,60*	0,59*	0,52*	0,51*	0,53*	0,60*	0,60*		
Blo	0,49*	0,27*	0,35*	0,29*	0,45*	0,39*	0,50*	0,48*	

Cargas dos fatores de primeira ordem nos fatores de segunda ordem					
	Modelo 3		Modelo 5		Modelo 6
	Geral		Geral		Geral
Num	0,72*	F1	0,85*	F1	0,78*
Per	0,70*	F2	0,78*	F2	0,98*
Flu	0,68*	F3	0,86*		
Mem	0,65*				
Par	0,68*				
Pre	0,71*				
Dim	0,76*				
Jul	0,76*				
Blo	0,89*				

Nota. * = $p < 0,01$.

6.4 Resultados das AFCs considerando os escores somados para os modelos 5 a 8: ajuste dos modelos, confiabilidade e cargas fatoriais

Os resultados das análises fatoriais confirmatórias, utilizando escores somados para os modelos 5, 6, 7 e 8, são detalhados na Tabela 77. Em termos de ajuste, todos os modelos submetidos à análise fatorial com escores somados alcançaram um ajuste aceitável para todos os índices avaliados. O que indica que esses modelos apresentam uma estrutura fatorial igualmente válida. Nota-se que o modelo 5, por sua vez, revelou-se um pouco inferior em comparação aos demais modelos testados.

Tabela 77. Ajuste dos modelos hipotéticos testados (AFC por escores somados)

Modelo	χ^2 [gl]	CFI	TLI	RMSEA [IC 90%]	SRMR
Modelo 5	129,669 [25]	0,992	0,988	0,053 [0,045; 0,063]	0,042
Modelo 6	64,985 [26]	0,997	0,996	0,032 [0,220; 0,042]	0,030
Modelo 7	64,985 [26]	0,997	0,996	0,032 [0,220; 0,042]	0,030
Modelo 8	86,925 [20]	0,995	0,991	0,048 [0,038; 0,058]	0,035

Nota. χ^2 = qui-quadrado; gl = graus de liberdade; IC = intervalo de confiança

A Tabela 78 mostra as estatísticas descritivas e a confiabilidade dos fatores. Foram necessárias correções em alguns modelos devido a problemas de convergência: no modelo 5, o F2 teve uma carga acima de 1 (um caso *Heywood*), levando à fixação de sua variância e carga em 1. No modelo 6, fixou-se a variância do F2 em 1 para ajustar a variância negativa. E, no modelo 8, as cargas dos itens de Julgamento e Precisão foram negativas, exigindo que suas cargas fossem limitadas a zero e carregadas apenas pelo fator geral. Esses modelos são apresentados nas Figuras 42 a 45.

Em geral, o modelo 8 não atingiu nos fatores específicos um nível aceitável no ômega de McDonald. Esse fenômeno pode ser atribuído à maior exigência do ômega para fatores com cargas fatoriais mais baixas. Os modelos 6 e 7 tiveram confiabilidade idêntica e cargas fatoriais também (Figuras 43 e 44), este resultado é justificado pela estrutura idêntica para os fatores de primeira ordem. A diferença entre os dois modelos está na presença de um fator geral de segunda ordem no modelo 6, que mostra índices de confiabilidade adequados e cargas fatoriais robustas, apontando mais uma vez para a presença de um fator geral de inteligência. No modelo 5 o fator 1 é o que apresenta menores índices de confiabilidade, mas mesmo assim considerados adequados ($>0,60$). Este menor índice pode ser em função da menor quantidade de itens presentes no F1.

Tabela 78. Confiabilidade e estatísticas descritivas dos fatores

	Modelo 5	Modelo 6	Modelo 7	Modelo 8
F1	M=0,70; DP=0,05	M=0,72; DP=0,11	M=0,72; DP=0,11	M=0,26; DP=0,16
	$\alpha = 0,62$	$\alpha = 0,77$	$\alpha = 0,77$	$\alpha = 0,72$
	$\Omega = 0,66$	$\Omega = 0,78$	$\Omega = 0,78$	$\Omega = 0,17$
	c.c. = 0,62	c.c. = 0,77	c.c. = 0,77	c.c. = 0,72
F2	M=0,75; DP=0,06	M=0,73; DP=0,05	M=0,73; DP=0,05	M=0,27; DP=0,19
	$\alpha = 0,76$	$\alpha = 0,80$	$\alpha = 0,80$	$\alpha = 0,76$
	$\Omega = 0,80$	$\Omega = 0,82$	$\Omega = 0,82$	$\Omega = 0,17$
	c.c. = 0,76	c.c. = 0,80	c.c. = 0,80	c.c. = 0,76
F3	M=0,70; DP=0,13	—	—	—
	$\alpha = 0,73$	—	—	—
	$\Omega = 0,73$	—	—	—
G	M=0,93; DP=0,06	M=0,90; DP=0,12	—	M=0,66; DP=0,10
	$\alpha = 0,82$	$\alpha = 0,75$	—	$\alpha = 0,85$
	$\Omega = 0,82$	$\Omega = 0,75$	—	$\Omega = 0,80$
	c.c. = 0,82	c.c. = 0,75	—	c.c. = 0,85

Nota. M=média; DP=desvio padrão; α =alfa de Cronbach; Ω =ômega de McDonald; c.c.=confiabilidade composta

Observando a magnitude das cargas fatoriais (Figuras 42 a 45), todos os modelos exibiram um padrão semelhante nos fatores específicos, os quais se mostraram importantes na explicação da variância dos escores somados. Contudo, a importância do fator geral se destaca nos modelos 5, 6 e 8, onde o fator g apresentou uma carga significativa, sendo no modelo 8 especialmente mais determinante na explicação dos escores somados do que os fatores específicos. Ainda sobre o modelo 8, destaca-se que os escores somados do teste Julgamento e Precisão carregaram negativamente em seus fatores específicos e, por essa razão, tiveram suas cargas estrangidas a zero em seus respectivos fatores, F1 e F2. Observa-se que a carga fatorial obtida pelo teste Julgamento nos diferentes modelos testados foi 0,89 para o fator de primeira ordem ou fator geral. Isso sugere que a característica avaliada pelo teste Julgamento mostra-se equivalente às habilidades mais amplas.

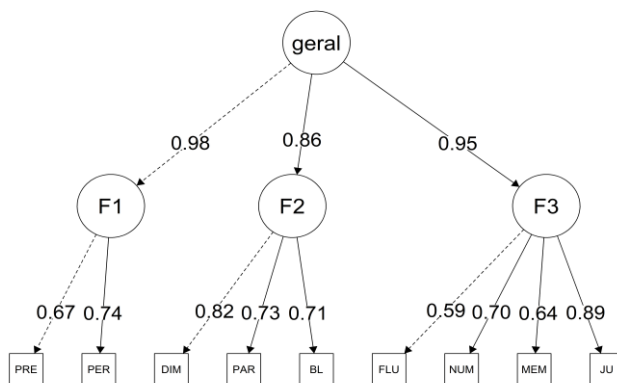


Figura 42. Estrutura fatorial do modelo 5

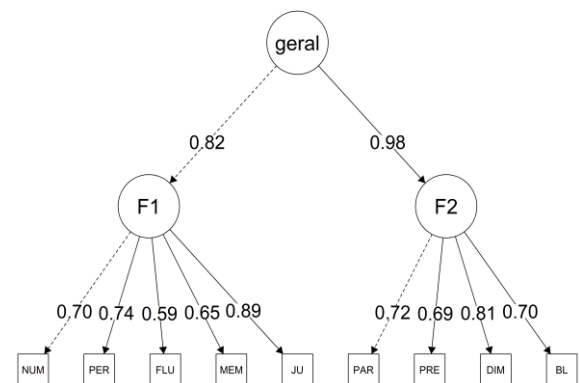


Figura 43. Estrutura fatorial do modelo 6

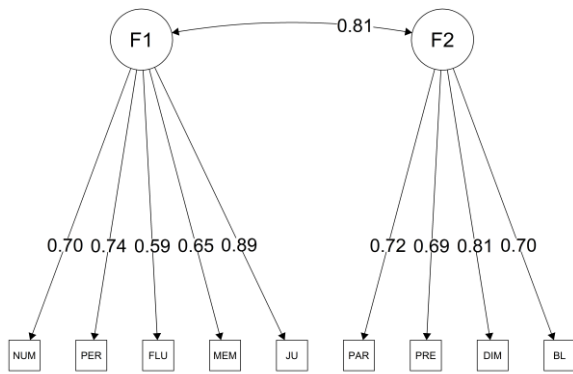


Figura 44. Estrutura fatorial do modelo 7

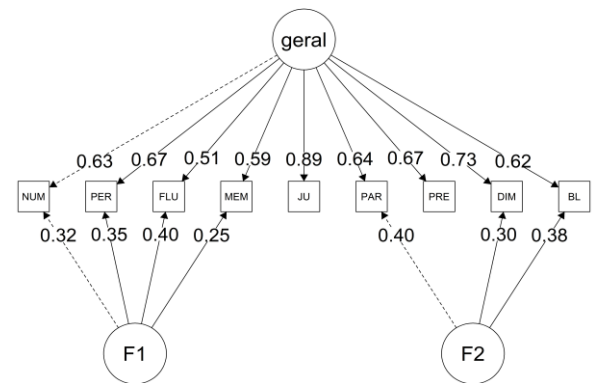


Figura 45. Estrutura fatorial do modelo 8

Em linhas gerais, fica evidente que os ajustes foram significativamente melhores na avaliação por escore somado do que na análise de itens agrupados. No entanto, comparar diretamente esses modelos é delicado, pois suas variáveis observáveis são distintas, o que impossibilita uma comparação equitativa. É importante destacar que os modelos testados por escore somado apresentam uma menor quantidade de parâmetros para serem estimados em comparação com os modelos de itens, o que tende a apresentar convergência mais fácil. Essa disparidade no número de parâmetros pode explicar as diferenças nos resultados obtidos, especialmente considerando que o tamanho amostral permaneceu o mesmo. Dado o grande número de itens e parâmetros a serem estimados nas AFCs de itens, seria necessária uma amostra consideravelmente maior do que para as análises pelos escores somados.

Os modelos que melhor se adequaram aos dados, fornecendo insights sobre a estrutura subjacente da Bateria TSP por meio da AFC de itens, foi o Modelo 1 de Thurstone, com fatores primários correlacionados, e o Modelo 3, alinhado à teoria de Carroll, que introduz um fator geral de inteligência. Esses modelos demonstraram estruturas fatoriais aceitáveis. Em ambos, foram incluídos para as análises todos os testes da Bateria TSP considerados individualmente. Nota-se que os índices de ajustes, a confiabilidade dos fatores e as cargas fatoriais foram muito próximos. No modelo 3, o fator geral demonstrou uma boa consistência em todos os parâmetros avaliados, evidenciando que um fator geral de inteligência se sustenta na estrutura analisada.

Os modelos testados que demonstraram os melhores índices, considerando os escores somados, foram a estrutura fatorial 6 e 7. Ambos os modelos apresentaram índices de ajuste adequados, confiabilidade e cargas fatoriais aceitáveis. Em ambos os modelos, os testes com estímulo verbal foram agrupados em F1, enquanto os testes com estímulo não verbal foram

agrupados em F2. Sugerindo, assim, que um melhor agrupamento para os testes da Bateria TSP seria em função do estímulo.

Para uma análise mais precisa sobre a relevância do fator geral, é interessante avaliar o modelo 8 testado. Nele foi utilizado o método de ortogonalização de fatores de Schmid e Leiman (1957) por meio de uma AFC. O modelo 8 foi examinado tanto com base nos itens individuais quanto nos escores somados dos testes da Bateria TSP. No modelo 8 da AFC de itens, os índices de confiabilidade foram aceitáveis para o fator geral, porém o ômega de McDonald não foi adequado para os fatores específicos. Por outro lado, na AFC de escores somados, houve uma melhoria nos índices de ajuste, sendo nessa análise considerados adequados. Notou-se que os índices de confiabilidade e as cargas fatoriais foram mais adequadas para o fator geral, indicando sua maior contribuição na explicação dos escores somados em detrimento dos fatores específicos. Outro ponto, é a maior carga fatorial atribuída ao teste Julgamento, condição já esperada em função das características do teste (Schneider & McGrew, 2012, 2018). Esses resultados destacam uma informação relevante sobre a quantidade de variância explicada por cada fator em relação aos testes da Bateria TSP, onde o fator geral apresenta-se de maneira relevante para a estrutura fatorial do teste.

Os demais modelos hierárquicos, 5 e 6 testados, demonstraram também resultados fortes das cargas fatoriais dos fatores de primeira ordem no fator geral de segunda ordem, introduzindo uma parcela da variância atribuída ao fator geral. Os índices de confiabilidade e as estatísticas descritivas dos fatores foram superiores para o fator geral no modelo 5, quando comparado aos fatores de primeira ordem.

A introdução de um fator geral de inteligência na concepção dos testes psicológicos de inteligência é respaldada por evidências teóricas e metodológicas (Laros, Jesus, & Karino, 2013; Schneider & McGrew, 2018). A possibilidade de propor um escore geral da inteligência a partir dos nove testes da Bateria TSP está alinhada com as práticas anteriores, como o escore ponderado proposto por King (1957) em seus estudos originais.

Em geral, observa-se que na AFC por itens, predominaram as estruturas baseadas nos testes analisados individualmente, como modelos semelhantes aos de Thurstone e de Carroll (modelos 1 e 3). Por outro lado, na AFC por escores somados, as estruturas fatoriais mais parcimoniosas foram aquelas com os testes agrupados com base no tipo de estímulo (verbal e não verbal) e, apresentaram semelhanças com as estruturas da inteligência propostas por Vernon e por Cattell e Horn (modelos 5 e 6). É importante ressaltar que não foi possível testar os modelos 1 e 3 pela AFC de escores somados. Além disso, como mencionado anteriormente, as AFC's conduzidas por itens e por escores somados não podem ser

comparadas diretamente. Embora haja divergências entre os modelos de AFC por itens e por escores somados, entende-se que cada abordagem oferece uma perspectiva valiosa sobre a estrutura fatorial da Bateria TSP. A própria história dos modelos de inteligência sugere que as teorias propostas pelos diferentes pesquisadores ao longo do tempo não são excludentes (Wasserman, 2018).

7. CONSIDERAÇÕES FINAIS

A presente tese representa um avanço no campo das investigações da Bateria TSP no Brasil. Os estudos utilizando a TRI e os modelos de inteligência foram os primeiros a serem realizados com esse conjunto de testes. Além disso, ela se mostra como uma importante contribuição para o estudo da inteligência, tanto pelos modelos teóricos testados quanto pelo entendimento da relação entre as habilidades cognitivas e as características sociodemográficas investigadas.

Foi objetivo da presente tese realizar novos estudos sobre as propriedades psicométricas da Bateria TSP. Em uma análise geral, os resultados obtidos fornecem informações importante sobre a Bateria TSP e apontam direções para futuras investigações. Embora as evidências apresentadas sejam satisfatórias de acordo com os *Standards* estabelecidos (AERA, APA, NCME, 2014), é crucial continuar produzindo novas evidências de diferentes fontes para garantir a qualidade da avaliação por meio dos testes da Bateria TSP.

Ao examinar os resultados das características dos itens e dos testes utilizando a Teoria de Resposta ao Item, pode-se observar um padrão consistente nos testes da Bateria TSP. Em geral, os itens tendem a ter níveis de dificuldade relativamente baixos, o que significa que não é necessário possuir habilidades altas (*theta* alto) nos fatores cognitivos avaliados para respondê-los corretamente. Além disso, ao analisar a curva de informação dos testes, a maioria dos instrumentos da Bateria TSP é mais precisa para sujeitos com níveis de habilidade mais baixos e, no máximo, médios. Isso sugere que a Bateria se destaca mais como testes de velocidade, no qual o desempenho do sujeito tende a ser mais afetado pela capacidade de realizar rapidamente a tarefa do que pela habilidade específica em si. Talvez essa condição possa fazer parte da proposta original da Bateria, conforme pode ser observado nas próprias instruções do teste:

Vocês irão fazer uma série de testes muito fáceis. É possível que todas as respostas pudessem ser acertadas se vocês tivessem tempo suficiente. Mas é aí que está a dificuldade destes testes, ou seja, o tempo para fazer cada um deles é de 5 minutos.

Isto quer dizer que não devem perder tempo quando uma questão for mais difícil: escolham a resposta que acharem melhor ou então pule essa questão. Não percam tempo com questões que pareçam difíceis, o importante é responder o maior número de questões possível.” (às vezes alguém pode perguntar se deve “chutar”; o aplicador responderá afirmativamente, mas sem muita ênfase, e se perguntado se “chutando” perde-se ponto, responderá que não) (Nascimento & Flores-Mendoza, 2007, p.50).

No entanto, isso não invalida as limitações e os impactos que podem ter na mensuração das habilidades cognitivas e na necessidade de testes com menos influência da velocidade e maior relevância para a mensuração da habilidade cognitiva. É fundamental a realização de futuros estudos com a Bateria TSP, incluindo estímulos com maior nível de complexidade, inserindo itens com maior grau de dificuldade. Além disso, os resultados para cada um dos instrumentos revelaram alguns itens com coeficientes inadequados, que deverão ser modificados ou removidos em futuras atualizações do teste.

Outro aspecto crítico relacionado aos testes da Bateria TSP está ligado à qualidade da impressão. Por exemplo, no teste de Precisão, que requer que o sujeito identifique igualdades, alguns itens apresentam pequenas imprecisões que podem confundir os participantes mais atentos. Em certos casos, os participantes verbalizaram que havia um ponto preto que não estava na figura alvo que deveria ser identificada, levando à incerteza sobre a resposta correta. Isso pode contribuir para um maior erro na medida e, conseqüentemente, para os índices de qualidade do item e do próprio instrumento inadequados.

Dentre as limitações apresentadas neste trabalho, destaca-se que, embora o banco de dados tenha sido construído registrando "certo" ou "errado" em cada item, na maioria dos testes, como Partes, Dimensão e Julgamento, há quatro opções de respostas disponíveis. Seria mais rigoroso e interessante realizar análises baseadas na resposta específica marcada pelo sujeito.

A investigação do modelo de estrutura da inteligência da Bateria TSP foi conduzida por meio de análises fatoriais confirmatórias para testar modelos baseados em teorias já estabelecidas, bem como alguns modelos alternativos. Foram realizadas AFCs de itens e por escores somados, com a limitação de não ser possível testar todos os modelos por ambos os métodos. Foi possível verificar que o modelo teórico de Thurstone permanece adequado para explicar a estrutura da Bateria TSP. No entanto, entre os outros modelos testados, foi observado que um fator geral se mostra consistente e relevante na explicação do desempenho nos testes da Bateria, tanto isoladamente quanto agrupados com base no tipo de estímulo (verbal e não verbal). Dois diferentes modelos mostraram-se adequados pelas análises de itens, semelhantes às teorias de Thurstone e Carroll. Além disso, dois modelos demonstraram

ajuste satisfatório quando analisados pelo score somado dos testes, condizentes com as teorias de Cattell e Horn, e Vernon.

Dessa maneira, para dar consistência aos resultados deste estudo, são necessárias análises futuras para testar os modelos de inteligência. Diversas técnicas podem ser empregadas, incluindo modelagem por equações estruturais exploratórias, bem como análises de invariância em subamostras divididas por escolaridade, idade e sexo. Isso permitirá verificar a estabilidade dos fatores em diferentes grupos amostrais e aprofundar a compreensão das relações entre os construtos avaliados.

É fundamental ressaltar que, no Brasil, o instrumento ainda não possui uma versão informatizada para aplicação e correção, ressaltando a importância de estudos para viabilizar essa possibilidade. Destaca-se a importância da criação da correção informatizada do teste em função da possibilidade de construção de um banco de dados nacional, com uma amostra ampliada das diferentes regiões do Brasil. Isso possibilitaria que pesquisas futuras abrangessem um maior número de sujeitos e uma amostra mais heterogênea.

REFERÊNCIAS

- Almeida, L. S., Nascimento, E., Lima, A. O. F., Vasconcelos, A. G., Akama, C. T., & Santos, M. T. (2010). Bateria de provas de raciocínio (BPR-5): estudo exploratório em alunos universitários. *Avaliação Psicológica*, 9(2), 155-162.
http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000200002&lng=pt&tlng=pt
- Ambiel, R. A. M., & Pacanaro, S. V. (2019). Da testagem à avaliação psicológica: aspectos históricos e perspectivas futuras. Em R. A. M. Ambiel et al. (Orgs.), *Avaliação Psicológica: guia de consulta para estudantes e profissionais de Psicologia* (pp.11-25). Artesã.
- American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Andrade, J. M., Laros, J. A., & Gouveia, V. V. (2010). O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Avaliação Psicológica*, 9(3), 421-435. <http://pepsic.bvsalud.org/pdf/avp/v9n3/v9n3a09.pdf>
- Andrés-Pueyo, A. (2006). Modelos psicométricos da inteligência. Em C. E. Flores-Mendoza & R. Colom (Orgs.), *Introdução à psicologia das diferenças individuais* (pp. 71-100). Artmed.
- Araujo, J. D., Gomes, C. M. A., & Jelihovschi, E. G. (2023). The factor structure of the Motivated Strategies for Learning Questionnaire (MSLQ): new methodological approaches and evidence. *Psicologia: Reflexão e Crítica*, 36(38).
<https://doi.org/10.1186/s41155-023-00280-0>
- Barbetta, P. A., Trevisan, L. M., Tavares, H., & Azevedo, T. C. A. M. (2014). Aplicação da Teoria da Resposta ao Item uni e multidimensional. *Estudos em Avaliação Educacional*, 25(57), 280-302. [10.18222/ea255720142832](https://doi.org/10.18222/ea255720142832)
- Bilash, I.; Zubek, J. P. (1960). The Effects of Age on Factorially "Pure" Mental Abilities. *Journal of Gerontology*, 15(2), 175–182. [doi:10.1093/geronj/15.2.175](https://doi.org/10.1093/geronj/15.2.175)
- Bishara, A. J., Li, J., & Conley, C. (2021). Informal versus formal judgment of statistical models: The case of normality assumptions. *Psychonomic Bulletin & Review*, 28, 1164–1182. <https://doi.org/10.3758/s13423-021-01879-z>
- Boccalandro, E. R. (2018). *G-38: Teste Não Verbal de Inteligência*. 6 ed., Vetor.

- Brown, R. E. (2016). Hebb and Cattell: The genesis of the theory of fluid and crystallized intelligence. *Frontiers in Human Neuroscience*, *10*, 606.
<https://doi.org/10.3389/fnhum.2016.00606>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications.
- Burgaleta, M., Head, K., Álvarez-Linera, J., Martínez, K., Escorial, S., Haier, R., & Colom, R. (2012). Sex differences in brain volume are related to specific skills, not to general intelligence. *Intelligence*, *40*(1), 60–68. <https://doi.org/10.1016/j.intell.2011.10.006>
- Caemmerer, J. M., Keith, T. Z., & Reynolds, M. R. (2020). Beyond individual intelligence tests: Application of Cattell-Horn-Carroll Theory. *Intelligence*, *79*, 101433.
<https://doi.org/10.1016/j.intell.2020.101433>
- Campos, C. R., Zaia, P., & Primi, R. (2019). Avaliação psicológica da inteligência. Em M. N. Baptista et al. (Orgs.), *Compêndio de Avaliação Psicológica* (pp. 349-363). Vozes.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York, NY: Cambridge University Press.
- Carroll, J. B. (2005). The Three-Stratum Theory of Cognitive Abilities. In D. P. Flanagan & P. L. Harrison (Orgs.), *Contemporary intellectual assessment: Theories, tests and issues* (2ed., pp. 69-76). New York: Guilford Press.
- Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., KwonHyun, K., Falk, C. F., & Meade, A. (2015). Package ‘mirt’. *Zugriff am*, *21*(01). <https://cran.r-project.org/web/packages/mirt/mirt.pdf>
- Chiodi, M. G. (2012). Estudo de adaptação e validação da bateria de habilidades cognitivas Woodcock-Johnson-III versão ampliada. 221 f. Tese (Doutorado em Psicologia) - Pontifícia Universidade Católica de Campinas, Campinas.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge.
<https://doi.org/10.4324/9780203771587>
- Conselho Federal de Psicologia. (2022). *Resolução CFP n. 31/2022*. Estabelece diretrizes para a realização de Avaliação Psicológica no exercício profissional da psicóloga e do psicólogo, regulamenta o Sistema de Avaliação de Testes Psicológicos - SATEPSI e revoga a Resolução CFP nº 09/2018. <https://atosoficiais.com.br/lei/avaliacao-psicologica-cfp?origin=instituicao>
- Couto, G., & Primi, R. (2011). Teoria de resposta ao item (TRI): conceitos elementares dos modelos para itens dicotômicos. *Boletim de Psicologia*, *61*(134), 1-15.
<http://pepsic.bvsalud.org/pdf/bolpsi/v61n134/v61n134a02.pdf>

- Dancey, C. P.; Reidy, J. (2019). *Estatística sem matemática para psicologia*. 7ed. Artmed.
- Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of pediatric psychology*, 34(9), 917-928. <https://doi.org/10.1093/jpepsy/jsp004>
- Edites (s.d.). *Bateria TSP: manual*. São Paulo: Edites.
- Fernandes, D. C., Rueda, F. J. M., & Sisto, F. F. (2009). Edad e inteligencia? variables relacionadas?. *Avaliação Psicológica*, 8(3), 303-311. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712009000300003&lng=pt&tlng=es.
- Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In *Encyclopedia of Special Education* (eds C.R. Reynolds, K.J. Vannest and E. Fletcher-Janzen). <https://doi.org/10.1002/9781118660584.ese0431>
- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, 3(4), 484-501. <https://doi.org/10.1177/2515245920951747>
- Flores-Mendoza, C. (2000). Diferenças intelectuais entre homens e mulheres: uma breve revisão da literatura. *Psicólogo inFormação*, 4(4), 25-34. https://www.fafich.ufmg.br/ladi/files/Art001_SexDifferences_2000.pdf
- Golino, H. F. (2017). A importância da psicometria na avaliação psicológica: conceitos básicos, fundamentos epistemológicos da validade uma análise ilustrativa do National Intelligence Tests. Em M. R. C. Lins & J. C. Borsa (Eds.), *Avaliação psicológica: Aspectos teóricos e práticos* (pp. 137-158). Vozes.
- Gomes, C. M. A. (2010). Estrutura fatorial da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCaLo). *Avaliação Psicológica*, 9(3), 449-459. <http://pepsic.bvsalud.org/pdf/avp/v9n3/v9n3a11.pdf>
- Gomes, C. M. A., & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira. *Avaliação Psicológica*, 6(2), 167-179. <http://pepsic.bvsalud.org/pdf/avp/v6n2/v6n2a07.pdf>
- Gomes, C. M. A., & Borges, O. N. (2009a). Propriedades psicométricas do conjunto de testes da habilidade visuo espacial. *Psico-USF*, 14(1), 19-34. <https://doi.org/10.1590/S1413-82712009000100004>
- Gomes, C. M. A., & Borges, O. (2009b). Qualidades psicométricas do conjunto de testes de inteligência fluida. *Avaliação Psicológica*, 8(1), 17-32. <http://pepsic.bvsalud.org/pdf/avp/v8n1/v8n1a03.pdf>

- Gomes, C. M. A., Golino, H. F., & de Souza Peres, A. J. (2020). Fidedignidade Dos Escores Do Exame Nacional Do Ensino Médio (ENEM). *Psico*, *51*(2), e31145.
<https://doi.org/10.15448/1980-8623.2020.2.31145>
- Gomes, C. M. A., Nascimento, D., & Araújo, J. (2021). Medindo a Inteligência Fluida: o Teste de Indução da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). *Research Gate*. 10.13140/RG.2.2.17087.84641/3
- Gottfredson, L. S. (2006). Consequências sociais das diferenças de grupo na capacidade cognitiva. Em C. E. Flores-Mendoza & R. Colom (Orgs.), *Introdução à psicologia das diferenças individuais* (pp. 433-456). Artmed.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados*. Porto Alegre: Bookman.
- Hauck-Filho, N., & Valentini, F. (2020). Coeficientes de fidedignidade e violações dos pressupostos essencialmente tau-equivalentes [Editorial]. *Avaliação Psicológica*, *19*(3), A-B. <http://dx.doi.org/10.15689/ap.2020.1903.ed>
- Haukoos, J. S., & Lewis, R. J. (2005). Advanced statistics: Bootstrapping confidence intervals for statistics with "difficult" distributions. *Academic Emergency Medicine*, *12*(4), 360-365.
<https://doi.org/10.1197/j.aem.2004.11.018>
- Hayes, A., & Coutts, J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. *But. Communication Methods and Measures*, *14*(1), 1-24.
<https://doi.org/10.1080/19312458.2020.1718629>
- Henderson, A. (2006). Testing experimental data for univariate normality. *Clinica Chimica Acta*, *366*(1), 112-129. <https://doi.org/10.1016/j.cca.2005.11.007>
- Horn, J. L., & Blankson, N. (2012). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Orgs.), *Contemporary intellectual assessment: Theories, tests and issues* (3ed., pp. 73-98). New York: Guilford Press.
- Horn, J. L., Donaldson, G., & Engstrom, R. (1981). Apprehension, Memory, and Fluid Intelligence Decline in Adulthood. *Research on Aging*, *3*(1), 33–84. <https://doi.org/10.1177/016402758131002>
- Horn, J. L., & Stankov, L. (1982). Auditory and visual factors of intelligence. *Intelligence*, *6*(2), 165–185. [https://doi.org/10.1016/0160-2896\(82\)90012-5](https://doi.org/10.1016/0160-2896(82)90012-5)
- Hu, L.-t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424-453. doi: 10.1037/1082-989X.3.4.424

- Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, 75(1), 28–42. <https://doi-org.ez27.periodicos.capes.gov.br/10.1037/0021-9010.75.1.28>
- Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, 2(3-4), 447-472. <https://doi.org/10.1037/1076-8971.2.3-4.447>
- JASP Team (2024). JASP (Version 0.18.3)[Computer software]. <https://jasp-stats.org/download/>
- Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33(4), 393–416. <https://doi.org/10.1016/j.intell.2004.12.002>
- Johnson, W., te Nijenhuis, J., & Bouchard Jr, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36(1), 81-95. <https://doi.org/10.1016/j.intell.2007.06.001>
- Jorgensen, T.D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *Useful tools for structural equation modeling* [R package semTools version 0.5-5]. <https://CRAN.R-project.org/package-semTools>
- Júnior, R. M. S, Nascimento, A. M., & Roazzi, A. (2019). Bateria de Provas de Raciocínio (BPR-5): Avaliação das qualidades psicométricas em adolescentes do Nordeste. *Amazonica – Revista de Psicopedagogia, Psicologia Escolar e Educação*, 23(1), 264-288. <https://periodicos.ufam.edu.br/index.php/amazonica/article/view/5174>
- Kalkbrenner, M. T. (2021). A Practical Guide to Instrument Development and Score Validation in the Social Sciences: The MEASURE Approach. *Practical Assessment, Research, and Evaluation*, 26(1), 1-18. <https://doi.org/10.7275/svg4-e671>
- Kaufman, A.S. (2001). WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence*, 29(2), 131-167. [https://doi.org/10.1016/S0160-2896\(00\)00046-5](https://doi.org/10.1016/S0160-2896(00)00046-5)
- Kaufman, A. S. & Horn, J. L. (1996). Age changes on tests of fluid and crystallized ability for women and men on the Kaufman Adolescent and Adult Intelligence Test (KAIT) at ages 17-94 years. *Archives of clinical Neuropsychology*, 11(2), 97-121. <https://doi-org.ez27.periodicos.capes.gov.br/10.1093/arclin/11.2.97>
- Kaufman, A. S., Kaufman, J. C., Liu, X., & Johnson, C. K. (2009). How do Educational Attainment and Gender Relate to Fluid Intelligence, Crystallized Intelligence, and

- Academic Skills at Ages 22–90 Years? *Archives of Clinical Neuropsychology*, 24(2), 153–163. <https://doi-org.ez27.periodicos.capes.gov.br/10.1093/arclin/acp015>
- Keith, T., Fine, J., Taub, G., Reynolds, M. & Kranzler, J. (2006). Higher-Order, Multi-Sample, Confirmatory Factor Analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What Does it Measure? *School Psychology Review*, 35(1), 108-127. <https://www.researchgate.net/publication/232489100>
- Keith, T. Z., Reynolds, M. R. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: evidence from the Woodcock-Johnson III tests of cognitive abilities. *Intelligence*, 36(6), 502-525. <https://doi.org/10.1016/j.intell.2007.11.001>
- Kell H. J., & Lang J. W. B. (2017). Specific Abilities in the Workplace: More Important Than *g*? *Journal of Intelligence*, 5(2), 13. <https://doi.org/10.3390/jintelligence5020013>
- Kerby, D. S. (2014). The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. *Comprehensive Psychology*, 3(1). <https://doi.org/10.2466/11.IT.3.1>
- King, J. E. (1957). Factored Aptitude Series of Business and Industrial Tests. *The Personnel and Guidance Journal*, 35(6), 351-358. <https://doi-org.ez27.periodicos.capes.gov.br/10.1002/j.2164-4918.1957.tb01782.x>
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2021). *MVN: Multivariate Normality Tests* (5.9). [R package]. <https://cran.r-project.org/web/packages/MVN/index.html>
- Lang, J. W. B., Kersting, M., Hülshager, U. R., & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested-factors model of cognitive abilities. *Personnel Psychology*, 63(3), 595–640. <https://doi.org/10.1111/j.1744-6570.2010.01182.x>
- Laros, J. A., Jesus, G. R., & Karino, C. A. (2013). Validação brasileira do teste não-verbal de inteligência SON-R 2½-7[a]. *Avaliação Psicológica*, 12(2), 233-242. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712013000200014
- Larson, E. C. (2019). A Meta-Analysis of Information Processing Measures of Intelligence, Performance, and Group Score Differences. *CUNY Academic Works*. https://academicworks.cuny.edu/gc_etds/3040
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>

- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320-341. doi: 10.1207/s15328007sem1103_2
- Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological methods*, *9*(3), 275. <http://dx.doi.org/10.1037/1082-989X.9.3.275>
- McGill, R. J., & Dombrowski, S. C. (2019). Critically Reflecting on the Origins, Evolution, and Impact of the Cattell-Horn-Carroll (CHC) Model. *Applied Measurement in Education*, *32*(3), 216–231. [https://doi-org.ez27.periodicos.capes.gov.br/10.1080/08957347.2019.1619561](https://doi.org.ez27.periodicos.capes.gov.br/10.1080/08957347.2019.1619561)
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: past, presente, and future. In D. P. Flanagan & P. L. Harrison (Orgs.), *Contemporary intellectual assessment: Theories, tests and issues* (2ed., pp. 136-181). New York: Guilford Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1-10. <https://doi.org/10.1016/j.intell.2008.08.004>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. <https://doi.org/10.1037/met0000144>
- Miot, H. A. (2017). Avaliação da normalidade dos dados em estudos clínicos e experimentais. *Jornal Vascular Brasileiro*. *16*(2), pp. 88-91. doi.org/10.1590/1677-5449.041117
- Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A Bifactor Exploratory Structural Equation Modeling Framework for the Identification of Distinct Sources of Construct-Relevant Psychometric Multidimensionality. *Structural Equation Modeling: A Multidisciplinary*, *23*(1), 116-139. doi: 10.1080/10705511.2014.961800
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, *31*(1), 7-16. [doi: 10.7334/psicothema2018.291](https://doi.org/10.7334/psicothema2018.291)
- Muniz, M., & Lins, M. R. C. (2018). Avaliação da inteligência aspectos não verbais. Em Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (Orgs.). *Avaliação psicológica da inteligência e da personalidade*. Artmed.
- Nascimento, E. & Flores-Mendoza, C. E. (2007). *Manual da Bateria TSP (revalidação)*. Edites.

- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101.
<https://doi.org/10.1037/0003-066X.51.2.77>
- Olea, J. & Abad, F. J. (2006). A medição das diferenças individuais. Em C. E. Flores-Mendoza & R. Colom (Orgs.), *Introdução à psicologia das diferenças individuais* (pp. 325-354). Artmed.
- Ortiz, S. O. (2015). CHC Theory of Intelligence. In S. Goldstein, D. Princiotta, & J. A. Naglieri (eds.), *Handbook of Intelligence: Evolutionary Theory, Historical Perspective, and Current Concepts* (pp. 209-227). Springer
- Pasquali, L. (2010). *Instrumentação psicológica: fundamentos e práticas*. Artmed.
- Pasquali, L. (2020). *TRI–Teoria de resposta ao item: Teoria, procedimentos e aplicações*. Editora Appris.
- Pasquali, L., & Primi, R. (2003). Fundamentos da teoria da resposta ao item: TRI. *Avaliação Psicológica*, *2*(2), 99-110.
http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712003000200002
- Peres, A. J. S., Lessa, J. P. A., & Valentini, F. Teoria de Resposta ao Item (TRI). Em M. N. Baptista et al. (Orgs.), *Compêndio de Avaliação Psicológica* (pp. 53-68). Vozes.
- Pilati, R., & Laros, J. A. (2007). Modelos de Equações Estruturais em Psicologia: conceitos e aplicações. *Psicologia: Teoria e Pesquisa*, *23*(2), 205-216. doi: 10.1590/S0102-37722007000200011
- Primi, R. (2003). Inteligência: avanços nos modelos teóricos e nos instrumentos de medida. *Avaliação Psicológica*, *2*(1), 67-77.
http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712003000100008&lng=pt&tlng=pt
- Primi, R., & Almeida, L. S. (2000). Estudo de validação da bateria de provas de raciocínio (BPR-5). *Psicologia: Teoria e Pesquisa*, *16*(2), 165-173. <https://doi.org/10.1590/S0102-37722000000200009>
- Primi, R., & Almeida, L. S. (2018). *BPR-5 – Bateria de Provas de Raciocínio: Estudos de Revisão das Normas*. Pearson Clinical Brasil.
- Primi, R., Almeida, L. S., Nakano, T. C., & Campos, C. R. (2022). Normatização da Bateria de Provas de Raciocínio (BPR-5) Usando a Teoria de Resposta ao Item. *Avaliação Psicológica*, *21*(2), 127-139. <https://dx.doi.org/10.15689/ap.2022.2102.20136.01>

- Primi, R., Correia, T. A., & Almeida, L. S. (2018). Bateria de Provas de Raciocínio (BPR-5). Em Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (Orgs.). *Avaliação psicológica da inteligência e da personalidade*. Artmed.
- Primi, R., Nakano, T. C., & Wechsler, S. M. (2012). Cross-battery factor analysis of the Battery of Reasoning Abilities (BPR-5) and Woodcock-Johnson Tests of Cognitive Ability (WJ-III): Bateria de Habilidades de Raciocínio (BPR-5) e Bateria de Habilidades Cognitivas Woodcock-Johnson III. *Temas em Psicologia*, 20(1), 121-132. <https://www.redalyc.org/pdf/5137/513751439010.pdf>
- R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.r-project.org/>
- Reppold, C. T., Zanini, D. A., & Noronha, A. P. P. (2019). O que é avaliação psicológica?. Em: M. N. Baptista et al. (Orgs.), *Compêndio de Avaliação Psicológica* (pp. 15-28). Vozes.
- Reynolds M. R., Keith T. Z., Flanagan D. P., Alfonso V.C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology*, 51(4), 535-55. <https://doi.org/10.1016/j.jsp.2013.02.003>
- Robbins, P. S., Judge, T. A., & Sobral, F. (2010). *Comportamento Organizacional: teoria e prática no contexto brasileiro*. Pearson Prentice Hall.
- Rosseel, Y., Jorgensen, T.D., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, Ed., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., & Du, H. (2022). *lavaan: Latent Variable Analysis* (0.6.11). [R package]. <https://cran.r-project.org/web/packages/lavaan/index.html>
- Rueda, F. J. M. (2013). *Bateria Psicológica para Avaliação da Atenção (BPA)*. 1ed. Vetor.
- Rueda, F. J. M., & Castro, N. R. de. (2013). Análise das Variáveis Idade e Sexo no Desempenho do Teste de Inteligência (TI). *Revista Psicologia: Teoria E Prática*, 15(2), 166–179. <https://editorarevistas.mackenzie.br/index.php/ptp/article/view/3829>
- Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, 16(5), 754–760. [doi:10.1017/S1355617710000706](https://doi.org/10.1017/S1355617710000706)
- Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis. A festschrift for Heinz Neudecker*, (pp. 233–247). London: Kluwer Academic Publishers.

- Schelini, P. W. (2006). Teoria das inteligências fluida e cristalizada: início e evolução. *Estudos de Psicologia (natal)*, 11(3), 323-332. <https://doi.org/10.1590/S1413-294X2006000300010>
- Schelini, P. W., & Wechsler, S. (2005). Bateria Multidimensional de Inteligência Infantil: desenvolvimento de instrumento. *Psico-USF*, 10(2), 129-139. <https://doi.org/10.1590/S1413-82712005000200004>
- Schelini, P. W., & Wechsler, S. (2006). Estudo da estrutura fatorial da Bateria Multidimensional de Inteligência Infantil. *Estudos de Psicologia*, 23(2), 105-112. <https://www.scielo.br/j/estpsi/a/Qh68cSnbZDf9ZMvHbTmFqDM/?lang=pt&format=pdf>
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53–61. <https://doi.org/10.1007/BF02289209>
- Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review*, 25(1), 12-27. <https://doi.org/10.1016/j.hrmr.2014.09.004>
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3ed., pp. 99–144). The Guilford Press.
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4ed., pp. 73–163). The Guilford Press.
- Sousa, L. A., & Braga, A. E. (2020). Teoria clássica dos testes e teoria de resposta ao item em avaliação educacional. *Revista de Instrumentos, Modelos e Políticas em Avaliação Educacional*, 1(1), e020002-e020002. 10.51281/impa.e020002.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- The jamovi project (2022). Jamovi. (Version 2.3) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- Tucker-Drob, E. M., Brandmaier, A. M., & Lindenberger, U. (2019). Coupled cognitive changes in adulthood: A meta-analysis. *Psychological Bulletin*, 145(3), 273–301. <https://doi.org/10.1037/bul0000179>
- Valentini, F., Gomes, C. M. A., Muniz, M., Mecca, T. P., Laros, J. A., & Andrade, J. M. (2015). Confiabilidade dos Índices Fatoriais da Wais-III Adaptada para a População

- Brasileira. *Psicologia - Teoria e Prática*, 17(2), 123–139.
http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872015000200010
- van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J. C., Colom, R., & Boomsma, D. I. (2006). Sex differences on the Dutch WAIS-III. *Intelligence*, 34(3), 273–289.
<https://doi.org/10.1016/j.intell.2005.08.002>
- Vetter, T. R. (2017). Fundamentals of Research Data and Variables: The Devil Is in the Details. *Anesthesia & analgesia*, 125(4), 1375-1380.
 doi:10.1213/ANE.00000000000002370
- Wasserman, J. D. (2018). A History of Intelligence Assessment: The Unfinished Tapestry. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4ed., pp. 901–911). The Guilford Press.
- Woodcock, R. W., Maricle, D. E., Miller, D. C., & McGill, R. J. (2018). Functional Cattell-Horn-Carroll Nomenclature for Practical Applications. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4ed., pp. 3–55). The Guilford Press.
- Zanin, D. S., Oliveira, K. da S., Oliveira, K. L. de, & Henklain, M. H. O. (2022). Desafios da Avaliação Psicológica no Brasil: Nova Realidade, Velhas Questões. *Avaliação Psicológica*, 21(4), 407-417. <https://dx.doi.org/10.15689/ap.2022.2104.24162>

Apêndice A – Aprovação do Comitê de ética em Pesquisa

UNIVERSIDADE FEDERAL DE
MINAS GERAIS**PARECER CONSUBSTANCIADO DO CEP****DADOS DO PROJETO DE PESQUISA****Título da Pesquisa:** Atualização das evidências de validade e das normas para testes psicológicos**Pesquisador:** ELIZABETH DO NASCIMENTO**Área Temática:****Versão:** 2**CAAE:** 08045918.9.0000.5149**Instituição Proponente:** UNIVERSIDADE FEDERAL DE MINAS GERAIS**Patrocinador Principal:** Financiamento Próprio**DADOS DO PARECER****Número do Parecer:** 3.358.718**Apresentação do Projeto:**

Este projeto pretende avaliar a validade de uma série de testes psicológicos visando a recomendação que esta validade seja testado a cada 15 anos. Segunda a pesquisadora: "esse projeto de pesquisa contempla atualmente a atualização dos testes Bateria Testes de Seleção Profissional (TSP), Medida de Prontidão Mental (MPM) e Destreza. Subprojetos futuros com o mesmo objetivo e procedimentos poderão ser integrados a esse protocolo de pesquisa. A investigação contemplará participantes adultos, de ambos os sexos e com diferentes níveis de instrução, residentes nos estados de Minas Gerais e Espírito Santo (500 participantes). Bancos de dados de respondentes de outras regiões do país também serão incorporados. Além de testes psicológicos, será aplicado um questionário sociodemográfico. Os dados serão analisados do ponto de vista psicométrico, estando previstas: investigação das propriedades dos itens (dificuldade e discriminação); levantamento de precisão por meio do coeficiente de consistência interna e estabilidade temporal; levantamento de evidências de validade (estrutura interna, validade convergente e discriminante, associação com variáveis relacionadas) e; elaboração das normas em Percentil. Espera-se que os resultados dessa pesquisa permitam que a comunidade profissional em psicologia possa continuar a utilizar instrumentos para fins de avaliação psicológica nos diferentes contextos." O projeto visa uma metodologia de aplicação dos testes a uma amostra de 500 pessoas. Esta aplicação "está prevista uma sessão

Continuação do Parecer: 3.358.718

com duração aproximada de 120 minutos para aplicação dos seguintes instrumentos: questionário pessoal e de saúde, Bateria TSP, MPM e Destreza. Para a subamostra que participará do estudo de levantamento de evidências de validade com construtos

relacionados, haverá uma segunda sessão de coleta de dados em que serão aplicados os testes Matrizes Progressivas de Raven - Escala Geral e TEACO-FF. Essa segunda sessão terá duração prevista de 60 minutos. Uma segunda subamostra será convidada para responder novamente aos testes Bateria TSP, MPM e Destreza para fins do estudo de estabilidade temporal. O intervalo entre a situação de testagem e retestagem nessa subamostra será de aproximadamente 15 dias." O critério de inclusão segundo o projeto são pessoas com escolaridade mínima de ensino médio em curso.

Objetivo da Pesquisa:

O objetivo principal do projeto segundo a autora é: "Atualizar as evidências de validade e as normas de interpretação dos escores de testes psicológicos."

Segundo a pesquisadora a pesquisa tem como objetivos secundários, "a) Analisar os parâmetros psicométricos dos itens; b) Estimar a consistência interna; c) Estimar a estabilidade temporal dos resultados; d) Levantar evidências sobre a estrutura interna; e) Levantar evidências de validade convergente e discriminante; f) Levantar evidências de validade com base em variáveis e construtos associados; g) Elaborar normas de interpretação dos escores."

Avaliação dos Riscos e Benefícios:

No resumo do projeto, a pesquisadora aponta que "a presente pesquisa oferece risco mínimo aos participantes voluntários. O risco pode estar relacionado ao cansaço decorrente da situação de testagem. Quando necessária, a aplicação poderá ser interrompida a qualquer momento. A pesquisa não prevê interrupções até que se obtenham os dados e os resultados que respondam aos objetivos. Não há risco biológico envolvido. A participação na pesquisa será interrompida caso o examinando demonstre cansaço, desconforto psicológico ou recusa em continuar as atividades por qualquer outro motivo." No TCLE e TALE apresentados cita os mesmos riscos e medidas a serem tomadas para amenizá-los.

Segundo a pesquisadora a pesquisa, o benefício da pesquisa é a: "A presente pesquisa está inserida na proposta de se desenvolver estudos que contribuam para o desenvolvimento da avaliação psicológica no Brasil, por meio da atualização das características psicométricas dos testes psicológicos. Os resultados desse estudo beneficiarão profissionais em situações de aplicação da psicologia que necessitem usar instrumentos psicológicos com qualidades psicométricas comprovadas. Beneficiam também os potenciais examinandos em situações de

Continuação do Parecer: 3.358.718

testagem psicológica, uma vez que os resultados alcançados por eles nos testes serão comparados com os de outros participantes com características sociodemográficas semelhantes, favorecendo a interpretações justas.”

Comentários e Considerações sobre a Pesquisa:

Pesquisa relevante para a área de Psicologia com previsão de término em 20/12/2019.

As solicitações do COEP foram atendidas: 1) Foi incluído no resumo de pesquisa a descrição mais detalhada da metodologia; 2) Foi modificado o TALE para ser um convite de participação diretamente para os participantes menores de 18 anos; 3) Foi acrescentado o TCLE para participantes maiores de idade; 4) Foi acrescentado aos TCLE/TALEs informações respectivas às outras etapas da pesquisa (a aplicação dos testes às subamostras).

Considerações sobre os Termos de apresentação obrigatória:

Foram apresentados os seguintes termos: folha de rosto, projeto completo, resumo do projeto, TCLE para os pais ou responsáveis dos participantes e TALE para os participantes, TCLE para os participantes maiores de 18 anos, modelo de carta de anuência, e parecer consubstanciado.

Recomendações:

Recomenda-se a aprovação do projeto de pesquisa. Por gentileza adicionar ao TALE e aos TCLEs texto que esclarece a forma de retorno dos resultados dos testes para os participantes.

Conclusões ou Pendências e Lista de Inadequações:

Somos favoráveis à aprovação do projeto "Atualização das evidências de validade e das normas para testes psicológicos" da pesquisadora responsável Profa. Dra. ELIZABETH DO NASCIMENTO.

Considerações Finais a critério do CEP:

Aprovado conforme parecer.

Tendo em vista a legislação vigente (Resolução CNS 466/12), o COEP-UFMG recomenda aos Pesquisadores: comunicar toda e qualquer alteração do projeto e do termo de consentimento via emenda na Plataforma Brasil, informar imediatamente qualquer evento adverso ocorrido durante o desenvolvimento da pesquisa (via documental encaminhada em papel), apresentar na forma de notificação relatórios parciais do andamento do mesmo a cada 06 (seis) meses e ao término da pesquisa encaminhar a este Comitê um sumário dos resultados do projeto (relatório final).

UNIVERSIDADE FEDERAL DE
MINAS GERAIS



Continuação do Parecer: 3.358.718

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1257170.pdf	17/04/2019 21:56:01		Aceito
Outros	Carta_resposta.docx	17/04/2019 21:55:17	ELIZABETH DO NASCIMENTO	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLETALEadolescenteTCLEadulto.pdf	16/04/2019 17:25:22	ELIZABETH DO NASCIMENTO	Aceito
Declaração de Instituição e Infraestrutura	Carta_Anuencia.pdf	07/02/2019 14:44:22	ELIZABETH DO NASCIMENTO	Aceito
Folha de Rosto	Folha_de_Rosto_CONEP_assinada.pdf	04/12/2018 00:51:45	ELIZABETH DO NASCIMENTO	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_COEP.doc	21/11/2018 02:18:00	ELIZABETH DO NASCIMENTO	Aceito
Parecer Anterior	Parecer_PPG.pdf	21/11/2018 02:16:10	ELIZABETH DO NASCIMENTO	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

BELO HORIZONTE, 30 de Maio de 2019

Assinado por:
Eliane Cristina de Freitas Rocha
(Coordenador(a))

Apêndice B - Carta de autorização da Instituição

CARTA DE AUTORIZAÇÃO

Eu, _____, declaro para os devidos fins que estou de acordo com a realização da pesquisa intitulada "Atualização das evidências de validade e das normas para testes psicológicos", sob a coordenação da professora Elizabeth do Nascimento, nas dependências desta organização, uma vez que há viabilidade de sua execução. Para tanto, subscrevo-me.

Nome

Cargo

Local e data

Apêndice C – Questionário Geral

Levantamento de informações

Data: ____ / ____ / 20____

Este questionário é individual e confidencial. Por favor, responda a **todas** as questões, pois a ausência de uma resposta pode invalidar sua avaliação.

Parte 1. Informações sócio-demográficas

Data de hoje: ____ / ____ / ____ Data de Nascimento: ____ / ____ / ____

Sexo 0 Feminino 1 Masculino

Idade: ____ anos

Cidade em que nasceu: _____ Estado: _____

Cidade atual: _____ Estado: () MG () Outro

Situação conjugal

1 Solteiro 3 Viúvo

2 Casado ou união estável 4 Divorciado/separado

Escolaridade:

1 Ensino médio completo ou incompleto 3 Superior completo em _____

2 Superior incompleto em _____ 4 Pós-graduação

Qual sua ocupação atual? (Se aposentado ou desempregado, informe a sua última ocupação)

Parte 2. Informações sobre saúde

1. Você tem alguma dificuldade de visão? [1] Sim [2] Não

1.1. Em caso afirmativo, você faz uso de óculos ou lentes? [1] Sim [2] Não

2. Você tem alguma dificuldade no movimento dos braços e das mãos? [1] Sim [2] Não

3. Você tem alguma doença:

Psiquiátrica: [1] Sim / Qual? _____ [2] Não

Neurológica: [1] Sim / Qual? _____ [2] Não

4. Já fez cirurgia no cérebro? [1] Sim [2] Não

5. Já () teve ou () tem crise epilética?

6. Já ficou inconsciente por mais de cinco minutos? [1] Sim [2] Não

7. Já teve algum acidente vascular cerebral (AVC)? [1] Sim [2] Não

8. Já teve isquemia cerebral? [1] Sim [2] Não

Apêndice D – Termo de Consentimento Livre e Esclarecido 1 - Adultos

Termo de Consentimento Livre e Esclarecido Adultos

Título do projeto de pesquisa: Atualização das evidências de validade e das normas para testes psicológicos

Prezado (a) participante,

este é um convite para você participar, voluntariamente, em uma pesquisa que se destina a investigar as qualidades técnicas de instrumentos psicológicos. Você pode escolher se quer participar ou não. Os pesquisadores vão lhe explicar como é a pesquisa, mas se você ainda tiver dúvidas, você pode procurar a pesquisadora responsável. Você tem toda liberdade para interromper sua participação na pesquisa, sem sofrer qualquer penalidade. Sua participação não envolverá despesas a você. A atividade que você irá fazer será responder a diferentes tarefas sobre memória, atenção, habilidade aritmética, fluência verbal e raciocínio. Também será apresentado um conjunto de perguntas com a finalidade de caracterizar os participantes da pesquisa em termos de idade, sexo, ocupação, local de residência e condições de saúde. A previsão de duração da tarefa é de aproximadamente 2 horas, o que poderá deixá-lo um pouco cansado, mas não irá lhe prejudicar, pois não há riscos para sua saúde. Os resultados da pesquisa serão utilizados em trabalhos científicos, publicados ou apresentados oralmente em eventos científicos, mantendo sua identidade em sigilo. Desta forma, você está ciente de que o seu nome não será mencionado em nenhuma circunstância, mantendo estrita confidencialidade e anonimato. Os dados coletados durante a realização das tarefas serão utilizados apenas para fins de pesquisa. O aceite para participar é fornecido por meio de sua assinatura em duas vias do presente Termo de Consentimento, sendo uma via para o pesquisador e outra via para o próprio participante. Em caso de dúvidas sobre objetivos específicos das tarefas realizadas, andamento das sessões, queixas ou comentários, você poderá entrar em contato com a pesquisadora Kellyane Madureira Figueiredo por e-mail kemfigueiredo@gmail.com e/ou telefone (31)98805-8225. O COEP-UFMG deverá ser consultado em caso de dúvidas relacionadas a questões éticas, no telefone (31)3409-4592 e/ou e-mail coep@prpq.ufmg.br. Estamos à disposição para quaisquer esclarecimentos e agradecemos a sua colaboração.

DECLARO QUE COMPREENDI OS OBJETIVOS E PROCEDIMENTOS DO ESTUDO ACIMA INFORMADO E, NÃO POSSUINDO DÚVIDAS ADICIONAIS E ESTANDO DE ACORDO COM A MINHA PARTICIPAÇÃO NELE, ACEITO PARTICIPAR. DECLARO QUE SEI QUE A QUALQUER MOMENTO POSSO REVOGAR ESTE ACEITE, SEM A NECESSIDADE DE PRESTAR QUALQUER JUSTIFICATIVA OU INFORMAÇÃO ADICIONAL. ESTOU CIENTE TAMBÉM DE QUE CASO NECESSITE, POSSO FAZER CONTATO COM AS PESQUISADORAS A QUALQUER INSTANTE.

As informações sobre e-mail e telefone deverão ser preenchidas para que tenha a devolução dos resultados desta pesquisa.

NOME

TELEFONE ()

EMAIL

ASSINATURA DO VOLUNTÁRIO: _____

Profa. Dra. Elizabeth do Nascimento Pesquisadora responsável pelo projeto Endereço: Av. Pres. Antônio Carlos, 6627 - Campus Pampulha CEP: 31270-901 / Belo Horizonte – MG
Telefone: (31) 3409-6275 E-mail: bethdonascimento@gmail.com Pesquisadora do Laboratório de Avaliação e Intervenção na Saúde (LAVIS) Departamento de Psicologia –
UFMG

Apêndice E – Termo de Consentimento Livre e Esclarecido 2 – Adolescentes

Termo de Consentimento Livre e Esclarecido Responsáveis pelos adolescentes

Título do projeto de pesquisa: Atualização das evidências de validade e das normas para testes psicológicos

Prezado (a) pais ou responsáveis,

este é um convite para você autorizar voluntariamente seu filho (ou menor de idade sob a sua responsabilidade) a participar de uma pesquisa que se destina a investigar as qualidades técnicas de instrumentos psicológicos. Você tem toda liberdade para interromper a participação do adolescente na pesquisa, sem sofrer qualquer penalidade. A participação dele não envolverá despesas a você. A atividade que ele irá fazer será responder a diferentes tarefas sobre memória, atenção, habilidade aritmética, fluência verbal e raciocínio. Também será apresentado e deverá ser preenchido pelo adolescente um conjunto de perguntas com a finalidade de caracterizar os participantes da pesquisa em termos de idade, sexo, ocupação, local de residência e condições de saúde. A previsão de duração da tarefa é de aproximadamente 2 horas, o que poderá deixá-lo um pouco cansado, mas não irá prejudicá-lo, pois não há riscos para a saúde dele. Ele também poderá interromper a participação quando quiser, sem que acarrete qualquer tipo de prejuízo. Os resultados da pesquisa serão utilizados em trabalhos científicos, publicados ou apresentados oralmente em eventos científicos, mantendo a identidade em sigilo. Desta forma, você está ciente de que o seu nome e/ou o nome do seu filho não serão mencionados em nenhuma circunstância, mantendo estrita confidencialidade e anonimato. Os dados coletados durante a realização das tarefas serão utilizados apenas para fins de pesquisa. O aceite para participar é fornecido por meio de sua assinatura em duas vias do presente Termo de Consentimento, sendo uma via para o pesquisador e outra via para o próprio participante e do Termo de Assentimento Livre e Esclarecido (TALE) para Participação de Adolescentes em Pesquisa.

Em caso de dúvidas sobre objetivos específicos das tarefas realizadas, andamento das sessões, queixas ou comentários, você poderá entrar em contato com a pesquisadora Kellyane Madureira Figueiredo por e-mail kemfigueiredo@gmail.com e/ou telefone (31)98805-8225. O COEP-UFMG deverá ser consultado em caso de dúvidas relacionadas a questões éticas, no

telefone (31)34094592 e/ou e-mail coep@prpq.ufmg.br. Estamos à disposição para quaisquer esclarecimentos e agradecemos a sua colaboração.

DECLARO QUE COMPREENDI OS OBJETIVOS E PROCEDIMENTOS DO ESTUDO ACIMA INFORMADO E, NÃO POSSUINDO DÚVIDAS ADICIONAIS ESTOU DE ACORDO COM A PARTICIPAÇÃO DO MEU FILHO (MENOR SOB MINHA RESPONSABILIDADE) NA PEQUISA. DECLARO QUE SEI QUE A QUALQUER MOMENTO POSSO REVOGAR ESTE ACEITE, SEM A NECESSIDADE DE PRESTAR QUALQUER JUSTIFICATIVA OU INFORMAÇÃO ADICIONAL. ESTOU CIENTE TAMBÉM DE QUE CASO NECESSITE, POSSO FAZER CONTATO COM AS PESQUISADORAS A QUALQUER INSTANTE.

As informações sobre e-mail e telefone deverão ser preenchidas para que tenha a devolução dos resultados do seu filho em relação a esta pesquisa.

NOME

TELEFONE ()

EMAIL

ASSINATURA DO RESPONSÁVEL PELO ADOLESCENTE:

Profa. Dra. Elizabeth do Nascimento

Pesquisadora responsável pelo projeto Endereço: Av. Pres. Antônio Carlos, 6627 - Campus

Pampulha CEP: 31270-901 / Belo Horizonte – MG Telefone: (31) 3409-6275 E-mail:

bethdonascimento@gmail.com Pesquisadora do Laboratório de Avaliação e Intervenção na Saúde (LAVIS) Departamento de Psicologia – UFMG

Apêndice F – Termo de Assentimento Livre e Esclarecido (TALE)

Termo de Assentimento Livre e Esclarecido (TALE) para participação de adolescentes em pesquisa

Prezado (a) participante,

estamos conduzindo uma pesquisa com o nome Atualização das evidências de validade e das normas para testes psicológicos. Esta pesquisa tem como objetivos validar e normatizar um instrumento criado para avaliação de habilidades cognitivas, como atenção, memória, percepção visual, fluência verbal, entre outras. Para a condução da pesquisa, precisaremos aplicar alguns questionários em um grupo de adolescentes. Este é um convite para você participar de nossa pesquisa. A coleta de dados acontecerá coletivamente nas escolas e instituições parceiras, também poderá acontecer individualmente você não pertença à escola/instituição parceiras, porém tenha juntamente com seu responsável manifestado interesse. O tempo gasto para preenchimento de todos os instrumentos é cerca de 120 minutos. Todos os dados coletados terão sua confidencialidade e anonimato assegurados com base nas normas previstas pelo Conselho de Ética e Pesquisa. Os dados serão utilizados exclusivamente para pesquisa e somente terão acesso a eles os pesquisadores envolvidos. O nome do adolescente não será identificado em nenhuma publicação que resultar deste estudo. Ressalta-se que você tem total direito de não consentir ou mesmo se retirar em qualquer fase da pesquisa, sem que isso lhes acarrete prejuízo de qualquer natureza. Não será cobrada nenhuma taxa ou qualquer outro valor aos responsáveis que consentirem com a participação na pesquisa. Os riscos envolvidos na pesquisa são mínimos. O risco máximo é de um possível cansaço. Diante de qualquer sinal de cansaço ou a pedido, a aplicação será interrompida a qualquer momento sem qualquer ônus você. A assinatura do Termo de Consentimento Livre e Esclarecido pelos pais/responsáveis e a assinatura desse documento por você é condição para a participação na pesquisa. Caso autorize a participação na pesquisa, solicitamos a gentileza de preencher e assinar esse termo no campo abaixo.

Caso tenha dúvidas ou necessite obter outras informações, favor entrar em contato com Elizabeth do Nascimento (31-99195-3393) ou Kellyane Madureira Figueiredo (31-988058225). Para maiores esclarecimentos de caráter ético você também poderá entrar em contato com o Comitê de Ética e Pesquisa da UFMG no endereço Av. Antônio Carlos, 6627, Unidade Administrativa II, 2º andar, Campus Pampulha, ou pelo telefone (31) 3409-4592. Agradecemos sua atenção e valiosa colaboração, subscrevendo-nos,

Dra. Elizabeth do Nascimento
Prof. Associada do Departamento de Psicologia da
UFMG
bethdonascimento@gmail.com

M^a. Kellyane Madureira Figueiredo
Professora e coordenadora do curso de Psicologia da
Faculdade do Futuro (FAF), Manhuaçu-MG
kemf_int@yahoo.com.br

Eu, _____, declaro que entendi todas as informações contidas neste Termo de Assentimento Livre e Esclarecido e que recebi uma cópia do mesmo. Concordo com os procedimentos da pesquisa intitulada Atualização das evidências de validade e das normas para testes psicológicos e concordo com a participação dele.

(cidade) _____, _____ de _____ de _____

Assinatura