



Universidade de Brasília

Instituto de Psicologia

Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações

Construção de uma bateria brasileira de inteligência com base na teoria Cattell–Horn–Carroll

Victor Vasconcelos de Souza

Brasília – DF

Fevereiro de 2018

Universidade de Brasília

Instituto de Psicologia

Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações

Construção e evidências de validade de uma bateria brasileira de múltiplas habilidades com  
base na teoria Cattell–Horn–Carroll

Victor Vasconcelos de Souza

Dissertação de Mestrado apresentada ao Programa  
de Pós-Graduação em Psicologia Social, do  
Trabalho e das Organizações (PPG-PSTO/UnB)  
como requisito para obtenção do título de mestre.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Cristiane Faiad

Brasília – DF

Fevereiro de 2018

Construção e evidências de validade de uma bateria brasileira de múltiplas habilidades com  
base na teoria Cattell–Horn–Carroll

Dissertação apresentada e avaliada por banca examinadora constituída por:

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Cristiane Faiad (Presidente)

Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações  
Universidade de Brasília

---

Prof. Dr. Everson Cristiano de Abreu Meireles (Membro titular)

Universidade Federal do Recôncavo Baiano

---

Prof. Dr. Sérgio Eduardo Silva de Oliveira (Membro titular)

Programa de Pós-Graduação em Psicologia Clínica e Cultura

Universidade de Brasília

---

Prof. Dr. Jacob Arie Laros (Membro suplente)

Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações

Universidade de Brasília

*“A verdadeira medida de um homem é como ele trata  
aquele que não pode fazer-lhe nada de bom.”*

Samuel Johnson

## **Agradecimentos**

Agradeço primeiramente a minha família, meus pais, José Antônio e Edna, e meus irmãos, Vinícius e Isabelle, que apoiaram toda a minha trajetória acadêmica, de todas as formas possíveis, desde a inspiração para traça-la.

À minha orientadora, Cristiane Faiad, que me acolheu para orientação com base em características abstratas que observou e não por demonstração de habilidades, permitindo que todo este trabalho tenha sido concebido e realizado com base pura e simplesmente na paixão pela perseguição do conhecimento.

Ao Cebraspe, e à Girlene, bem como minhas colegas Renata Manuelly e Caroline, que ajudaram sempre quando foi necessário, e até mesmo quando não foi, oferecendo todo o apoio possível, de emocional a logístico durante todo o percurso.

Aos alunos do grupo de pesquisa, que deram feedback importante durante o processo de construção dos instrumentos apresentados, e ajudaram nos momentos de aplicação repletos de esforço físico e resiliência. Em especial ao Lucas, que contribuiu mais diretamente para a construção de um dos instrumentos.

À Ana Cláudia, sem a qual este trabalho teria sido realizado com muito menos prazer e determinação. À Laura, ao Víthor e à Mayara, e tantos outros amigos cuja paciência para ouvir possibilitou que eu seguisse mesmo nos momentos mais difíceis.

Agradeço aos meus colegas de pós-graduação, Carlos, Wladimir e Tatiana, com os quais nem sempre foram compartilhados os bons momentos, mas definitivamente foram compartilhados os maus, tornando-os mais leves.

Por fim, agradeço aos professores, servidores e estagiários do PSTO, bem como aos professores Sérgio Oliveira, Everson Meirelles e Jacob Laros por aceitarem fazer parte da banca.

## Índice Geral

Agradecimentos.....	5
Índice Geral.....	6
Lista de Tabelas.....	9
Lista de Figuras .....	11
Resumo Geral.....	12
Abstract .....	13
Apresentação .....	14
Uma análise crítica da produção dos últimos dez anos da literatura brasileira de avaliação da inteligência .....	17
Resumo.....	18
Abstract .....	19
As quatro ondas.....	21
O estado da arte da avaliação psicométrica de inteligência e estudos internacionais .....	22
Método .....	25
Resultados e Discussão .....	28
Considerações finais.....	33
Referências.....	35
Construção e evidências de validade da Bateria de Múltiplas Habilidades com base na teoria CHC .....	39
Resumo.....	40
Abstract .....	41
O processo de desenvolvimento das teorias de inteligência .....	42

A teoria Cattell–Horn–Carroll .....	45
A prática de avaliação psicológica no Brasil .....	47
Método .....	49
Seleção das habilidades.....	49
Construção dos testes da bateria .....	50
Estudo de evidências de validade.....	52
Procedimentos .....	53
Resultados .....	54
Informação Geral Verbal (K0).....	55
Raciocínio Geral Sequencial (RGS) .....	56
Memória Visual (MV).....	58
Velocidade Perceptiva (VP).....	59
Visualização (Vz).....	60
Discussão .....	62
Referências.....	64
Evidências de validade com base na relação teste-critério de uma bateria brasileira baseada na teoria CHC.....	
Resumo.....	70
Abstract .....	71
Método .....	76
Participantes .....	76
Instrumentos .....	76
Procedimentos .....	78
Resultados .....	79

Matriz de correlações .....	80
Ajuste do modelo de equações estruturais .....	81
Discussão .....	84
Referências.....	87
Considerações Finais.....	91
Lista Geral de Referências .....	94

## Lista de Tabelas

### Manuscrito 1

Tabela 1. Critérios possíveis para classificação dos artigos pesquisados.	27
Tabela 2. Frequência e porcentagem dos artigos por classificação de artigo.	28
Tabela 3. Frequência e porcentagem dos objetivos dos artigos pesquisados.	29
Tabela 4. Frequência e porcentagem da concepção teórica de inteligência utilizada nos artigos.	31
Tabela 5. Frequência e porcentagem dos testes de inteligência utilizados nos artigos.	31
Tabela 6. Periódicos de origem e suas classificações na avaliação de periódicos da CAPES de 2015.	33

### Manuscrito 2

Tabela 1. Testes psicológicos de raciocínios e inteligência registrados no SATEPSI, segundo Reppold et al. (2017).	48
Tabela 2. Parâmetros de dificuldade e discriminação do teste K0, em ordem decrescente de discriminação.	56
Tabela 3. Parâmetros de dificuldade e discriminação do teste RGS, em ordem decrescente de discriminação.	57
Tabela 4. Parâmetros de dificuldade e discriminação do teste MV, em ordem decrescente de discriminação.	59
Tabela 5. Parâmetros de dificuldade e discriminação dos itens do teste VZ, em ordem decrescente de discriminação.	61

**Manuscrito 3**

Tabela 1. Habilidades amplas da teoria Cattell–Horn–Carroll. Uma tabela completa, com todas as habilidades específicas, pode ser encontrada em Schneider & McGrew (2013).	72
Tabela 2. Relações esperadas na matriz de correlações.	79
Tabela 3. Correlação entre os testes de habilidades da BHM, BPR-5, AA e TMR.	80
Tabela 4. Valores das medidas de encaixe dos modelos confirmatórios especificados para análise.	82

## Lista de Figuras

### Manuscrito 1

Figura 1. Critérios de inclusão e de exclusão para o estudo de revisão bibliográfica. 26

### Manuscrito 2

Figura 1. Organização das habilidades amplas da teoria CHC conforme grupamento funcional e conceitual (adaptado de Schneider e McGrew, 2012, p. 135). 47

Figura 2. Modelo de construção da bateria de testes. 50

Figura 3. Exemplo de item do teste K0. 55

Figura 4. Exemplo de item do teste RGS. 57

Figura 5. Exemplos de estímulos visuais para o teste de memória visual. 58

Figura 6. Exemplo de item do teste de visualização. 61

### Manuscrito 3

Figura 1. Representação gráfica do modelo utilizado na análise. 81

Figura 2. Coeficientes da análise por equações estruturais do modelo 1. 82

Figura 3. Coeficientes da análise por equações estruturais do modelo 2. 83

## Resumo Geral

O objetivo geral desta dissertação foi a construção de uma bateria de testes de inteligência com base na teoria Cattell-Horn-Carroll. No estudo 1, foi realizada uma avaliação crítica da literatura brasileira a partir de um modelo teórico de maturação da literatura proposto por Kamphaus. Foram analisados 91 artigos publicados em periódicos brasileiros. Uma quantidade significativa (22,9%) utilizaram testes medindo apenas o fator geral de inteligência, apesar de a teoria CHC ter sido a mais utilizada (28,9%). Também foi frequente a utilização de testes sem o uso destas interpretações, como o WISC-III e o BPR-5 (ambos com 18,5% cada). No geral, os artigos encontrados não discutiram seus achados à luz desta teoria de destaque, não alcançando completamente as tendências mais recentes de produção segundo Kamphaus. Parte do motivo pode ser a pequena quantidade de testes que permitem a interpretação com base nas habilidades previstas no modelo CHC. Neste sentido, no estudo 2, uma bateria com cinco testes de inteligência (Informação Geral Verbal, Raciocínio Geral Sequencial, Memória Visual, Velocidade Perceptiva e Visualização) foi construída com base em habilidades desta teoria. Todos os testes foram aprovados por especialistas após pequenas alterações. Os testes foram aplicados em 104 agentes de segurança pública, sendo que o teste Visualização também foi aplicado em 38 estudantes de universidade pública. Os escores foram analisados com base na teoria de resposta ao item com dois parâmetros, e revelaram dificuldade média de -0,64, e discriminação média de 1,32. No entanto, ainda não havia evidências de validade com base na relação com variáveis externas. No estudo 3, foram analisados os escores da bateria aqui construída juntamente com os escores dos testes RA, RV e RE da BPR-5, o teste de Atenção Alternada do BPA e o TMR. As relações hipotetizadas entre os testes foram confirmadas na matriz de correlações em magnitude e direção. Por fim, um modelo de equações estruturais foi construído com base no modelo teórico da CHC, e obteve coeficientes adequados.

*Palavras-chave:* construção de teste, teoria CHC, avaliação de inteligência.

### **Abstract**

The general objective of this dissertation was the construction of a battery of intelligence tests based on the Cattell-Horn-Carroll theory. In the first study, the Brazilian literature was critically reviewed based on a theoretical model proposed by Kamphaus. We analyzed 91 articles published on Brazilian periodicals. Although the most widely used theory was CHC (28,9%), a significant amount, 22,9%, used tests measuring only the general intelligence factor. In any case, the use of tests without a CHC interpretation, such as WISC-III and BPR-5, was frequent (each accounted for 18,5% of the tests used). In general, the articles found did not discuss their findings in light of this prominent theory, failing to fulfill the most recent tendencies according to Kamphaus. Part of the reason might have been the small number of instruments based on CHC theory with validity evidence for Brazilian samples. In this sense, in study 2, a battery of five intelligence tests (General Verbal Information, General Sequential Reasoning, Visual Memory, Perceptual Speed and Visualization) was built based on the abilities from this theory. All tests were approved by specialists after small modifications. The tests were administered to 104 public security agents. The scores were analyzed using a 2-parameter item response theory model and revealed overall mean difficulty -0.64 and mean discrimination 1.32. However, there hadn't been analyzed validity evidence based on the relationship with external variables. In study 3, the scores were analyzed along with the RA, RV, and RE tests of BPR-5, as well as the Alternating Attention test from BPA and the visual memory test TMR. The hypothesized relationships were confirmed in the table of correlations, both in magnitude and direction. Finally, a structural equation model was built based on CHC theoretical model, and presented adequate coefficients.

*Keywords:* test construction, CHC theory, intellectual assessment.

## Apresentação

O campo da avaliação da inteligência é frequentemente alvo de ataques, mesmo de dentro da psicologia. Há considerável crítica de setores de tradição continental da psicologia do desenvolvimento que sugerem que a avaliação da inteligência não deva nem mesmo ser realizada. Os argumentos contra a avaliação utilizados nestes setores são essencialmente três: a) o escopo dos testes de inteligência é muito restrito; b) os escores são enviesados pela visão de que habilidades são consideradas parte da inteligência e; c) os escores não querem dizer nada acerca das habilidades importantes para a vida na sociedade contemporânea. Do ponto de vista da psicometria, esta crítica, na verdade, diz respeito à validade da testagem. Em outras palavras, é uma crítica sobre a interpretação dos escores obtidos pelos respondentes—escores estes refletem a qualidade dos testes utilizados. Se nós quisermos justificadamente continuar a realizar avaliação de inteligência, devemos produzir testes que sejam capazes de medir uma alta gama de habilidades, e que estejam inseridos em uma teoria substancial acerca da inteligência, ou seja, que tragam informações adicionais acerca do que os escores representam no contexto de desenvolvimento e das relações específicas entre as habilidades do indivíduo avaliado. Em outras palavras, a construção de um procedimento de avaliação da inteligência mais contextualizada depende da utilização de uma teoria mais ampla e informativa.

Historicamente, cada teoria de inteligência que se destacou surgiu em resposta às questões deixadas pela anterior. Desta forma, foi possível observar uma maturação destas teorias ao longo do tempo. As críticas realizadas a uma teoria, se pertinentes, levavam à incorporação de melhorias na próxima. Atualmente, a teoria que tem ganhado visibilidade na literatura internacional de avaliação de inteligência é a teoria Cattell–Horn–Carroll (CHC). Conforme o esperado, esta teoria surgiu em resposta a diversos problemas enfrentados pelas teorias que a antecederam. Por assimilar vantagens de diferentes modelos em uma única

teoria, os autores gozam de um relativo consenso quanto à qualidade da teoria. Este consenso é verificado, por exemplo, pelo desenvolvimento de baterias de testes de inteligência com base nesta teoria, como a Woodcock–Johnson, ou das propostas de interpretação de baterias já consagradas, como as de Wechsler, por meio desta teoria. A preocupação com a utilização de uma teoria substancial na interpretação de perfis de pontos fortes e fracos (*profiles of strengths and weaknesses*, PSW) estudados por meio destes novos testes, no entanto, é recente, mesmo na literatura internacional. Cabe perguntar, então, se a literatura de inteligência no Brasil já alcança este nível de rigor teórico.

A presente dissertação se divide em três manuscritos. O primeiro manuscrito se propõe responder a esta pergunta, ou seja, a avaliar o momento atual da literatura brasileira de avaliação da inteligência. Para isso, foi utilizado como parâmetro um modelo teórico proposto no capítulo inicial escrito por Kamphaus et al. (2012). Este modelo sugere que a história da avaliação da inteligência pode ser dividida em quatro ondas. Por meio de uma revisão sistemática da produção acerca da avaliação da inteligência em periódicos brasileiros, foi realizada uma classificação dos artigos com base no tipo de atividade realizada e a perspectiva teórica utilizada no estudo. Estes dados permitiram a classificação dos artigos conforme à tradição de uma das ondas propostas por Kamphaus. Apesar do impacto da teoria CHC nas discussões, grande parte dos instrumentos utilizados, como o *Wechsler Adult Intelligence Scale*, utilizam subescores atóricos, conforme crítica de Kaufman já desde 1979. Para que instrumentos mais contextualizados sejam utilizados, no entanto, é necessário que estejam disponíveis para uso no Brasil testes de inteligência com amplas evidências de validade em amostras brasileiras.

Neste contexto o segundo manuscrito apresenta a construção de cinco testes de habilidades, baseados na teoria CHC e que compõem o fator g de inteligência. A bateria de testes construída leva o nome de Bateria de Múltiplas Habilidades, uma vez que, na teoria

CHC, o fator g de inteligência é composto por diversas habilidades amplas. Para contextualizar a escolha da teoria CHC como base deste manuscrito, também é apresentada uma breve história das teorias psicométricas de inteligência em sua introdução.

O segundo manuscrito, além de apresentar o processo de construção da bateria, traz também evidências de validade acerca da precisão dos itens propostos. Estas evidências estão na forma da análise dos parâmetros dos itens da bateria por meio da Teoria de Resposta ao Item (TRI). Neste artigo, devido às características dos itens e da amostra, foi utilizado o modelo dicotômico da TRI de dois parâmetros, discriminação e dificuldade.

O terceiro manuscrito descreve a relação dos cinco testes construídos no segundo manuscrito, e que compõem a Bateria de Múltiplas Habilidades, e outros cinco testes: três da Bateria de Provas de Raciocínio, acerca de raciocínio espacial, raciocínio abstrato e raciocínio verbal, um da Bateria de Funções Mentais, o Teste de Memória Visual (TMR), e um instrumento da Bateria de Provas de Atenção (Alternada). A correlação de testes já conhecidos com testes que estão sendo produzidos é um método comum em estudos de validação. Neste artigo, duas perspectivas de validação são tomadas. Em uma, a partir da correlação, são descritos os testes que servem como critério e os demais testes em que deveria se observar correlações segundo a teoria CHC. Na outra, ocorre o encaixe de um modelo confirmatório que traz os escores dos diversos testes organizados na mesma hierarquia proposta pela teoria CHC.

Espera-se que, ao final dos três manuscritos, seja possível a formulação de um argumento de evidência de validade para a Bateria de Múltiplas Habilidades. Considerando o panorama atual de avaliação da inteligência no Brasil, há enormes benefícios na disponibilização de medidas que sejam capazes de cobrir os construtos, conforme são descritos na teoria mais atual: a CHC.

**MANUSCRITO 1:**

Uma análise crítica da produção dos últimos dez anos da literatura brasileira de avaliação da  
inteligência

A critical review of the Brazilian literature of the last ten years on intellectual assessment

### Resumo

A avaliação psicológica da inteligência é um campo que conta com uma teoria recente de alto impacto acadêmico, a teoria Cattell–Horn–Carroll (CHC). No entanto, não há discussão sobre as contribuições que a literatura brasileira faz neste contexto. Foi realizada uma avaliação da literatura brasileira a partir do modelo das quatro ondas de inteligência proposto por Kamphaus em 91 artigos publicados em periódicos brasileiros. Apesar de 22,9% utilizarem testes que se espelham apenas no fator geral de inteligência, a teoria mais usada (28,9%) foi a CHC. Mesmo assim, há grande utilização de testes sem o uso destas interpretações, como o WISC-III e o BPR-5 (ambos com 18,5% cada). A avaliação da inteligência no Brasil carece de artigos que debatam com base em uma teoria sólida as implicações dos achados das pesquisas com estes testes, que é característica da literatura atual de inteligência.

*Palavras-chave:* teoria Cattell–Horn–Carroll, avaliação da inteligência, quatro ondas.

### **Abstract**

The Cattell–Horn–Carroll theory (CHC) has recently taken prominence between factorial theories of intelligence. The international impact of this theory has been studied, both within academic and professional contexts of intellectual assessment. However, there has been no discussions about such an impact on Brazilian literature. We found 91 articles on Brazilian periodicals regarding intelligence assessment and evaluated them in consonance with Kamphaus’s four waves of intelligence assessment. Despite 22,9% articles having used a simplistic idea of general intelligence, the most widely used theory (28,9%) was CHC. However, the use of atheoretical tests without employing such interpretations, such as WISC-III and BPR-5 (both with 18,5%) was widespread. The area of intellectual assessment in Brazil did not show much concern with discussing the findings of their results in light of a solid theory, which is characteristic of the latest wave of intellectual assessment.

*Keywords:* Cattell–Horn–Carroll theory, intellectual assessment, four waves.

A avaliação psicológica da inteligência faz parte do histórico da psicometria, tendo sido alvo das primeiras investigações desta área (Kamphaus, Winsor, Rowe, & Kim, 2012; Pasquali, 2011), e mantendo-se relevante até os dias atuais com a maturação das técnicas de pesquisa (como neuroimagem; cf., Khundrakpan et al., 2017). Desde as investigações pioneiras, duas grandes questões têm permeado a discussão. A primeira é a existência ou não de um fator geral de inteligência, inicialmente proposto por Spearman com o fator *g* (Spearman, 1927). A segunda é a dimensionalidade da inteligência, cujo desenvolvimento teórico acompanhou o desenvolvimento de técnicas de definição de dimensionalidade de uma série de variáveis, como a análise fatorial (Schneider, & McGrew, 2012).

Considerando as variadas formas que a inteligência pode tomar, não é de surpreender que um grande número de teorias psicométricas para a inteligência tenha sido proposto ao longo do tempo (cf., Pasquali, 2011). Uma vez que cada teoria surgia em resposta às questões deixadas pela anterior, foi possível observar uma maturação destas teorias ao longo do tempo. Este desenvolvimento culminou numa teoria que hoje goza de relativo consenso, chamada Cattell–Horn–Carroll (CHC; Alfonso, Flanagan & Radwan, 2005).

Estabilizadas as preocupações com a hierarquização e a dimensionalidade da inteligência, surge um novo problema: a definição precisa deste construto. Kamphaus et al. (2012) divide a história da avaliação psicométrica da inteligência em quatro fases, chamadas ondas. A título de exemplo, na onda mais recente, a “quarta onda” da avaliação de inteligência, passam a dominar as tentativas de aplicar uma teoria mais substantiva aos testes de inteligência, que até então haviam se mantido relativamente inertes frente às discussões teóricas.

Mas será que a pesquisa recente em avaliação psicológica da inteligência, no Brasil, está acompanhando o progresso destas teorias? O objetivo deste manuscrito é avaliar e classificar o processo de maturação da prática de avaliação psicométrica da inteligência.

## As quatro ondas

Segundo Kamphaus et al. (2012), a história da interpretação dos testes de inteligência pode ser dividida em quatro ondas distintas, vinculadas a um processo de maturação da teoria e da medida. Na primeira, a avaliação da inteligência é limitada à quantificação em termos gerais. Neste sentido, não havia uma teoria formal acerca da relação entre os escores obtidos e as habilidades decorrentes disto. Neste primeiro momento, no entanto, a interpretação era feita por conveniência conforme fosse o objetivo da testagem, geralmente a classificação do indivíduo em categorias definidas com base em amplitudes de QI de forma arbitrária, indo de “idiota” (até então um termo médico, referente aos QIs de 0 a 24), a “precoce” (referente aos QIs de 175 para cima; Wechsler, 1944). No período a que se refere esta onda, Spearman (1927) desenvolveu sua teoria acerca do fator geral de inteligência, a qual propôs que a inteligência pode ser entendida como decorrente de um único fator. Esta teoria representou bem os escores que até então eram utilizados. Os escores finais eram a soma dos escores de todos os testes, implicitamente sugerindo que seriam resultado de um único construto psicológico.

Conforme a utilização dos testes psicológicos passou a se difundir, principalmente como método de avaliação e seleção em contextos escolares e organizacionais, começou a haver demanda para que estes testes revelassem informações que estivessem ligadas a habilidades específicas dos indivíduos. Assim, a segunda onda de avaliação da inteligência é caracterizada pela criação de perfis de pontos fortes ou fracos (*profiles of strengths and weaknesses*, PSW) por meio dos escores de tarefas específicas ou de grupos de tarefas dentro dos testes de inteligência. Neste período, Wechsler (1944) foi responsável por propor agrupamentos das tarefas de seus testes, que chamou de subescores. Nas versões mais primitivas, os subescores obtidos eram o QI Verbal e o QI de Desempenho. No entanto, com o avanço das técnicas de processamento estatístico, surgiram dúvidas quanto a validade

destes subescores. Kaufman (1979) sugere que a falta de teoria por trás dos testes de inteligência limita a utilidade dos perfis com base nestes subescores, dado que exige estes escores de uma interpretação que agregue conhecimento, de forma que seja possível utilizar o conhecimento já obtido como forma de embasar inferências ou intervenções.

Kaufman (1979) foi pioneiro nesta terceira onda, na qual os estudos predominantemente se preocuparam com a formação de perfis com base em escores psicométricos. Neste sentido, apesar do objetivo dos trabalhos que caracterizaram esta onda se manter na formação de perfis, os subescores que formavam estes perfis eram provenientes de fatores psicométricos com maiores evidências de validade e de fidedignidade.

Já os estudos da quarta onda de avaliação de inteligência, referente ao momento em que a literatura se encontra atualmente, são marcados pela busca da ligação entre a avaliação psicológica e a definição precisa da inteligência como construto. Em outras palavras, fazem um questionamento mais substanciado da teoria que precede ou sustenta a medida e prezam, agora com melhores ferramentas estatísticas, pela qualidade dessa medida. Schneider e McGrew (2012) sugerem, em um modelo paralelo da história da avaliação da inteligência, que a fase atual também tem como preocupação a utilização de estudos de neurociências. No entanto, esta sugestão não é tão diferente de Kamphaus et al. (2012), visto que estes estudos têm como objetivo último definir os limites e as dimensões do construto de inteligência com maior rigor teórico-conceitual.

### **O estado da arte da avaliação psicométrica de inteligência e estudos internacionais**

A teoria CHC tem tido bastante espaço nas publicações em avaliação da inteligência. Alfonso, Flanagan e Radwan (2005) relatam o impacto da teoria nas medições de habilidades cognitivas e na interpretação do desempenho dos indivíduos em avaliações. Os autores relatam uma melhora na medição de pelo menos cinco habilidades amplas (raciocínio fluido – *Gf*, memória de curto-prazo – *Gsm*, armazenamento e recuperação de longo-prazo – *Glr*,

raciocínio auditivo – *Ga* e velocidade de processamento – *Gs*) por influência da teoria.

Outros estudos demonstraram a utilização cada vez mais frequente da teoria no delineamento de estudos científicos (Loughman, Bowden, & D'Souza, 2014; Nettelbeck & Burns, 2010; Zajac, Burns, & Nettelbeck, 2012).

Mais recentemente, James, Jacobs e Roodenburg (2015) descreveram os impactos da teoria CHC na avaliação psicológica realizada na Austrália. Os autores reforçam o impacto que a teoria teve na literatura da avaliação de inteligência; no entanto, relatam uma lacuna entre a teoria e a prática no entendimento da inteligência, por parte dos psicólogos clínicos. Neste estudo, apenas metade dos psicólogos utilizavam a CHC na atuação prática, e apenas um quarto dos profissionais possuía grande familiaridade com esta teoria.

Além da facilidade introduzida pelo teste Woodcock–Johnson III (WJ-III), a realização de estudos que utilizam a teoria CHC foi facilitada por outros fatores. A versão mais recente de duas das baterias de inteligência mais utilizadas, WISC–IV e WAIS–IV, tiveram escores interpretados a luz de cinco habilidades amplas previstas na teoria CHC e obtiveram melhores encaixes (Kranzler, Floyd, Benson, Zabolski, & Thibodaux, 2016a; Weiss, Keith, Zhu & Chen, 2013). Outro estudo examinou a validade da utilização do modelo em outro teste amplamente utilizado na Europa, o *Kaufman Assessment Battery for Children*, segunda edição (Kaufman, 2004; McGill, 2015). Este estudo foi realizado com 2.025 crianças e adolescentes selecionados por amostragem aleatória, que tiveram seus escores dentro de dimensões da inteligência propostos pela teoria CHC calculados e aplicados à predição de conquista acadêmica.

Uma implicação prática na avaliação psicológica com propósito clínico, por exemplo, foi a utilização de abordagens de definição de PSWs em transtornos de aprendizagem (cf. Flanagan, Ortiz & Alfonso, 2007). Os perfis foram desenvolvidos a partir de baterias de testes de inteligência, para identificar com mais precisão os aspectos cognitivos do transtorno

de aprendizagem—seja ele na área quantitativa, de escrita ou de leitura, por exemplo. James et al. (2015) sugerem que este tipo de utilização é de grande importância, visto que tanto as amplas habilidades quanto as estreitas explicam variância de desempenho acadêmico não explicada pelo QI.

Este método foi questionado por Flanagan e Schneider (2016), devido à baixa fidedignidade dos subescores tanto de baterias como WISC, quanto do próprio WJ-III. No entanto, Kranzler et al. (2016a), em resposta a esta crítica, citou a proposta de Woodcock (1990) de que, frente ao enorme número mesmo de habilidades amplas—comparado ao que se praticava com a teoria *Gf–Gc*, apenas a utilização de várias baterias, numa avaliação transversal, seria capaz de avaliar suficientemente um grande número destas habilidades com qualquer confiabilidade. Esta proposta, nomeada de avaliação por baterias cruzadas (*cross battery assessment*, ou *XBA*), resultou na realização de estudos mais confiáveis que tinham como objetivo o diagnóstico de especificidades dos transtornos de aprendizagem, estudos que foram nomeados de PSW XBA (ex.: Kranzler et al., 2016b). Mesmo assim, estudos de apenas uma bateria continuam sendo realizados, e ainda há aceitação de seus resultados quando há cuidado para evitar os problemas levantados por Flanagan e Schneider (2013; c.f. Koriakin et al., 2017).

Ainda no campo da avaliação psicológica em sua utilização clínica, o desenvolvimento do método proposto por Woodcock (1990) é minuciado no livro *Essentials of Cross-Battery Assessment*, de Flanagan et al. (2007). Na clínica neuropsicológica, Jewsbury, Bowden e Dugg (2016) relatam a capacidade do modelo para abarcar preocupações na avaliação cognitiva, como a avaliação das funções do executivo central.

Já no campo da psicologia educacional, o modelo foi utilizado para entender os mediadores cognitivos da compreensão de leitura no desenvolvimento precoce (Decker, Fillmore, & Roberts, 2017). Os autores aplicaram o WJ-III *Normative Update* (versão do

teste WJ-III com normas atualizadas) a 835 crianças em diferentes estágios de desenvolvimento de leitura. Os resultados sugerem que diferentes habilidades específicas são mais influentes dependendo do estágio do desenvolvimento. Estes resultados explicaram inconsistências nos resultados de estudos anteriores que possuíam pequenas diferenças metodológicas (Floyd et al., 2007; Floyd et al. 2012; Hajovsky et al. 2014).

A vasta literatura que utiliza os fatores propostos na teoria CHC é largamente internacional. Tendo em vista a importância do alinhamento da teoria à utilização de testes psicológicos na ciência, o presente estudo teve como objetivo avaliar e classificar os esforços recentes nas publicações da área de avaliação de inteligência do Brasil, e diagnosticar se os esforços mais recentes têm acompanhado as tendências de publicação internacionais de acordo com o modelo teórico de Kamphaus et al. (2012).

## **Método**

Foi realizada uma série de buscas sistemáticas em dois sites que agregam artigos de psicologia em português, Pepsic (Periódicos Eletrônicos em Psicologia) e SciELO (*Scientific Electronic Library Online*). As buscas foram realizadas com a palavra-chave “inteligência”, acompanhada de “avaliação” ou “teste”. Esta busca foi permitida em todos os campos de indexagem: título, área, palavras-chave, etc. Uma outra expressão de busca foi utilizada para denotar o ano. Foram pesquisados todos os anos de 2006 a 2017. Artigos que não tivessem ligação com psicologia (por ex., utilizavam a inteligência apenas como uma variável num estudo médico) ou que não realizavam ou tratavam de nenhum tipo de avaliação da inteligência foram descartados. A Figura 1 apresenta os critérios utilizados para inclusão e exclusão dos manuscritos avaliados.

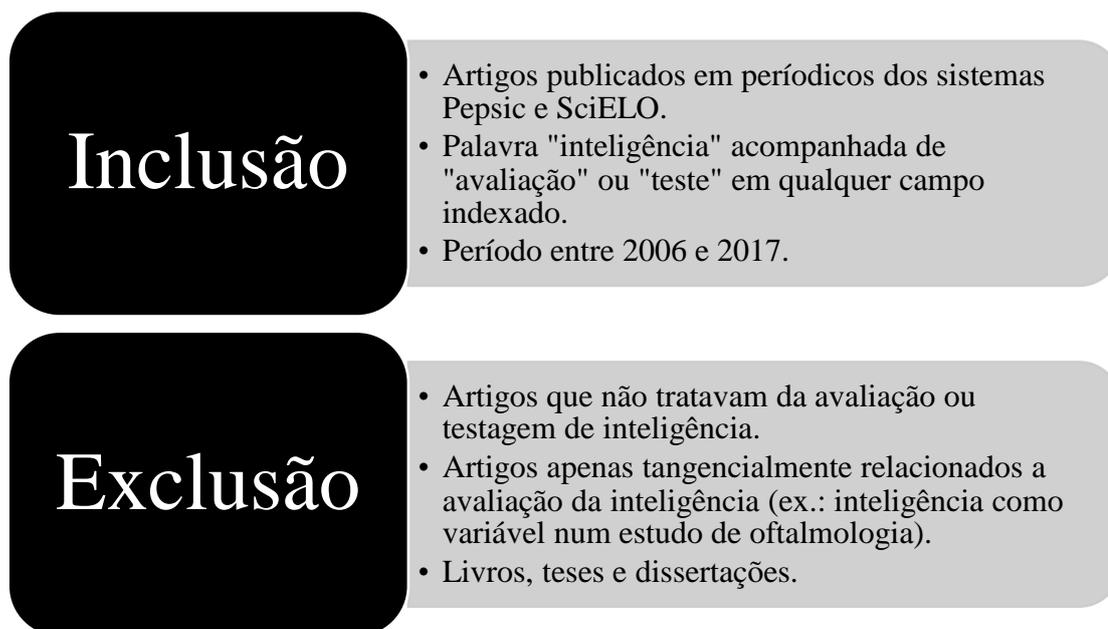


Figura 1.

*Critérios de inclusão e de exclusão para o estudo de revisão bibliográfica.*

Os dados foram analisados com base em classificação previamente proposta por Borges-Andrade e Pagotto (2010) para artigos de clima organizacional, tendo sido realizadas adaptações para cobrir as especificidades dos artigos publicados na área de psicometria. A planilha de classificação possibilitava a entrada de ano, título, revista, tipo de artigo, objetivo, testes utilizados e concepção teórica do artigo. Os critérios para as variáveis que os permitiam são descritos na Tabela 1.

Tabela 1.

*Critérios para classificação dos artigos pesquisados.*

Campo	Critérios Possíveis
Ano	2006–2017
Tipo de artigo	Pesquisa empírica, relato de experiência, resumo/revisão
Objetivo	Apresentação de evidências de validade e/ou fidedignidade Associação de variáveis Revisão teórica Perfil clínico Sugestão de método Normatização Adaptação de instrumento Construção de teste Crítica Exploração
Concepção teórica	Teorias psicométricas, neurocientíficas ou desenvolvimentistas de inteligência

Os artigos classificados como “pesquisa empírica” são aqueles que explicitavam relações entre variáveis como seu objetivo. Os artigos classificados como “relato de experiência” são aqueles que tratavam de experiências pontuais, utilizando amostras não-representativas e sem pretensão de estabelecer relações entre variáveis. Os artigos classificados como “resumo” ou “revisão” apenas agregavam informações e as sintetizavam ou tiravam conclusões de outros artigos, sem trazer novos dados.

Para “objetivo do artigo” foram considerados “apresentação de evidências de validade e fidedignidade”; “associação de variáveis”, que também tratava de diferenças entre grupos, correlações, etc.; “revisão teórica”; “perfil clínico”, para artigos que utilizavam subescores de testes de inteligência em participantes encaixados em categorias clínicas para informar sobre estas categorias; “sugestão de método”, para artigos metodológicos ou que traziam sugestões por meio de relato de experiência; “normatização”; “adaptação de instrumento”, para artigos que faziam adaptações transcultural de instrumento; “construção de teste”; “crítica”, para artigos que faziam alguma crítica de teoria ou metodologia; e “exploração” para artigos que

apenas relatavam alguma experiência, sem necessariamente definir variáveis, defender a utilização de método, etc.

Por fim, para “concepção teórica dos artigos”, foram consideradas na seguinte ordem de prioridade: a teoria explicitamente assumida pelo artigo; a implicitamente assumida pelo artigo com base nas análises e nos instrumentos; a concepção teórica dos instrumentos. Os artigos não foram classificados se não tratavam de assuntos que tinham a concepção teórica nem mesmo implícita nas análises realizadas.

### **Resultados e Discussão**

A pesquisa inicial retornou 143 artigos. Foram excluídos os artigos que não possuíam a avaliação da inteligência como objetivo principal ou, pelo menos, parte do método do artigo. Desta forma, restaram 91 artigos. Deste total, 75 (82,4%) artigos traziam relatos de pesquisa empírica estruturada. Um artigo (1,1%) era apenas relato de experiência de aplicação de testes. Quinze (16,4%), ou simplesmente visavam resumir o estado da área, ou revisavam a literatura e propunham alguma contribuição teórica, conforme apresentado na Tabela 2.

Tabela 2.

*Frequência e porcentagem dos artigos por classificação de artigo.*

Tipo de artigo (n = 91)	Frequência	Porcentagem
Pesquisa empírica	75	82,4
Resumo/revisão	15	16,5
Relato de experiência	1	1,1

Os objetivos mais específicos dos artigos foram divididos entre Apresentação de Evidências de Validade e Fidedignidade, com 42 artigos, 19 artigos de Associação de Variáveis, que tentavam associar variáveis entre sujeitos ou entre grupos, e 13 foram revisão teóricas. Doze artigos retornaram perfis clínicos de inteligência em diversos grupos. Quatro

fizeram revisões defendendo a utilização de uma teoria ou método, e dois relataram coletas de larga escala para Normatização de Teste. Os dados são apresentados na Tabela 3.

Tabela 3.

*Frequência e porcentagem dos objetivos dos artigos pesquisados.*

Objetivo (n = 91)	Frequência	Porcentagem
Apresentação de Evidências de Validade e Fidedignidade	41	45,0
Associação de Variáveis	19	20,9
Revisão/Teórico	13	14,3
Perfil Clínico	12	13,2
Metodologia	3	3,3
Normatização	2	2,2
Estudo de Caso	1	1,1

Os dados indicam que a produção brasileira tem sido dominada pela publicação de estudos de argumentação de validade de instrumentos, que compõem 45% de todos os estudos acerca de avaliação da inteligência publicados nos parâmetros desta bibliometria. A quantidade relativamente pequena de estudos com outros fins torna difícil a tarefa de classificar o momento pelo qual passa a literatura nacional de avaliação da inteligência, tendo-se como referência o modelo das ondas proposto por Kamphaus et al. (2012).

Quanto à concepção teórica dos artigos analisados, de forma a compará-los ao modelo de Kamphaus et al. (2012), foram consideradas aquelas trazidas nos manuais dos testes utilizados nos artigos. Caso o artigo explicitasse uma outra visão teórica, distinta do teste utilizada, optou-se por contabilizar apenas esta outra. Conforme apresentado na Tabela 4, a classificação de concepção teórica não se aplicou para 8 artigos, que apenas traziam resumos de outras análises. Nestes casos, não havia um posicionamento do autor sobre uma teoria que direcionasse suas análises. A teoria CHC foi o modelo mais utilizado, com 24 artigos. Em segundo lugar, 19 artigos utilizaram o fator g de Spearman, 14 artigos utilizaram a teoria de

três estratos. A teoria de inteligência emocional de Mayer e Salovey foi utilizada por 8 artigos. Cinco artigos utilizaram a concepção teórica de Cattell, a teoria *Gf-Gc*.

Na comparação dos referenciais teóricos com base na proposta de Kamphaus et al. (2012), dividida em quatro ondas, utilizou-se como critério de classificação a seguinte descrição. Na primeira onda, situada aproximadamente entre 1920 e 1940, a inteligência foi quantificada pelo uso de escores gerais. Schneider e McGrew (2012) sugerem que a utilização do modelo *g+* de Spearman–Holzinger foi mais proeminente entre meados de 1900 e 1940, período correspondente à primeira onda proposta. Já a segunda e a terceira onda foram identificadas pelo objetivo de criação de perfis e a utilização de teorias de inteligência multidimensionais. No entanto, a segunda baseou-se apenas nas divisões de subescores de Wechsler, enquanto a terceira associou-se às teorias multidimensionais provenientes de estudos psicométricos, como a teoria de três estratos de Carroll. Finalmente, quarta onda foi definida por teorias que trouxeram contribuições teóricas mais substanciais, como a teoria Cattell–Horn–Carroll. Assim, a partir da análise dos artigos, os dados indicaram uma predominância de 28,9% dos manuscritos na teoria CHC, embora em segundo lugar tenham sido identificados 22,9% dos manuscritos baseados na teoria de Spearman - que, adversamente, corresponde à primeira onda. Esse dado é um importante indicador da inconsistência presente na produção brasileira, que ora utiliza uma teoria de consenso internacional, ora utiliza uma base teórica provinda do início do século XX, embora a predominância de estudos esteja em conformidade com a tendência internacional de desenvolvimento da teoria CHC.

Tabela 4.

*Frequência e porcentagem da concepção teórica de inteligência utilizada nos artigos.*

Concepção Teórica de Inteligência dos Artigos (n = 83)	Frequência	Porcentagem
CHC	24	28,9
Fator g	19	22,9
Três estratos	14	16,9
Neurociências/Cognitiva	12	14,5
Gf-Gc	7	8,4
Mayer e Salovey	7	8,4

Com relação aos testes de inteligência utilizados, estão relatados na Tabela 5 apenas os testes que foram utilizados mais de uma vez. Além disso, um mesmo artigo pode ter utilizado mais de um teste. Desta forma, o número total (n = 53) não corresponde ao número de artigos que utilizaram instrumentos, mas o número de ocorrência dos mesmos. Houve predominância do uso do WISC-III e BPR-5, com 10 artigos cada, seguidos do MSCEIT, um teste de inteligência emocional, 5. O Desenho da Figura Humana, o único teste expressivo, foi utilizado como medida de desenvolvimento em 4 artigos. O Raven, WISC-IV, WAIS-III e o WJ-III também foram utilizados em 4 artigos.

Tabela 5.

*Frequência e porcentagem dos testes de inteligência utilizados nos artigos.*

Testes de Inteligência (n = 53)	Frequência	Porcentagem
WISC-III	10	18,9
BPR-5	10	18,9
MSCEIT	5	9,4
DFH	4	7,5
Raven	4	7,5
WISC-IV	4	7,5
WJ-III	4	7,5
WAIS-III	4	7,5
SON-R	3	5,7
WASI	3	5,7
R-2	2	3,8

A grande frequência de testes com uma perspectiva multidimensional de inteligência, como o WISC e o WAIS, o BPR-5 e o WJ-III, é consequência do fato de que grande parte dos estudos que não tinham como objetivo validação tiveram como objetivo a criação de perfis clínicos, que dependem da divisão da inteligência em diferentes escores para possibilitar a identificação das forças e das fraquezas dos indivíduos testados. Na segunda onda, o trabalho seminal de Rapaport, Gil e Schafer (1945-1946), baseado nos subescores da escala de Wechsler–Bellevue (Wechsler, 1939), foi um exemplo proeminente deste tipo de artigo de análise de perfis clínicos. Este pioneirismo é uma contribuição para o papel social da avaliação de inteligência na sociedade, mas ultimamente os esforços específicos realizados não foram baseados em alguma teoria sólida, de origem psicométrica e com boas evidências de validade. De fato, os subescores das escalas Wechsler são criticados por sua baixa fidedignidade até hoje, mesmo após revisões da escala (Valentini et al., 2015).

Na análise da literatura brasileira, no entanto, foi possível observar que os trabalhos de perfis clínicos frequentemente utilizaram testes como o WISC e WAIS em suas terceiras e quartas revisões. A terceira onda foi caracterizada pela análise de perfil sob a perspectiva psicométrica, e a técnica de análise fatorial passa a ter um maior impacto na interpretação dos escores dos testes de inteligência. Neste sentido, os testes utilizados já situam a produção brasileira em consonância com a terceira onda, uma vez que a terceira versão das escalas de Wechsler (1991) já possui uma possibilidade de interpretação proposta por Kaufman (1975), e a quarta versão já possui alterações no manual que permitam a interpretação com base na teoria CHC (Wechsler, 2004). Além destes testes, o segundo teste mais utilizado, o BPR-5, também é proveniente de uma interpretação teórica, a de Thurstone (1938).

Na literatura brasileira, 12 artigos do total se dedicaram a formação de perfis clínicos com base em testes. Desconsiderando os artigos associados a apresentação de evidências de validade e normatização, os artigos de perfil clínico representam 25% de todos os artigos

publicados dentro dos parâmetros desta revisão sistemática de literatura, sugerindo uma importante influência desta onda na produção brasileira analisada.

Dentre as revistas de origem dos artigos com mais de uma publicação relevante, apenas duas não são exclusivamente da psicologia e tratam de interfaces: Revista de Psicopedagogia e Revista Neuropsicologia Latinoamericana. As revistas com maior número de publicação foram Avaliação Psicológica (25%) e Psicologia: Teoria e Prática (13,10%), conforme apresentado na Tabela 6.

Tabela 6.

*Periódicos de origem e suas classificações na avaliação de periódicos da CAPES de 2016.*

Periódico (n = 91)	Qualis 2016	Frequência	Porcentagem
Avaliação Psicológica	A2	23	25,3
Psicologia: Teoria e Prática	A2	12	13,2
Boletim da Psicologia	B1	8	8,8
Rev. Psicopedagogia	B2	7	7,7
Temas em Psicologia	A2	7	7,7
Psic: revista da Vetor Editora	--	4	4,4
Psico-USF	A2	3	3,3
Psicologia: Ciência e Profissão	A2	3	3,3
Arquivos Brasileiros de Psicologia	A2	3	3,3
Psicologia Hospitalar	B2	2	2,2
Psicologia em Pesquisa	B1	2	2,2
Revista Neuropsicologia Latinoamericana	--	2	2,2

### **Considerações finais**

Com vista à evolução dos estudos realizados no Brasil, existem diversas dificuldades associadas a realização de pesquisa que se encaixaria na quarta onda, na perspectiva de Kamphaus et al. (2012), ou na “nova geração” proposta por Schneider e McGrew (2012), que inclui a utilização de métodos neuropsicológicos para a construção de teorias de maior clareza conceitual. Apesar da utilização de testes neuropsicológicos já ocorrer no Brasil, estudos internacionais mais robustos têm utilizado equipamentos demasiadamente caros, como máquinas de tomografia computadorizada e de ressonância magnética. Além disso,

restrições no campo do código de ética quanto as remunerações de participantes tornam estes estudos proibitivos, visto que demandam tempo e deslocamento dos participantes para laboratório especializados. De qualquer forma, as perspectivas são positivas, visto que 12 artigos utilizaram abordagens cognitivas, que é uma disciplina que compõe as neurociências, representando 14,5% de todos os artigos classificados. Por fim, sugere-se que a maturação da produção brasileira seja avançada pela realização de estudos mais robustos teoricamente, com maior discussão das consequências conceituais para o estudo da inteligência.

Apesar da clara importância da utilização de modelos mais modernos e robustos na literatura acadêmica, este trabalho não foi capaz de mensurar as consequências práticas do uso da teoria CHC fora da academia, no contexto geral de avaliação psicológica, tendo em vista o papel social que este tipo de avaliação carrega. O trabalho de James et al. (2015), que analisa a utilização de teoria na testagem psicológica da inteligência na Austrália, sugere que possa haver uma grande dissonância entre estes dois contextos de avaliação. A “nova geração” de estudos, proposta por Schneider e McGrew (2012) para a década de 2010, propõe a realização de trabalhos futuros deste gênero.

Apesar de a taxonomia proposta pela teoria CHC ser útil para o entendimento de um fenômeno complexo como a inteligência (Schneider & McGrew, 2012), pouco é de fato conhecido sobre a relação entre processos cognitivos e as tarefas mais complexas que são utilizadas em testes psicológicos. Mesmo as descrições mais extensivas das 16 habilidades amplas presentes no modelo CHC tomam apenas 20 páginas do texto de Schneider e McGrew (2012), no total. Com o objetivo de se alinhar às perspectivas internacionais de estudo, novos estudos em avaliação da inteligência deverão trazer uma perspectiva neuropsicológica para a teoria CHC. Os pesquisadores devem formular estudos que, mesmo que não tenham como objetivo primário o desenvolvimento da teoria, tragam reflexões acerca destas relações para

eventual construção de contribuições teóricas robustas da literatura brasileira para a literatura internacional de avaliação de inteligência.

### Referências

- Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The Impact of the Cattell–Horn–Carroll Theory on Test Development and Interpretation of Cognitive and Academic Abilities. In D. Flanagan, P. Harrison. (eds.) *Contemporary Intellectual Assessment, Second Edition: Theories, Tests, and Issues*. New York: Guilford Press.
- Becker, K.A (2003). History of the Stanford–Binet Intelligence scales: Content and psychometrics. *Stanford–Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No. 1*.
- Borges-Andrade, J. E., & Pagotto, C. P. (2010). O Estado da Arte da Pesquisa Brasileira em Psicologia do Trabalho e Organizacional. *Psicologia: Teoria e Pesquisa, 26*(especial), 37-50.
- Decker, S. L., Fillmore, P. T., & Roberts, A. M. (2017). Cognitive Mediators of Reading Comprehension in Early Development. *Contemporary School Psychology, 4*(1), 3–13. <https://doi.org/10.15540/nr.4.1.3>
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of Cross-Battery Assessment (2nd edition)*. Michigan: Wiley & Sons.
- Flanagan, D. P., & Schneider, W. J. (2016). Cross-Battery Assessment? XBA PSW? A case of mistaken identity: A commentary on Kranzler and colleagues “Classification agreement analysis of Cross-Battery Assessment in the identification of specific learning disorders in children and youth.” *International Journal of School and Educational Psychology, 4*(3), 137–145.

- Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell- Horn-Carroll cognitive abilities and their effects on reading decoding skills: *g* has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly*, *22*, 200–233.
- Floyd, R. G., Meisinger, E., Gregg, N., & Keith, T. (2012). An explanation of reading comprehension across development using models from Cattell-Horn-Carroll theory: support for integrative models of reading. *Psychology in the Schools*, *49*(8), 725–743.
- Hajovsky, D., Reynolds, M. R., Floyd, R. G., Turek, J. J., & Keith, T. Z. (2014). A multigroup investigation of latent cognitive abilities and reading achievement relations. *School Psychology Review*, *43*(4), 385–406.
- James, L., Jacobs, K. E., & Roodenburg, J. (2015). Adoption of the Cattell–Horn–Carroll model of cognitive abilities by Australian psychologists. *Australian Psychologist*, *50*(3), 194–202. <https://doi.org/10.1111/ap.12110>
- Jewsbury, P. A., Bowden, S. C., & Duff, K. (2016). The Cattell–Horn–Carroll Model of Cognition for Clinical Assessment. *Journal of Psychoeducational Assessment*, *1*(June), 1–21. <https://doi.org/10.1177/0734282916651360>
- Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2012). A History of Intelligence Test Interpretation. In D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues (3rd ed.)* (p. 99-144). New York: Guilford.
- Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman assessment battery for children (2nd ed.)*. Circle Pines, MN: American Guidance Service.
- Khundrakpan et al. (2017). Imaging structural covariance in the development of intelligence. *Neuroimage*, *144*, 227–240.
- Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016a). Classification agreement analysis of Cross-Battery Assessment in the identification of specific learning disorders in children and youth. *International Journal of School &*

*Educational Psychology*, 4(3), 124–136.

<https://doi.org/10.1080/21683603.2016.1155515>

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016b). Cross-Battery Assessment pattern of strengths and weaknesses approach to the identification of specific learning disorders: Evidence-based practice or pseudoscience? *International Journal of School & Educational Psychology*, 4(3), 146–157.

Koriakin, T., White, E., Breaux, K. C., DeBiase, E., O'Brien, R., Howell, M., ... Courville, T. (2017). Patterns of Cognitive Strengths and Weaknesses and Relationships to Math Errors. *Journal of Psychoeducational Assessment*, 35(1–2).

<https://doi.org/10.1177/0734282916669909>

Loughman, A., Bowden, S. C., & D'Souza, W. (2014). Cognitive functioning in idiopathic generalised epilepsies: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, 43, 20–34. doi:10.1016/j.neubiorev.2014.02.012

McGill, R. J. (2015). Interpretation of KABC-II Scores: An Evaluation of the Incremental Validity of Cattell-Horn-Carroll (CHC) Factor Scores in Predicting Achievement. *Psychological Assessment*, 27(4), 1417–1426. <https://doi.org/10.1037/pas0000127>

Nettelbeck, T., & Burns, N. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences*, 48, 379–384. doi:10.1016/j.paid.2009.10.032

Pasquali, L. (2011). *Psicometria: Teoria dos Testes na Psicologia e na Educação*. Petrólis: Vozes.

Rapaport, D., Gill, M. M., & Schafer, R. (1945–1946). *Diagnostic psychological testing (Vols. 1–2)*. Chicago: Yearbook Publishers.

- Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In, D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues (3rd ed.)* (p. 99-144). New York: Guilford.
- Spearman, C. (1927). *The abilities of man*. London: MacMillan & Co.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Valentini, F., Gomes, C. M. A., Muniz, M., Mecca, T. P., Laros, J. A., Andrade, J. M. de (2015). Confiabilidade dos índices fatoriais da Wais-III adaptada para a população brasileira. *Revista Psicologia: Teoria e Prática*, 17(2), 123–139.
- Wechsler, D. (1944). *The measurement of adult intelligence (3rd ed.)*. Baltimore: Williams & Wilkins.
- Wechsler, D. (1991). *The Wechsler intelligence scale for children—third edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). *The Wechsler intelligence scale for children—fourth edition*. London: Pearson.
- Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*, 31, 94–113. doi:10.1177/0734282913478030
- Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, 8, 231– 258.
- Zajac, I., Burns, N., & Nettelbeck, T. (2012). Do purpose-designed auditory tasks measure general speediness? *International Journal of Intelligence Science*, 2, 23–31. doi:10.4236/ijis.2012.22004

**MANUSCRITO 2:**

Construção e evidências de validade da Bateria de Múltiplas Habilidades com base na teoria

CHC

Construction and validity evidence of five psychological tests based on the CHC theory

### Resumo

Uma bateria com cinco testes de inteligência foi construída a partir de um levantamento acerca de possíveis tarefas e suas respectivas habilidades avaliadas com base na teoria Cattell–Horn–Carroll. As habilidades cobertas foram Informação Geral Verbal, Raciocínio Geral Sequencial, Memória Visual, Velocidade Perceptiva e Visualização. Os testes passaram por análises de especialistas e foram aplicados em 104 agentes de segurança pública, sendo o último teste aplicado também em 38 estudantes de universidade pública. Os escores foram analisados com base na teoria de resposta ao item com dois parâmetros, e revelaram qualidade adequada para a maior parte dos itens, que tiveram dificuldade média de -0,64, e discriminação média de 1,32. Estes resultados apontam inicialmente para adequação da bateria, que poderá vir a ser utilizada como um instrumento para avaliação de habilidades específicas, alinhado à uma teoria psicométrica de destaque. Futuras coleta para avaliação de evidências de validade concorrente são sugeridas.

*Palavras-chave:* construção de instrumento, inteligência, Cattell–Horn–Carroll, teoria de resposta ao item.

### **Abstract**

A battery composed of five intelligence tests was constructed based on a review of possible tasks and their respective abilities, as proposed by the Cattell–Horn–Carroll theory. The abilities covered were General Verbal Information, General Sequential Reasoning, Visual Memory, Perceptual Speed and Visualization. The tests were evaluated by subject matter experts and were then administered to 104 public security agents, with the Visualization test being applied to an additional 38 students. The scores were analyzed through item response theory, with a two-parameter model. The model revealed that the items were of adequate quality, with mean difficulty -0.64 and mean discrimination 1.32. These results indicate that the battery of tests could be used for clinical assessment and as an instrument of research of specific intelligence abilities from a robust psychometric intelligence theory after further validity evidence is collected.

*Keywords:* test construction, intelligence, Cattell–Horn–Carroll, item response theory.

Considerando as variadas perspectivas constitutivas pelas quais a inteligência pode ser estudada, não é de surpreender que um grande número de teorias psicométricas para a inteligência tenha sido proposto ao longo do tempo, sendo possível observar uma maturação nestas teorias (Almeida, Guisande, Primi, & Ferreira, 2008). Este processo culminou numa teoria que goza de relativo consenso hoje em dia, denominada Cattell–Horn–Carroll (CHC; Alfonso, Flanagan & Radwan, 2005, Aken, van der Heijden, Oomens, 2017).

A despeito do impacto da teoria CHC no estudo psicométrico da inteligência ser definitivamente reconhecido em diversos contextos de estudo, não existem instrumentos psicométricos de inteligência criados com base nesta teoria em utilização no Brasil. Este trabalho teve como objetivo o desenvolvimento e a coleta de evidências iniciais de validade de cinco testes de inteligência que formam a Bateria de Múltiplas Habilidades.

### **O processo de desenvolvimento das teorias de inteligência**

Apesar da primeira teoria de inteligência ser frequentemente atribuída a Spearman, Greenwood (2015) sugere que Herbert Spencer foi o primeiro a propor formalmente uma teoria científica substancial acerca da inteligência, em 1855. Segundo Spencer, a inteligência seria a “capacidade do indivíduo de responder a estímulos”. Ao final do século XIX, a linha científica psicométrica teve como precursor Francis Galton por meio do estudo de habilidades individuais. Galton estabeleceu uma cultura científica pela qual se entende que conhecer um construto depende de medi-lo (Silva, 2011). Este legado foi seguido por Cattell, que passou a investigar medidas de percepção e habilidades motoras (Kamphaus et al., 2012).

No desenvolvimento da psicometria como ciência moderna, ao início do século XX, pesquisadores da psicologia do desenvolvimento começaram a se interessar na avaliação das funções cognitivas com o objetivo de classificar crianças para séries escolares mais apropriadas (Pasquali, 2011). Este movimento fez parte do contexto sociohistórico da França, que foi pioneira na ideia de promover o ensino formal a todas as crianças. No entanto, as

crianças de mesma idade possuíam diferentes aptidões, algo que foi identificado como um problema pela sociedade. Neste contexto, Binet e Simon (1905) se dedicaram ao desenvolvimento do primeiro teste empírico e prático de inteligência para uso aplicado na classificação de estudantes. A primeira versão da escala de Binet–Simon foi altamente influente pelo pioneirismo e ampla utilidade.

Neste primeiro momento de testagem psicológica da inteligência, os escores retornados pelos instrumentos eram escores compostos. Isto significa que a inteligência era representada por um único número, calculado a partir da conjunção de todas as tarefas realizadas (Nicolas, Andrieu, Croizet, Sanitioso, & Burman, 2013). Esta prática revela uma tendência de consideração da inteligência por unidimensionalidade, ou seja, a ideia de que a inteligência era uma soma de habilidades, mas que como construto, deveria ser entendida como uníssona. Esta perspectiva da época foi confirmada na teoria de Spearman (1904) acerca do fator que este autor denominou *fator g*. O escore final de testes de inteligência foi utilizado ora com uma conotação médica, diagnosticando aqueles com “deficiência intelectual”, ora para seleção, de forma que a era dos testes de inteligência (1910–1930) viu os testes psicológicos de inteligência sendo utilizados para classificar recrutas na Primeira Guerra Mundial, conforme seus resultados nos testes *Army Alpha* e *Army Beta* (Yoakum, & Yerkes, 1920; Pasquali, 2011)

Apesar da utilidade de simplificar a inteligência como um escore único, outras teorias surgiram com o objetivo de diferenciar tipos de raciocínio. Schneider e McGrew (2012) apontam que, durante esta época, Spearman passou a realizar análises bifatoriais—análises que permitiam que a variância fosse explicada ao mesmo tempo por fatores em dois níveis hierárquicos diferentes—considerando fatores específicos além do fator *g*, na teoria denominada de Spearman–Holzinger (Spearman, 1939). No mesmo ano, Wechsler (1939) implementou em seu teste o cálculo de subescores, ou seja, escores que agrupavam

funcionalmente um conjunto menor de testes. Por meio destes escores que representavam a interpretação de Wechsler para diferentes funções dos subtestes, passou a ser possível a criação de perfis clínicos, que seriam capazes de realizar um diagnóstico mais específico. Em outras palavras, Rapaport et al. (1945-1946) sugeriram que a avaliação intrasujeito das diferenças dos subescores produz informações úteis para o diagnóstico e indicações psicoterapêuticas.

Wechsler (1949) popularizou o uso destes escores individuais de tarefas de Compreensão, Aritmética, Informação, Dígitos, Semelhança, entre outros. No entanto, a utilização destes escores carregou uma pressuposição adjacente: um entendimento multidimensional do construto de inteligência. Este entendimento passou a ser mais bem explorado a partir da maturação da técnica de análise fatorial e de sua utilização por Cohen (1959). Este autor estabeleceu uma estrutura de três fatores utilizada até hoje na versão mais nova do WAIS, o WAIS-IV.

Tanto a interpretação de Wechsler (1949) quanto os resultados do estudo de Cohen (1959) possuem uma limitação apontada em trabalho seminal de Kaufman (1979): a ausência de teoria. Esta ausência nega ao profissional a correta interpretabilidade e subsequente tratamento, uma vez que não há relação estabelecida entre os escores e o processo de desenvolvimento psicossocial de um dado paciente. Kaufman (1979) sugeriu, então, uma interpretação com base em clusters de subtestes baseados em divisões de teorias de inteligência já bem estabelecidas.

Thurstone (1938) protagonizou a teorização multifatorial da inteligência, propondo uma divisão altamente influente do construto em sete fatores, que denominou “habilidades mentais primárias”. Estas habilidades compreendiam tipos diferentes de raciocínio: compreensão verbal, fluência, facilidade numérica, visualização espacial, memória associativa, velocidade perceptiva e raciocínio. No entanto, ainda não havia consenso entre as

teorias de inteligência acerca do entendimento último da inteligência como possuindo ou não um fator geral (Pasquali, 2011).

Cattell (1943) foi responsável por formular uma das teorias mais influentes entre as que consideravam mais de um fator de inteligência, a teoria *Gf–Gc*. Na sua visão, a inteligência poderia ser entendida como uma relação complexa entre *Gf*, a inteligência fluida, e *Gc*, a inteligência cristalizada. A inteligência fluida seria responsável pela resolução criativa de problemas: isto é, o sujeito deveria criar, sem base em treinamentos específicos, uma resolução para um determinado problema. A inteligência cristalizada seria a capacidade do sujeito de resolver problemas com base em experiências passadas. Horn (Horn & Cattell, 1966), aluno de Cattell, com base em bancos de dados que se tornavam cada vez mais abrangentes, observou a necessidade de aumentar o poder explicativo da teoria, e expandiu esses dois fatores para nove. Esta teoria ficou conhecida como *Gf–Gc* estendida.

### **A teoria Cattell–Horn–Carroll**

A teoria Cattell–Horn–Carroll (CHC) é a mais recente das formas de entendimento fatorial da inteligência, tendo resultado da integração da teoria de três estratos de Carroll com a teoria *Gf–Gc* estendida (Horn, 1965; Horn, & Cattell, 1966). A teoria de três estratos de Carroll (1993) foi responsável por conciliar importantes contribuições de outros autores. Ao mesmo tempo admite o fator *g*, uma quantidade parcimoniosa de fatores intermediários, e a separação de habilidades específicas que não são, a rigor, redutíveis aos fatores intermediários (Schneider, & McGrew, 2012). O resultado desta integração, a teoria CHC, está hoje em sua terceira revisão, qualificada como “versão 2.2”. O modelo possui 16 tipos de raciocínio no segundo estrato, chamados habilidades amplas. Por sua organização em diferentes estratos, Schneider e McGrew (2012) argumentam que a teoria é uma taxonomia. Os autores defendem a importância da taxonomia, alegando que “um sistema útil de

classificação modela como nós vemos fenômenos complexos iluminando distinções consequenciais e obscurecendo diferenças triviais” (Schneider, & McGrew, 2012, p. 99).

Como argumento a favor da flexibilização da teoria com base no processamento de novos dados, McGrew (2007), na primeira revisão da teoria—a versão 2.0, recomenda que resistamos à tentação de cristalizar as categorias da CHC, e, de fato, já avaliava a adição de quatro outras habilidades às 10 da época (Schneider & McGrew, 2012). Estas características levaram a teoria ao sucesso, tendo alcançado bastante influência na avaliação psicológica da inteligência (Alfonso, Flanagan, & Radwan, 2005; Aken, van der Heijden, Oomens, 2017).

A forma de organização da teoria CHC se assemelha àquela proposta por Carroll (1993). A teoria possui um fator geral que é composto pelas 16 habilidades amplas: raciocínio fluido (*Gf*), memória de curto prazo (*Gsm*), armazenamento e recuperação de longo prazo (*Glr*), velocidade de processamento (*Gs*), velocidade psicomotora (*Gps*), conhecimento–compreensão (*Gc*), conhecimento específico de domínio (*Gkn*), leitura e escrita (*Grw*), conhecimento quantitativo (*Gq*), processamento visual (*Gv*), processamento auditório (*Ga*), habilidades olfatórias (*Go*), habilidades táteis (*Gh*), habilidades cenestésicas (*Gk*), habilidades psicomotoras (*Gp*). Estas habilidades foram ainda divididas em grupamentos funcionais e conceituais, conforme organização na Figura 1.

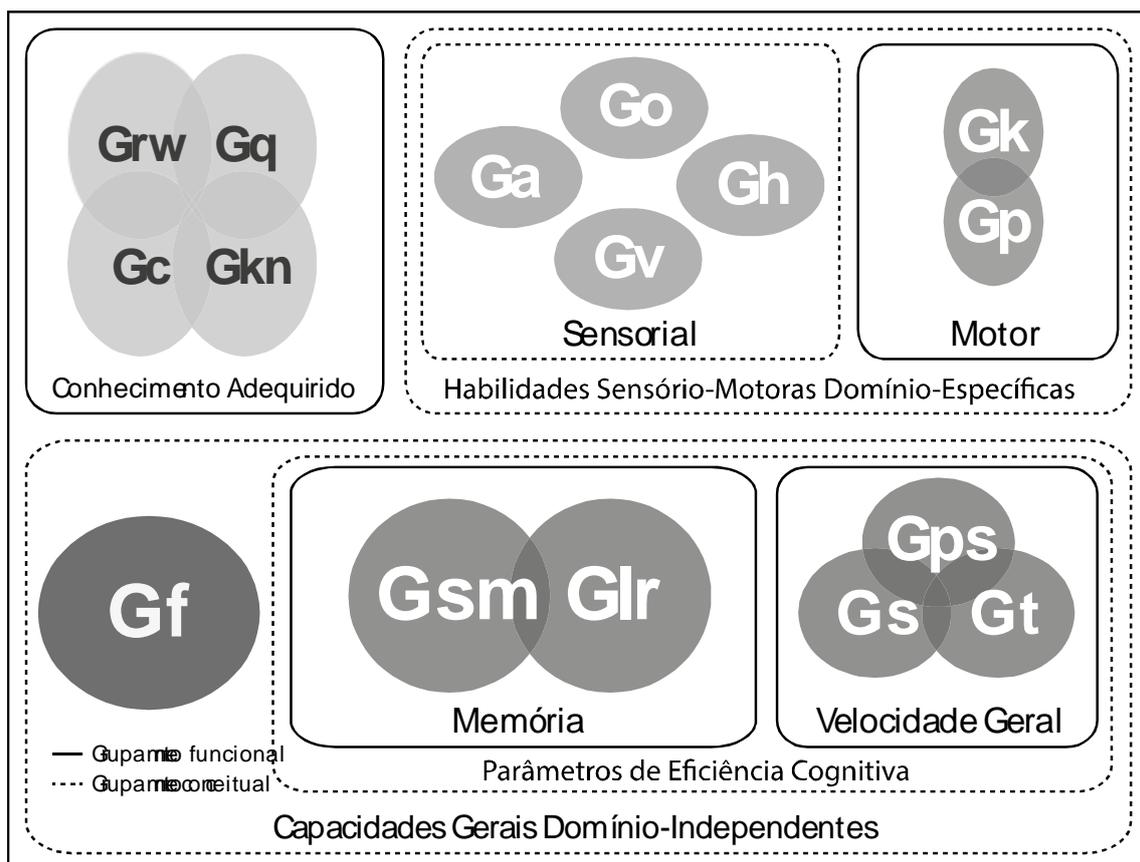


Figura 1.

*Organização das habilidades amplas da teoria CHC conforme agrupamento funcional e conceitual (adaptado de Schneider e McGrew, 2012, p. 135).*

Os maiores impactos desta teoria na prática brasileira estão presentes nos testes que são mais frequentemente utilizados, os testes de Wechsler (WISC e WAIS) e a Bateria de Provas de Raciocínio (Reppold et al., 2017; manuscrito 1, s.d.). Há estudos com o fim de relacionar ambos estes testes com a teoria CHC (Wechsler, 2003; Primi, Nakano & Wechsler, 2012), que podem ser então interpretados à luz desta teoria.

### **A prática de avaliação psicológica no Brasil**

No Brasil, a avaliação psicológica é atividade privativa do psicólogo, conforme a Lei 4.119 (1962). O Conselho Federal de Psicologia é responsável por definir as diretrizes da avaliação, sendo o uso de testes regulamentado pela Resolução CFP nº 002 (2003). Os testes psicológicos são avaliados por uma comissão *ad hoc* e têm seu parecer publicado no sistema

de avaliação de testes psicológicos, o SATEPSI e exigem do autor um grande investimento de pesquisa que atenda—pelo menos—aos critérios mínimos de qualidade para sua aprovação.

A atividade de construção ou adaptação de testes psicológicas tem um alto custo, exige um investimento de tempo e conhecimento especializado. Desta forma, não é de se surpreender que, dentre outras medidas, poucos testes de inteligência sejam produzidos todos os anos no Brasil. Reppold et al. (2017) fazem uma revisão dos manuais aprovados pelo SATEPSI para o ano de 2016. A tabela abaixo apresenta os testes classificados como raciocínio ou inteligência.

Tabela 1.

*Testes psicológicos de raciocínios e inteligência registrados no SATEPSI, segundo Reppold et al. (2017).*

Nome	Ano de aprovação
Matrizes Progressivas Avançadas de Raven	2014
Teste Não-Verbal de Inteligência SON-R 1/2-7	2011
Teste Não Verbal de Inteligência Geral - BETA-III	2011
Teste de Habilidade para Trabalho Mental (HTM)	2009
TONI-3: Teste de Inteligência Não Verbal	2006
Desenho da Figura Humana (DFH)	2005
Teste Não-Verbal de Raciocínio para Crianças (TNVRI)	2005
Teste Não Verbal de Inteligência (G36)	2003
Teste Verbal de Inteligência (V-47)	2003
Escala de Inteligência Wechsler para Crianças 3ª edição (WISC-3)	2002
Escala de Inteligência Wechsler para Crianças 4ª edição (WISC-4)	2002
Teste Não Verbal de Inteligência (G38)	2002
Teste Equicultural de Inteligência (TEI)	2002
Bateria Geral de Funções Mentais (BGFM)	2002
Escala de Maturidade Mental Colúmbia	2001
Teste Não-Verbal de Inteligência para Crianças (R-2)	2000
Bateria de Provas de Raciocínio (BPR-5)	2000

Em uma avaliação entre a relação entre os instrumentos aprovados e as teorias que os subsidiam, observa-se que o impacto da teoria CHC no entendimento da inteligência no

âmbito acadêmico não é garantia de impacto na prática profissional de avaliação psicológica no Brasil. A presença de testes psicológicos cadastrados no SATEPSI e capazes de avaliar dimensões desta teoria ainda é baixa. Dentre os testes apresentados na Tabela 1, nenhum foi construído com base no modelo. Há um artigo sobre a relação entre Bateria de Provas de Raciocínio e o Woodcock-Johnson III, um teste baseado no modelo CHC (Primi, Nakano & Wechsler, 2012). No entanto, os escores dos testes não se comportaram como esperado, tendo se saturado majoritariamente no fator de raciocínio fluido. Além disso, não há evidências de validade para o teste Woodcock-Johnson III em amostras brasileiras, comprometendo as conclusões tiradas com base neste teste como critério.

Há, ainda, estudos do teste SON-R (Rêgo, s.d.) que apesar de não ter sido criado a partir dessa teoria, mostram engajamento na adaptação dos testes e suas interpretações para melhor abarcar suas dimensões. No entanto, estes estudos ainda não estão concluídos. Neste contexto, o presente estudo teve como objetivo a construção de cinco testes de inteligência que formem uma bateria alinhada com as habilidades da teoria CHC. Adicionalmente, foi avaliada a qualidade dos itens que compõem estes testes para utilização como evidência de validade.

## **Método**

### **Seleção das habilidades**

Uma revisão de instrumentos que medem raciocínios multidimensionais nacionais e internacionais revelou um número elevado de possíveis tarefas para medida da inteligência. Assim, inicialmente foi elaborada uma planilha que identificasse a que classificação na teoria Cattell–Horn–Carroll, cada uma das tarefas se destinava. Cada instrumento foi avaliado por meio das descrições das habilidades de Schneider e McGrew (2012).

Como forma de identificar quais dessas habilidades seriam mais importantes para a construção da bateria, utilizou-se como critério a análise da planilha realizada pelo Grupo de

Pesquisa em Avaliação Psicológica no Contexto da Segurança Pública e Privada (Grupo Perfil), da Universidade de Brasília. Esta planilha continha uma lista de habilidades demandadas na seleção de profissionais de segurança pública. Para a construção dos testes, foram selecionadas tarefas associadas a cinco das habilidades apontadas nesta planilha. A seleção teve ainda, como critério, a inclusão de variância de pelo menos quatro tipos de habilidades amplas para medição do fator  $g$ , no primeiro estrato. Desta forma, também foi utilizado como critério que as habilidades selecionadas pertencessem a agrupamentos funcionais diferentes, conforme a Figura 1. O modelo resultante da seleção é apresentado na Figura 2.

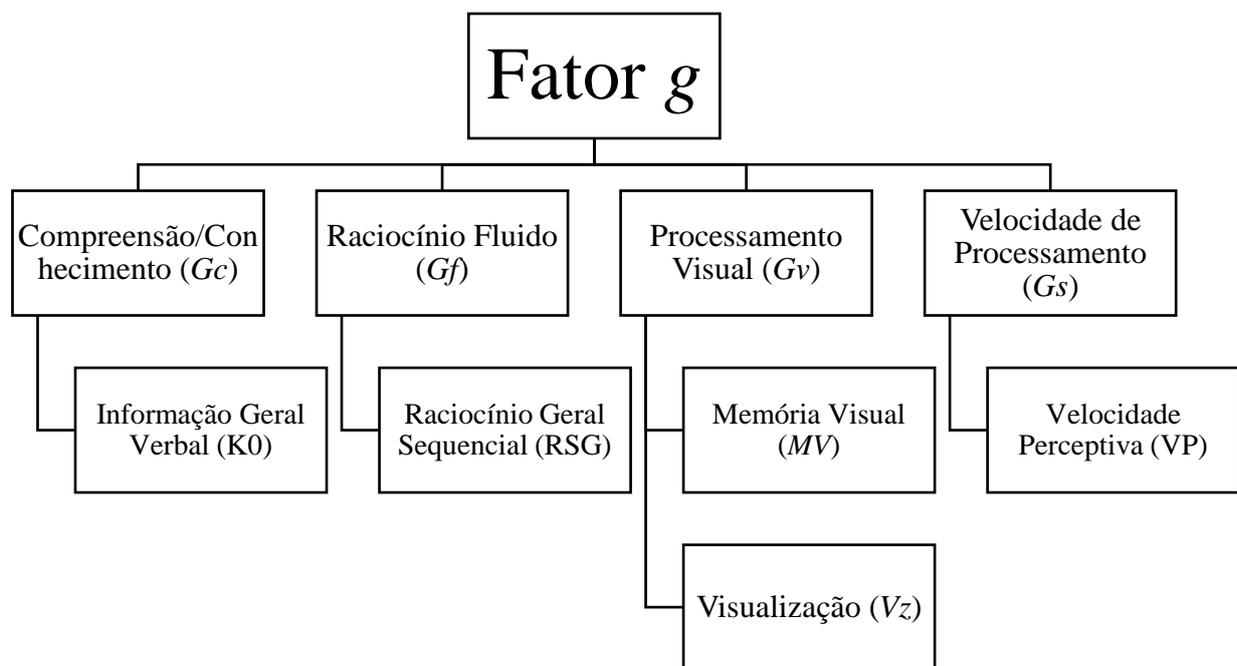


Figura 2.

*Habilidades selecionadas para a construção da bateria de testes no terceiro nível hierárquico, correspondente às habilidades específicas.*

### **Construção dos testes da bateria**

Os cinco testes foram construídos com o objetivo de mensurar três habilidades amplas, por meio de cinco habilidades específicas, baseadas no modelo CHC. Considerando a

necessidade de aplicação em grupo, os testes foram construídos com instruções extensivas e exemplos. Assim, os objetos de avaliação de cada instrumento foram:

1. Informação Geral Verbal (K0): informação geral verbal é uma habilidade situada dentro de compreensão/conhecimento ( $Gc$ ), e tem como objetivo avaliar o conhecimento acerca de termos gerais por meio de analogias com sentido de palavras.
2. Raciocínio Geral Sequencial (RGS): são habilidades específicas situadas na habilidade ampla de raciocínio fluido ( $Gf$ ). O processo de aprendizagem por indução é responsável pelo reconhecimento de padrões, e o raciocínio geral sequencial pela aplicação destes padrões a novas situações ou estímulos.
3. Memória Visual (MV): habilidade específica situada na habilidade ampla de processamento visual ( $Gv$ ). A memória visual é a capacidade de lembrar estímulos visuais por períodos curtos de tempo.
4. Velocidade Perceptiva (VP): habilidade específica situada na habilidade ampla de velocidade de processamento ( $Gs$ ). As tarefas clássicas de atenção são bons exemplos de velocidade perceptiva, visto que, conhecendo o tempo de aplicação, é possível avaliar a velocidade de reação do participante por meio da quantidade de estímulos marcados num dado período de tempo.
5. Visualização (Vz): habilidade específica situada na habilidade ampla de processamento visual ( $Gv$ ). A visualização é a capacidade de perceber padrões complexos de manipulação espacial e simular mentalmente como eles pareceriam após transformação.

A partir dessa definição, foram construídos itens para representarem cada um dos objetos definidos. Para isso, também foram utilizados como base os instrumentos nacionais e

internacionais, previamente avaliados, que mensuram mesmas medidas ou medidas correlatas à cada medida.

### **Análise de Especialistas**

Uma vez construídos, todos os testes foram submetidos a análise de pesquisadores na área de medidas (entre graduandos, pós-graduandos e docentes) que responderam todos os testes e, posteriormente, avaliaram cada um dos instrumentos por meio de um grupo focal. Os critérios avaliados foram de: a) tipo de estímulo do item; b) qualidade gráfica; c) dificuldade para resposta, segundo experiência na execução do instrumento; d) correção do gabarito do instrumento. Após correções, os testes novamente foram avaliados e aprovados por três juízes especialistas em sua forma final. Os juízes aprovaram todos os itens, fazendo apenas considerações sobre melhorias na página de instruções, que foi retificada. As versões finais dos testes não se encontram em anexo por estarem sujeitas a direitos autorais.

### **Estudo de evidências de validade**

#### **Participantes**

Participaram desse estudo 105 agentes de segurança pública, sendo 26 mulheres (24,8%). A idade média foi de 36,89 anos,  $DP = 8,40$ . Os participantes tinham, em média, 16,21 anos de serviço na instituição,  $DP = 9,51$ . A maior parte dos agentes era casado (67,6%) ou solteiro (22,9%). Quase todos os agentes possuíam ensino superior (74,2%), seguidos por 13,3% com ensino médio completo. Para acréscimo nos estudos de validade, o teste Visualização foi respondido, ainda, por mais 38 estudantes de universidade pública de diversos cursos. A média de idade dos estudantes foi de 22,21 anos,  $DP = 5,62$ . O sexo de 51,3% dos estudantes é masculino. A maior parte dos estudantes relatou ser solteira (92,3%).

#### **Instrumentos**

1. Questionário de informações pessoais.

2. A Bateria de Múltiplas Habilidades (BMH), composta por: Informação Geral Verbal, Raciocínio Geral Sequencial, Memória Visual, Velocidade Perceptiva e Visualização, a serem descritos como resultados do primeiro estudo.

### **Procedimentos**

Inicialmente foi realizado um estudo do tempo de aplicação de cada um dos instrumentos. Os testes foram aplicados em estudantes universitários e, considerando o objetivo de coleta de evidências de validade e a necessidade de resposta de todos os itens, foi escolhido para aplicação o tempo mais longo de resposta como critério para aplicação.

Para os estudos de evidências de validade, os cinco instrumentos aqui produzidos foram aplicados em agentes de segurança pública. O procedimento de aplicação constituiu-se na reunião dos participantes em salas de aplicação coletiva, nas quais os aplicadores seguiram um folheto de instruções redigido para padronizar as situações de testagem. Inicialmente, os aplicadores explicaram brevemente o propósito da pesquisa. Após a explicação, os participantes receberam o TCLE e o questionário sociodemográfico, este último de preenchimento optativo. Após assiná-lo, os testes foram entregues um a um, e a entrega do próximo se iniciava após o término do tempo do instrumento anterior. A cada instrumento, os aplicadores explicavam o propósito do instrumento e acompanhavam a leitura das instruções e realização das tarefas. Caso houvesse dúvida sobre a execução da tarefa, esta dúvida era sanada.

Ao final de cada instrumento, os participantes deveriam guarda-lo em plástico providenciado com este propósito. Ao fim da aplicação, os plásticos foram recolhidos. A administração dos demais participantes do instrumento de Visualização foi realizada individualmente ou em grupo.

## Resultados

As análises estatísticas foram realizadas com o IBM SPSS Statistics 23, e o R 3.4.1. A modelagem de traço latente, ou teoria de resposta ao item, é uma análise que permite a qualificação de um teste por meio da definição de parâmetros dos itens (van der Linder, Hambleton, 1997). Por este motivo, esta análise frequentemente é utilizada para coleta de evidências de validade (por ex., Almeida, & Primi, 2000), sendo considerada uma análise fatorial confirmatória para itens binários (Baker, & Kim, 2004).

O modelo escolhido para encaixe de TRI foi o de dois parâmetros (2PL). Este modelo tem como pressuposto a unidimensionalidade e a independência local, e descreve os itens por meio dos parâmetros de discriminação e de dificuldade (van der Linder, Hambleton, 1997). Para conferência do pressuposto de unidimensionalidade, foi avaliado o critério da razão entre o primeiro autovalor em relação ao número total de itens, que deve ser maior que 0,2 para considerar o instrumento “essencialmente unidimensional” (Maydeu-Olivares, 2014). Cuidados para a independência local foram tomados na construção dos itens.

O modelo de três parâmetros (3PL) não foi escolhido pelo princípio da parcimônia, uma vez que demanda um número maior de sujeitos para convergência e poder estatístico e não possui retornos neste caso: neste estudo, os participantes foram instruídos a responder honestamente, tendo ficado claro que não haveria consequências para os participantes caso obtivessem baixos escores. Além disso, o parâmetro adicional, o acerto ao acaso, é fixo e de baixo valor nos testes de múltipla escolha com cinco opções de resposta (Bock, & Aitkin, 1981). Os valores dos parâmetros ficaram livres para variar, visto que não foram especificadas restrições (*constraints*).

Foram construídos cinco testes cobrindo cinco habilidades específicas, compondo três habilidades amplas da teoria Cattell–Horn–Carroll. Os testes construídos avaliam: Informação Geral Verbal (K0), Raciocínio Geral Sequencial (RGS), Memória Visual (MV),

Velocidade Perceptiva (VP) e Visualização (Vz). A seguir serão apresentados os dados de cada instrumento, separadamente.

### **Informação Geral Verbal (K0)**

Os itens no teste de visualização são de múltipla escolha e foram construídos com base na tarefa de analogias de palavras. Neste teste, o respondente deve reconhecer os dois primeiros termos em negrito e a relação semântica entre eles. A partir disto, deverá mentalmente aplicar esta mesma regra semântica para no terceiro item. Por fim, deverá marcar a opção de resposta que corresponde ao termo. A Figura 3 possui um exemplo de item deste teste.

**Ex. Hoje está para Amanhã assim como Ontem está para**  
 a Anteontem      b Esta semana      c Este mês      d Anual      e Hoje

Figura 3.

*Exemplo de item do teste K0.*

O teste K0 foi composto por 45 itens de múltipla-escolha com 5 opções de resposta. Durante a análise de especialistas, dois itens foram removidos por terem sido considerados inadequados, restando 43 itens. Foram ainda removidos 15 itens por obterem carga fatorial inadequada, inferior a 0,3 (Tabatchnik & Fidell, 2007). Os escores variaram entre 5 e 26,  $M = 19,59$ ,  $DP = 4,31$ . A medida de fidedignidade lambda-2 de Guttman (1945) foi de 0,8, considerada desejável (DeVellis, 2012). A unidimensionalidade do instrumento foi verificada pela relação entre o primeiro autovalor e o número total de itens, que resultou em 0,2, considerado limítrofe, no entanto, aceitável, para a classificação de “essencialmente unidimensional”. (Maydeu-Olivares, 2013). O encaixe do modelo de TRI de dois parâmetros obteve  $\text{Log.Lik} = -1149,473$ .

Excluindo os itens outliers, a dificuldade dos itens variou entre -4,07 e 2,73. A dificuldade média dos itens, excluindo outliers, foi de -1,29. Considerando que a dificuldade

de 0 é adequada para a média da população, o teste pode ser considerado fácil. Dos 43 itens produzidos, 11 obtiveram dificuldade menor do que -1,96. Itens abaixo deste ponto são discriminativos apenas abaixo do percentil 5 de escores, sendo então considerados muito fáceis. A discriminação média dos itens foi de 1,21. Baker (2001) sugere que a discriminação entre 0,65 e 1,34 seja considerada “moderada”. A tabela 2 possui a lista completa de dificuldades e discriminações dos itens.

Tabela 2.

*Parâmetros de dificuldade e discriminação do teste K0, por ordem decrescente de discriminação.*

Item	<i>b</i>	<i>a</i>	Item	<i>b</i>	<i>a</i>
K0_q013	-0,63	4,64	K0_q037	0,89	0,97
K0_q010	-0,57	2,02	K0_q008	-2,75	0,96
K0_q006	-1,01	2,01	K0_q025	-2,29	0,89
K0_q012	-1,12	1,83	K0_q033	0,56	0,87
K0_q009	-1,50	1,68	K0_q026	-0,77	0,81
K0_q023	-2,48	1,53	K0_q028	-3,05	0,78
K0_q018	-0,27	1,51	K0_q030	-3,00	0,71
K0_q027	-1,41	1,51	K0_q016	-4,87	0,69
K0_q029	-1,23	1,49	K0_q031	1,88	0,66
K0_q032	-0,16	1,38	K0_q036	-1,26	0,65
K0_q019	-2,78	1,17	K0_q020	-4,58	0,63
K0_q004	-2,39	1,16	K0_q007	0,16	0,62
K0_q001	2,73	1,06	K0_q022	-4,07	0,56
K0_q005	1,07	0,97	K0_q003	-49,55	0,02

### **Raciocínio Geral Sequencial (RGS)**

Os itens de RGS foram construídos com base numa tarefa de analogias de figuras. Um exemplo de item deste teste pode ser encontrado na Figura 4. Nesta tarefa, o respondente deve observar a relação estabelecida entre o primeiro e o segundo estímulo. Após esta identificação, o respondente deverá transformar a terceira figura com base nesta mesma regra, e escolher a resposta que corresponda a esta transformação. Trinta itens foram construídos, e

todos foram aprovados pelos especialistas. Foram removidos 6 itens por não alcançarem a carga fatorial mínima (Tabatchnik, & Fidell, 2007).

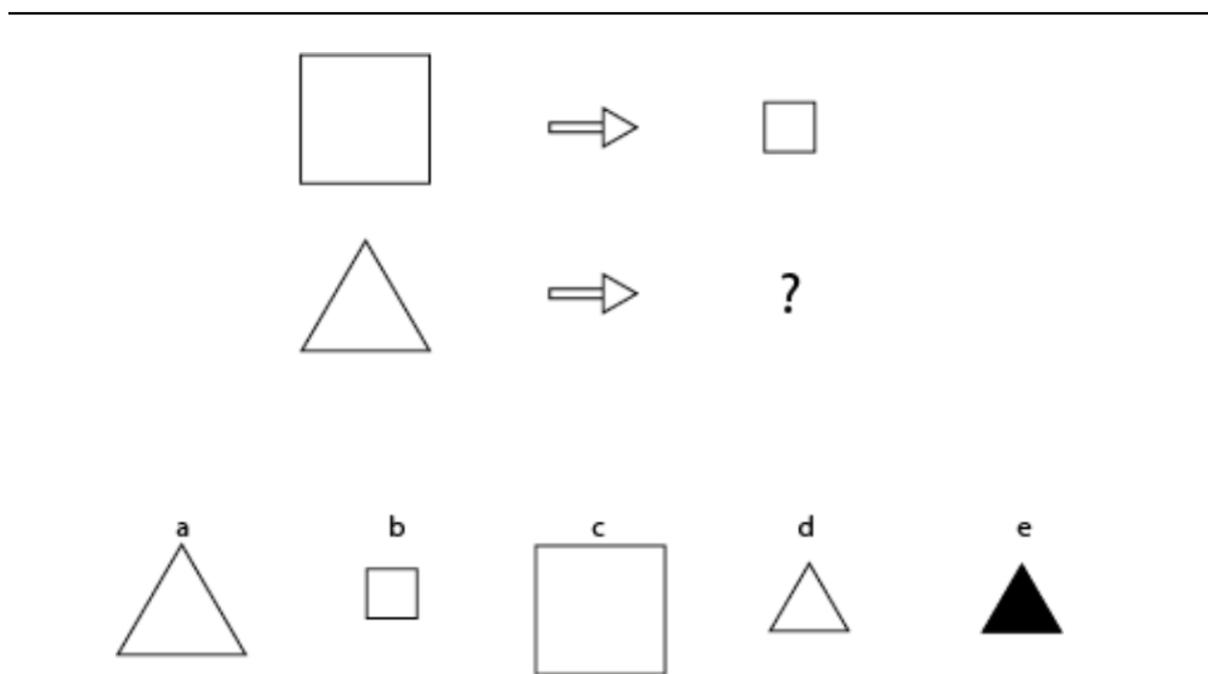


Figura 4.

*Exemplo de item do teste RGS.*

Os escores do teste RGS variaram entre 1 e 22,  $M = 13,90$ ,  $DP = 4,48$ . A medida de fidedignidade lambda-2 de Guttman foi de 0,76, considerada adequada (DeVellis, 2012). A unidimensionalidade do instrumento foi verificada pela relação entre o primeiro autovalor e o número total de itens, que resultou em 0,23, considerado “essencialmente unidimensional”. (Maydeu-Olivares, 2013). O encaixe do modelo de TRI de dois parâmetros obteve  $\text{Log.Lik} = -1289,006$ .

A dificuldade dos itens, excluindo os itens outliers, variou entre -2,91 e 1,88. Dos trinta itens produzidos, apenas um obteve dificuldade menor do que -1,96. A dificuldade média dos itens foi de -0,2, considerada adequada para a média da população. A discriminação média dos itens foi de 1,09, considerada moderada (Baker, 2001). A Tabela 3 relata as dificuldades e discriminações de cada um dos itens.

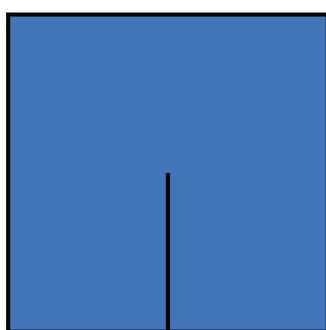
Tabela 3.

*Parâmetros de dificuldade e discriminação do teste RGS, em ordem decrescente de discriminação.*

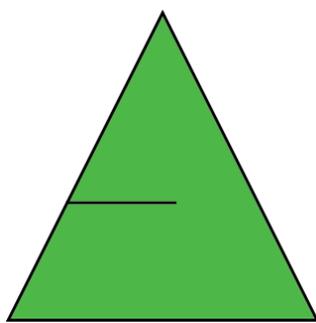
Item	<i>b</i>	<i>a</i>	Item	<i>b</i>	<i>a</i>	Item	<i>b</i>	<i>a</i>
rgs_q024	-0,87	3,12	rgs_q029	0,84	1,12	rgs_q017	0,03	0,83
rgs_q008	-0,77	1,99	rgs_q005	-0,77	1,10	rgs_q021	0,71	0,80
rgs_q003	-0,66	1,54	rgs_q001	-0,60	1,10	rgs_q007	0,48	0,71
rgs_q014	-0,75	1,44	rgs_q023	-0,93	0,93	rgs_q010	-0,01	0,69
rgs_q025	-1,63	1,26	rgs_q018	0,39	0,92	rgs_q002	-2,91	0,68
rgs_q016	-0,80	1,19	rgs_q027	1,05	0,90	rgs_q020	1,88	0,68
rgs_q028	-0,67	1,18	rgs_q004	0,09	0,88	rgs_q019	1,20	0,68
rgs_q013	-1,33	1,12	rgs_q022	1,10	0,84	rgs_q012	0,04	0,47

### Memória Visual (MV)

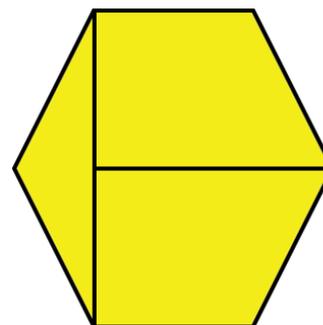
Na tarefa de memória visual, os respondentes tinham um minuto para memorizar uma grade com 12 figuras. As figuras apresentavam três graus de complexidade: forma, cor e desenho, vide exemplo abaixo. Após o tempo de memorização, deveriam abrir o caderno de respostas e marcar as figuras que estivessem presentes na grade de memorização. O escore foi calculado com base nos acertos. A figura 4 exemplifica os estímulos utilizados.



Quadrado azul com traço  
vertical



Triângulo verde com traço  
horizontal



Hexágono amarelo com dois  
traços

Figura 5.

*Exemplos de estímulos visuais para o teste de memória visual.*

A grade na qual os participantes respondiam era composta por 24 itens. Os escores no teste de Memória Visual variaram entre 8 e 24,  $M = 17,45$ ,  $DP = 5,7$ . A medida de fidedignidade lambda-2 de Guttman (1945) foi de 0,86, considerado desejável (DeVellis, 2012). A unidimensionalidade do instrumento foi verificada pela relação entre o primeiro autovalor e o número total de itens, que resultou em 0,38, considerado “essencialmente unidimensional” (Maydeu-Olivares, 2013).

O encaixe do modelo de TRI de dois parâmetros obteve  $\text{Log.Lik} = -948,096$ . Excluindo os itens outliers, a dificuldade dos itens variou entre -2,09 e -0,07. Cinco itens obtiveram dificuldade menor do que -1,96, sendo considerados muito fáceis. A dificuldade média dos itens foi -1,26, de forma que o teste pode ser considerado fácil. A discriminação média dos itens, excluindo os outliers, foi de 1,65, considerada “alta” (Baker, 2001). Uma lista completa dos parâmetros pode ser encontrada na tabela 4.

Tabela 4.

*Parâmetros de dificuldade e discriminação do teste MV, em ordem decrescente de discriminação.*

Item	<i>b</i>	<i>a</i>	Item	<i>b</i>	<i>a</i>
mv_q022	-1,426	30,75	mv_q001	-1,568	1,918
mv_q008	-0,779	15,382	mv_q010	-2,448	1,658
mv_q013	-1,068	5	mv_q020	-1,277	1,526
mv_q018	-0,824	4,127	mv_q019	-0,826	1,055
mv_q006	-0,888	3,464	mv_q003	-0,965	0,924
mv_q005	-1,074	2,999	mv_q002	-1,362	0,693
mv_q009	-1,764	2,841	mv_q012	-3,851	0,573
mv_q007	-0,907	2,777	mv_q011	-0,072	0,525
mv_q024	-0,779	2,686	mv_q004	-0,649	0,506
mv_q017	-0,919	2,666	mv_q021	-0,382	0,492
mv_q023	-0,82	2,612	mv_q014	-2,412	0,46
mv_q016	-2,09	2,591	mv_q015	-14,407	0,147

### **Velocidade Perceptiva (VP)**

A tarefa de velocidade perceptiva foi construída com base no tempo de percepção de estímulos visuais. Em um tempo pré-determinado, os respondentes deveriam navegar

visualmente uma série de linhas de estímulos e marcar aqueles que correspondiam aos três estímulos presentes no topo da folha, em ordem. Foram computados os acertos quando os respondentes encontravam os estímulos na ordem estipulada, além dos erros e omissões. As omissões não foram computadas após a última marcação dos participantes.

Os acertos variaram entre 0 e 46,  $M = 18,65$ ,  $DP = 13,69$ . Em média, os participantes também tiveram entre 5,95 omissões,  $DP = 3,34$ . Por fim, foram computados em média 5,16 erros,  $DP = 7,33$ .

### **Visualização (Vz)**

Os itens do teste de visualização foram construídos com base na tarefa de rotação de cubos. Foram identificadas e corrigidas as fraquezas das tarefas de rotação de cubos já existentes com relação à teoria CHC, nomeadamente, a utilização de estímulos posicionados em ângulos fisicamente impossíveis—que comprometia a validade ecológica, e a utilização de faces demasiadamente simples, que facilitavam o processo de visualização, visto que permitiam número muito limitado de rotações (Campos, 2005).

O teste de visualização é formado por itens de múltipla-escolha, no qual um cubo é apresentado de uma perspectiva e sofre rotação por duas ou três vezes, apresentadas nos cubos em sequência. O respondente deve identificar o tipo de rotação sofrida pelo cubo e aplica-la mentalmente ao próximo cubo da sequência, considerando as opções de resposta.

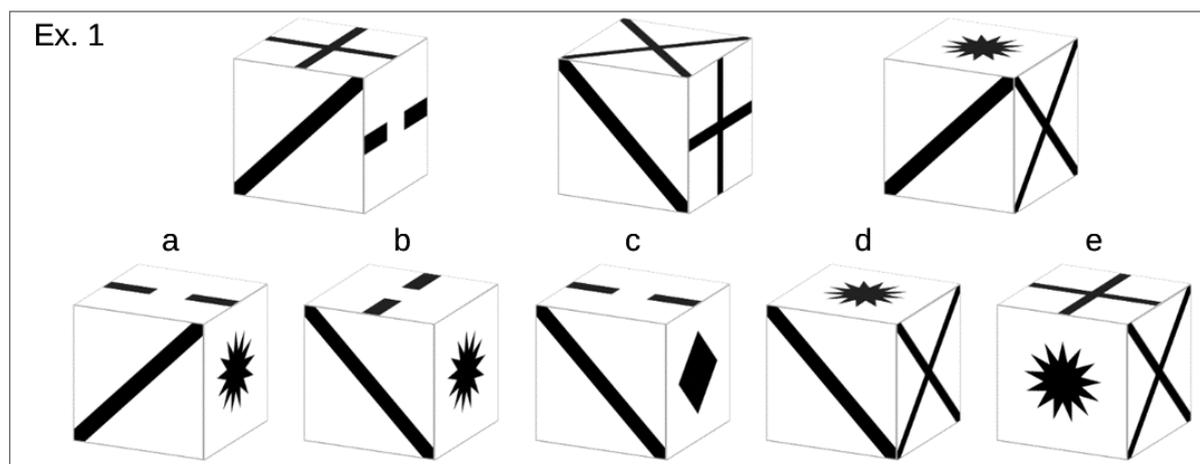


Figura 6.

*Exemplo de item do teste de visualização.*

Foram construídos 35 itens para o teste Vz. Sete itens foram descartados por terem sido considerado inadequados pela análise de juízes. Seis itens foram considerados inadequados devido a baixas cargas fatoriais (Tabatchnik & Fidell, 2007). O instrumento foi respondido por 141 participantes. Os participantes obtiveram média de 13,03 pontos, com desvio-padrão 5,69, variando entre 0 e 22. A fidedignidade, pelo lambda-2 de Guttman (1945), foi calculada em 0,93, considerada adequada (DeVellis, 2012). A unidimensionalidade do instrumento foi verificada pela relação entre o primeiro autovalor e o número total de itens, que resultou em 0,31, considerado “essencialmente unidimensional” (Maydeu-Olivares, 2013).

O encaixe do modelo de TRI de dois parâmetros obteve  $\text{Log.Lik} = -1726,386$ . Nenhum item teve dificuldade inferior a -1,96, que seria considerado muito fácil. A dificuldade média dos itens foi de -0,33. A discriminação média dos itens foi de 1,65, considerada “alta” (Baker, 2001). Todas as dificuldades e discriminações podem ser encontradas na tabela 5.

Tabela 5.

*Parâmetros de dificuldade e discriminação dos itens do teste de Visualização, em ordem decrescente de discriminação.*

Item	<i>b</i>	<i>a</i>	Item	<i>b</i>	<i>a</i>
vz_q008	-0,35	3,56	vz_q009	0,26	1,44
vz_q001	-0,89	2,82	vz_q017	-0,88	1,40
vz_q005	-0,74	2,51	vz_q018	0,66	1,31
vz_q002	-0,56	2,35	vz_q029	0,31	1,29
vz_q010	-0,23	1,86	vz_q030	-0,77	1,19
vz_q013	-0,14	1,83	vz_q020	0,00	1,12
vz_q016	0,04	1,81	vz_q031	-0,69	1,08
vz_q012	-0,12	1,68	vz_q004	-0,26	1,07
vz_q015	0,19	1,66	vz_q026	-0,99	1,04
vz_q006	-0,45	1,63	vz_q025	-0,50	1,04
vz_q007	-0,60	1,54	vz_q003	-0,47	1,01

## **Discussão**

Este estudo teve como objetivo a construção e coleta de evidências de validade de cinco testes de habilidades brasileiros com base na teoria Cattell-Horn-Carroll. Foram construídos cinco testes de inteligência, cada um correspondente a uma habilidade específica do modelo escolhido. As características do teste foram conhecidas a partir dos parâmetros dos itens calculados por meio do modelo de TRI de 2 parâmetros.

Em uma análise geral, cada um dos instrumentos apresentou alguns itens com coeficientes inadequados, e deverão ser removidos em futuras aplicações do teste. Apesar da baixa dificuldade de alguns itens não necessariamente dever resultar na eliminação destes itens, em situações típicas de testagem, não é frequente a construção de testes com precisão da discriminação a níveis abaixo do percentil 95 (Hom, Walsh, Doran, & Lott, 2017). Uma possível explicação para a facilidade dos itens, que não foram construídos com base em níveis de dificuldade menores do que as dos testes no mercado, é que a amostra utilizada é majoritariamente de pessoas com ensino superior.

Parte da demanda da construção da bateria foi a criação de testes de maior dificuldade. No entanto, diversos itens foram acertados por grande parte da amostra, resultando em níveis de dificuldade baixíssimos. Futuros estudos de construção de testes devem estudar a utilização de estímulos com maior nível de confusão por meio da inserção de diversos estímulos, um dos fatores que influencia a dificuldade (Campos, 2005). No teste RGS, por exemplo, poderiam ter sido utilizados mais elementos ou mais transformações em cada uma das imagens. No teste Vz, a partir dos dados observa-se a necessidade de que os cubos sofram rotações em direções adicionais. No teste K0, as palavras escolhidas podem ser mais raras, e poderiam ser estabelecidas relações mais abstratas entre elas.

Apesar destas evidências de validade iniciais serem adequadas, os *Standards for Educational and Psychological Testing* (AERA, APA, & NMCE, 2014) enfatizam que o processo de validação é contínuo, e que devam ser buscadas evidências de validade de fontes variadas. No presente estudo, apenas evidências de validade com base no conteúdo e na qualidade dos itens são apresentadas. No entanto, este tipo de evidência não possibilita que se aponte os processos que levam à resposta do item. Apesar da teoria CHC não trazer relações formais entre processos cognitivos e as habilidades previstas, suas categorias e descrições são largamente influenciadas pelos avanços nas ciências cognitivas (Schneider, & McGrew, 2012), e seus construtos podem ser razoavelmente atribuídos aos processos que são elencados nas respostas dos itens das habilidades, uma vez que a teoria faz a diferenciação em “agrupamentos funcionais” (Schneider, & McGrew, 2013). Uma vez que os processos envolvidos na resolução dos itens não são conhecidos pela análise fatorial ou teoria de resposta ao item, não é possível concluir que os instrumentos propostos sejam capazes de medir aquilo que se propõem (APA, AERA, & NMCE, 2014). Estudos futuros devem trazer evidências de validade com base na relação com variáveis critério, em outros testes que meçam construtos que sejam análogos aos aqui propostos.

Neste estudo, não foram coletadas evidências de validade com base nos itens do instrumento de velocidade perceptiva. Este instrumento possui estrutura de resposta incompatível com as análises comumente utilizadas. No entanto, futuras análises de evidências de validade com base na relação deste teste com tarefas parecidas devem ser suficientes para a aceitação deste instrumento.

Por fim, a nomeação da bateria (Bateria de Múltiplas Habilidades) teve como objetivo abarcar a perspectiva da inteligência conforme o modelo CHC. Apesar de modelos como Thurstone (1935) tratarem a divisão da inteligência em diferentes tipos de raciocínio, as categorias propostas por Schneider & McGrew (2012) excedem as concepções tradicionais de inteligência, trazendo aspectos como memória visual ou velocidade de processamento. Apesar da tradição de estudo da memória e da atenção (frequentemente operacionalizada como velocidade de processamento) como campos independentes de pesquisa, a realidade do processamento cognitivo não obedece às divisões conceituais e epistemológicas, e é possível encontrar mérito na proposição de que estas habilidades compõem a inteligência, conforme a organização ontológica destas habilidades compõe os processos dos quais se deriva o construto da inteligência.

### **Referências**

- AERA, APA, & NMCE. (2014). *Standards for Educational and Psychological Testing*. Washington: AERA Publications.
- Aken, L. van, Heijden, P. T. van der, Oomens, W., Kessels, R. P. C., & Egger, J. I. M. (2017). Predictive Value of Traditional Measures of Executive Function on Broad Abilities of the Cattell–Horn–Carroll Theory of Cognitive Abilities. *Assessment*, 1073191117731814.
- Almeida, L. S., Guisande, M. A., Primi, R., Ferreira, A. (2008). Construto e medida da inteligência: contributos da abordagem fatorial. Em: A. Candeias, L. Almeida, A.

- Roazzi & R. Primi (orgs.). *Inteligência: Definição e medida na confluência de múltiplas concepções*. São Paulo: Casa do Psicólogo.
- Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland College Park, MD.
- Baker, F. & Kim, S-H. (2004). *Item Response Theory* (2nd ed). New York: Marcel Dekker.
- Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux [A new method for the diagnosis of the intellectual level of abnormal persons]. *L'Année Psychologique*, 11, 191–244.
- Bock, R.D., Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46(4), 443–459.
- Campos, H. D. (2005). *Análise de conteúdo e sua relação com a dificuldade dos itens da bateria de provas de raciocínio (BPR-5)*. Dissertação de Mestrado, Universidade São Francisco.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40, 153–193.
- Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology*, 23, 285–299.
- DeVellis, R.F. (2012). *Scale development: Theory and applications*. Los Angeles: Sage.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10 (4), 255-282.
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic and developmental study of the structure among primary mental abilities*. Unpublished doctoral dissertation, University of Illinois, Champaign.

- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253–270.
- Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2012). A History of Intelligence Test Interpretation. In D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues (3rd ed.)* (p. 99-144). New York: Guilford.
- Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley- Interscience.
- Lei 6.119, de 27 de agosto de 1962 (1962). Diário Oficial da União. Seção 1. 05/09/1962. p. 9253.
- Lima, R. M. F. L. (s.d.). *Normatização e Validação do Teste não-verbal de Inteligência SON-R 6-40 para o Brasil*. Defesa de tese de doutorado não publicada. Universidade de Brasília.
- Maydeu-Olivares, A. (2013). Goodness of fit assessment of item response theory models. *Measurement*, 11, 71–101.
- Nicolas, S., Andrieu, B., Croizet, J.-C., Sanitioso, R. B., & Burman, J. T. (2013). Sick? Or slow? On the origins of intelligence as a psychological object. *Intelligence*, 41(5), 699–711.
- Pasquali, L. (2011). *Psicometria: Teoria dos Testes na Psicologia e na Educação*. Petrólis: Vozes.
- 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.
- Rapaport, D., Gill, M. M., & Schafer, R. (1945–1946). *Diagnostic psychological testing (Vols. 1–2)*. Chicago: Yearbook Publishers.

- Resolução CFP nº 002. (2003). Define e regulamenta o uso, a elaboração e a comercialização de testes psicológicos e revoga a Resolução CFP nº 025/2001. Conselho Federal de Psicologia.
- Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In, D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues (3rd ed.)* (p. 99-144). New York: Guilford.
- Schneider, W. J., & McGrew, K. (2013). The Cattell–Horn–Carroll (CHC) Model of Intelligence v. 2.2. *Institute of Applied Psychometrics*. Acessado em 08 de dezembro de 2017. Disponível em <http://www.iapsych.com/chcv2.pdf>.
- Silva, M. C. V. M. (2011). *História dos Testes Psicológicos: Origens e Transformações*. 1ª Edição, São Paulo: Vetor.
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15 (2), 201–292. doi:10.2307/1412107. JSTOR 1412107.
- Spearman, C. (1939). Thurston's work reworked. *Journal of Educational Psychology*, 30, 1–16.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.)*. Boston, MA: Allyn & Bacon/Pearson Education.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- van der Linder, W. J., & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.
- Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *The Wechsler intelligence scale for children—fourth edition*. London: Pearson.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army Mental Tests*. New York: H. Holt.

**MANUSCRITO 3:**

Evidências de validade com base na relação teste-critério de uma bateria brasileira baseada na  
teoria CHC

Validity evidence based on external criteria of a Brazilian battery based on the CHC theory

### Resumo

A teoria Cattell-Horn-Carroll tomou prominência na literatura de avaliação da inteligência devido a sua capacidade de abarcar hierarquicamente tanto um fator geral, quanto um número parcimonioso de habilidades amplas, sem perder a especificidade das tarefas encaradas por meio das chamadas habilidades específicas. A Bateria de Múltiplas Habilidades foi construída com base nesta teoria para possibilitar a testagem de três habilidades amplas, por meio de cinco testes que medem cinco habilidades específicas. No entanto, ainda não havia evidências de validade com base na relação com outros testes análogos. A bateria foi respondida com os testes RA, RV e RE da BPR-5, o TMR e o Atenção Alternada do BPA por 104 agentes de segurança pública. A matriz de correlações entre os escores confirmou largamente em magnitude e direção as relações hipotetizadas. Por fim, um modelo de equações estruturais foi construído com base no modelo teórico da CHC, e obteve coeficientes adequados.

*Palavras-chave:* Cattell–Horn–Carroll, validade convergente, validade de critério, modelagem por equações estruturais.

### **Abstract**

The Cattell–Horn–Carroll theory recently took prominence in intellectual assessment literature due to its capacity of hierarchically admitting a general factor and a parsimonious number of broad abilities without losing the task specificity with a more comprehensive number of narrow abilities. The Multiple Abilities Battery was constructed with basis on this theory to allow the measuring of three broad abilities through the measuring of five narrow abilities. However, there weren't yet test-criteria validity evidence over the relationship with analogous tests. The battery was administered along with the RA, RV and RE tests of the BPR-5, the TMR and the Alternating Attention test of the BPA by 104 public security agents. The correlation matrix largely confirmed the hypothesized relationships in magnitude and direction. Finally, a structural equation model was built based on the CHC theoretical model, and had adequate coefficients.

*Keywords:* Cattell–Horn–Carroll, convergent validity, construct validity, confirmation models.

A avaliação psicológica da inteligência está cada vez mais atrelada ao modelo de três estratos composto por um fator geral, 16 habilidades amplas e pelo menos 70 habilidades específicas conhecido como o modelo de Cattell–Horn–Carroll (CHC; Schneider, & McGrew, 2012; Schneider, & McGrew 2013). O impacto deste modelo é notável na academia e no desenvolvimento de testes psicológicos de inteligência (Jewsbury, Bowden, & Duff, 2017). No entanto, estudos acerca da prática dos psicólogos em situação real de avaliação psicológica, sugerem que os psicólogos não parecem seguir esta tendência (James, Jakobs, & Roodenburg, 2015).

A teoria CHC surgiu em resposta a diversos movimentos na área da avaliação da inteligência (Schneider, & McGrew, 2012). Uma das questões mais amplamente debatidas, segundo Kamphaus et al. (2012), seria a existência ou não de um fator geral de inteligência, que seria responsável pelo aprendizado ou desenvolvimento de qualquer habilidade específica. Spearman (1904) representa o primeiro lado desta discussão, defendendo em seu artigo seminal “*General Intelligence, Objectively Measured*” a existência de um fator geral de inteligência, o qual denominou *g*. Adversamente, um exemplo de teoria que dividia a inteligência em diferentes tipos de raciocínio foi proposto por Thurstone (1938), que propunha a existência de sete construtos independentes.

Outra questão debatida seria a precisão da dimensionalidade da inteligência como construto, que notadamente foi alvo de discussões pela necessidade de aplicar a parcimônia, mas não ignorar a relativa independência de habilidades específicas. No início do século do século XXI, a teoria de raciocínio fluido e raciocínio cristalizado (*Gf–Gc*), proposta por Cattell (1943) e estendida por Horn (Horn & Cattell, 1966), foi integrada a teoria dos três estratos de Carroll (1993). Desta forma, a existência do fator *g* não ameaçava a independência dos fatores, e a existência das habilidades específicas permitia parcimônia no número de habilidades amplas.

A teoria CHC resultou desta integração, e após uma série de revisões, consta hoje no que os autores chamam de uma “versão 2.2” (Schneider, & McGrew, 2013). Na versão atual, a teoria conta com 16 habilidades específicas compondo o fator geral (Tabela 1). Estas habilidades não são propriamente os fatores descritos em outras teorias como raciocínio, mas integram habilidades que até então vinham sendo tratadas como construtos independentes da inteligência, como a memória visual, integrada à habilidade ampla de Processamento Visual ( $Gv$ ) e a Velocidade de Percepção - uma habilidade ampla ( $Gs$ ).

Tabela 1.

*Habilidades amplas da teoria Cattell–Horn–Carroll. Uma tabela completa, com todas as habilidades específicas, pode ser encontrada em Schneider & McGrew (2013).*

Agrupamentos	Habilidades Amplas
<b>Funcional</b>	
1. Conhecimento Adquirido	Conhecimento Quantitativo ( $Gq$ ) Leitura e Escrita ( $Grw$ ) Compreensão–Conhecimento ( $Gc$ ) Conhecimento de Domínio Específico ( $Gkn$ )
2. Memória	Memória de Curto–Prazo ( $Gsm$ ) Armazenamento e Reobtenção de Longo–Prazo ( $Glr$ )
3. Velocidade	Velocidade de Processamento ( $Gs$ ) Tempo de Reação e Decisão ( $Gps$ )
4. Motor	Velocidade Psicomotora ( $Gps$ ) Habilidades Cinestésicas ( $Gk$ ) Habilidades Psicomotoras ( $Gp$ )
<b>Conceitual</b>	
1. Capacidades Gerais Domínio–Independentes	Raciocínio Fluido ( $Gf$ ) Memória de Curto–Prazo ( $Gsm$ ) Armazenamento e Reobtenção de Longo–Prazo ( $Glr$ )
2. Habilidades Específicas Sensorio–Motoras	Processamento Visual ( $Gv$ ) Processamento Auditório ( $Ga$ ) Habilidades Olfatórias ( $Go$ ) Habilidades Táteis ( $Gh$ ) Habilidades Cinestésicas ( $Gk$ ) Habilidades Psicomotoras ( $Gp$ )

No Brasil, a testagem psicológica é regulamentada pela Resolução CFP nº 002 (CFP, 2003). Este documento prevê que a utilização de testes psicológicos avaliados pelo SATEPSI depende da aprovação de uma comissão consultiva, composta por especialistas na área. O psicólogo deve utilizar, necessariamente, testes com evidências de validade no Brasil, avaliados por este órgão. Esta necessidade de aprovação, que destoa de práticas de outros países, pode ser um dos motivos pelo qual uma revisão dos testes aprovados no SATEPSI por Reppold et al. (2017) revelou que a avaliação da inteligência até agora conta com um número reduzido de testes.

Dentre os testes com parecer favorável, nenhum foi construído com o objetivo de cobrir as habilidades amplas da teoria CHC. A ausência de testes que cubram esta teoria compromete a validade da testagem realizada, uma vez que a interpretabilidade dos escores obtidos é comprometida pela falta de uma teoria substancial sobre a inteligência (Kaufman, 1979). Adicionalmente, considerando o efeito Flynn (2009), é de se esperar que os testes de inteligência eventualmente se tornem fáceis demais para a população para o qual é criado, que passa a demandar a existência de testes com maior capacidade de discriminação dentre as mais altas aptidões.

Neste contexto, a Bateria de Múltiplas Habilidades foi desenvolvida com o objetivo de preencher este nicho. Estudos de evidências de validade iniciais desta bateria, com base na qualidade dos itens, foram apresentados no manuscrito 2. Estas evidências foram produzidas a partir do encaixe de um modelo de teoria de resposta ao item. No entanto, de acordo com os parâmetros estabelecidos pelos *Standards for Educational and Psychological Testing*, considera-se que o processo de validação é contínuo, e que a apresentação de apenas um tipo de evidência de validade constituiria dificuldade no julgamento quanto à validade de um instrumento (AERA, APA, & NMCE, 2014).

Um dos tipos de evidências de validade mais classicamente utilizadas é com base na relação do teste com um critério externo (APA, AERA & NMCE, 2014). Este tipo de evidência de validade, outrora chamado validade de critério, é produzido a partir da computação da relação matemática entre os escores de um e de outro instrumento. Caso os instrumentos tenham objetivos similares, nomeia-se validade convergente. Caso tratem de construtos independentes, nomeia-se validade divergente (Pasquali, 2007).

Cronbach e Meehl (1955), em trabalho seminal, descreveram uma das formas que vieram a se tornar o tipo de evidência de validade mais frequente na validação de instrumentos psicológicos. A técnica, nomeada matriz multitraço-multimétodo, envolve a criação de uma matriz quadrada, que, quando utilizada como evidência de validade com base na relação do teste com o critério, descreve os índices de correlação entre os diferentes métodos utilizados. Byrne (2017) sugere que este método seja utilizado como validade de construto, uma vez que esta matriz situa o construto não apenas por sua correlação, mas pelas hipóteses acerca das relações entre o construto na rede nomológica.

Técnicas estatísticas mais recentes conseguem estabelecer relações ainda mais específicas. A modelagem por equações estruturais, por exemplo, tem sido utilizada como uma forma de mostrar a relação entre os escores de diferentes testes e entre os construtos que se diz medir, com base na teoria que embasa estes construtos (Finney, DiStefano, & Kopp, 2017). No campo da avaliação de inteligência, Primi, Nakano e Wechsler (2012) realizaram um estudo no qual a Bateria de Provas de Raciocínio (BPR-5), que foi baseada no modelo de Thurstone (1939), foi encaixada no modelo Cattell–Horn–Carroll juntamente com o teste Woodcock–Johnson III (WJ-III), que foi construído com base neste mesmo modelo teórico. Neste estudo, os autores tiveram como objetivo a coleta de evidências de validade cruzada. Para tal, apresentaram uma matriz de correlação entre os testes, bem como o modelo que utilizaram por meio da análise fatorial confirmatória. Os resultados não confirmaram o

modelo hipotetizado; no entanto, a bateria Woodcock–Johnson III é americana, e não possui evidências de validade em amostras brasileiras publicadas em artigo científico ou manual disponível publicamente.

A Bateria de Múltiplas Habilidades já obteve evidências de validade com base na qualidade dos itens, por meio do encaixe de modelos de teoria de resposta ao item (presente no manuscrito 2). No entanto, ainda não havia a produção de evidências de validade baseada em outras relações previstas nos Standards (2014). O presente estudo teve como objetivo obter e interpretar evidências de validade mais extensivas da bateria, com base na relação dos testes propostos com os demais testes. Outra importante fonte de evidência a ser apresentada é um modelo confirmatório para a composição do escore g com base na estimação das habilidades por meio dos testes e seus critérios.

## **Método**

### **Participantes**

Responderam aos testes 105 agentes de segurança pública, dos quais 26 eram mulheres (24,8%). Em média, os participantes tinham 36,89 anos, DP = 8,40. Os participantes tinham, em média, 16,21 anos de serviço, DP = 9,51. A maior parte dos agentes era casado (67,6%) ou solteiro (22,9%). Quase todos os agentes possuíam ensino superior (74,2%), e a maior parte dos demais (13,3%) possuía ensino médio completo.

### **Instrumentos**

Além de um questionário de informações pessoais contendo informações como idade, escolaridade e estado civil, foi utilizada uma versão do Termo de Consentimento Livre e Esclarecido (TCLE), contendo informações acerca da pesquisa, contato dos profissionais responsáveis e esclarecimentos acerca da não-obrigatoriedade e possibilidade de interrupção de participação na pesquisa.

Dentre os testes utilizados, está a Bateria de Múltiplas Habilidades (BMH) (construída no manuscrito 2), composta por: Informação Geral Verbal, Raciocínio Geral Sequencial (GRS), Memória Visual, Velocidade Perceptiva e Visualização. Estes testes recebem a nomeação de BMH-K0, BMH-RGS, BMH-MV, BMH-VP e BMH-Vz.

1. Informação Geral Verbal (K0): informação geral verbal é uma habilidade situada dentro de compreensão/conhecimento ( $G_c$ ), e tem como objetivo avaliar o conhecimento acerca de termos gerais por meio de analogias com sentido de palavras. Em estudo descrito no manuscrito 2, este teste obteve índice de fidedignidade de 0,8 pelo lambda-2 de Guttman (1945).
2. Raciocínio Geral Sequencial (RGS): habilidade específica situadas na habilidade ampla de raciocínio fluido ( $G_f$ ). O processo de indução é responsável pelo reconhecimento de padrões, e o raciocínio geral sequencial pela aplicação destes padrões. Em estudo descrito no manuscrito 2, este teste obteve índice de fidedignidade de 0,76 pelo lambda-2 de Guttman (1945).
3. Memória Visual (MV): habilidade específica situada na habilidade ampla de raciocínio visual ( $G_v$ ). A memória visual é a capacidade de lembrar estímulos visuais por períodos curtos de tempo. Em estudo descrito no manuscrito 2, este teste obteve índice de fidedignidade de 0,86 pelo lambda-2 de Guttman (1945).
4. Velocidade Perceptiva (VP): habilidade específica situada na habilidade ampla de velocidade de processamento ( $G_s$ ). As tarefas clássicas de atenção são bons exemplos de velocidade perspectiva, visto que, conhecendo o tempo de aplicação, é possível avaliar a velocidade de reação do participante por meio da quantidade de estímulos marcados num dado período de tempo.

5. Visualização (Vz): habilidade específica situada na habilidade ampla de raciocínio visual. A visualização é a capacidade de perceber padrões complexos e simular mentalmente como eles pareceriam após transformação. Em estudo descrito no manuscrito 2, este teste obteve índice de fidedignidade de 0,93 pelo  $\lambda$ -2 de Guttman (1945).
6. Três subtestes da Bateria de Provas de Raciocínio (BPR-5; Almeida, & Primi, 2000), nomeadamente o teste de raciocínio abstrato (RA), o teste de raciocínio verbal (RV) e o teste de raciocínio espacial (RE). A BPR-5 foi criada com base no modelo multidimensional de inteligência proposto por Thurstone (1938), e os testes RA, RV e RE utilizam tarefas de identificação de padrões não-verbal e verbal, e de rotações de cubos, respectivamente.
7. Teste de Memória Visual (TMR; Tonglet, 2007): utiliza uma tarefa de reconhecimento de figuras de diferentes combinações de cores e formas a fim de verificar a memória visual de curto ou longo prazo.
8. Bateria de Provas de Atenção, de Atenção Alternada (BPA: AA, Rueda, 2013): utiliza uma tarefa na qual o sujeito deve buscar uma série de estímulos, alternadamente, dentre uma linha que inclui estes estímulos dentre outros símbolos similares que servem como distratores.

## **Procedimentos**

Em dois momentos, os cinco instrumentos da BMH foram aplicados com outros cinco instrumentos de habilidades (BPR-5 RA, RE e RV, TMR e AA) e três instrumentos de outros construtos, em aplicação coletiva. Esta aplicação ocorreu em duas tardes consecutivas. Na primeira tarde responderam 79 participantes, e, na segunda, 26.

O procedimento de aplicação constituiu-se na reunião dos participantes em salas de aplicação coletiva, nas quais os aplicadores seguiram um documento padronizado de

instruções. De acordo com as instruções, os aplicadores explicaram brevemente o propósito da pesquisa. Após a explicação, os participantes receberam o TCLE e o questionário sociodemográfico, este último de preenchimento optativo. Após assiná-lo, os testes foram entregues um a um, e a entrega do próximo se iniciava após o término do tempo do instrumento anterior. A cada instrumento, os aplicadores explicavam o propósito do instrumento e acompanhavam a leitura das instruções e realização das tarefas. Caso houvesse dúvida sobre a execução da tarefa, esta dúvida era sanada antes do início do tempo de aplicação.

Ao final de cada instrumento, os participantes deveriam guarda-lo em plástico providenciado com este propósito. Ao final da aplicação, os plásticos foram recolhidos. Os participantes que assim desejassem poderiam registrar seus e-mails para receberem uma síntese dos resultados do estudo.

Para realização das análises estatísticas foram utilizados o IBM SPSS Statistics 23, e o R 3.4.1, com pacotes *lavaan*, *semTools* e *psych*, por meio do RStudio.

## **Resultados**

Os dados foram explorados a fim de encontrar *outliers* e valores ausentes. Foram produzidos, ainda, estatísticas descritivas sobre as características dos participantes. Os escores dos participantes foram calculados com base na teoria clássica dos testes, e verificados por anomalias.

Foi montada uma matriz de correlações de Pearson com valores de significância de uma cauda. Conforme a relação estabelecida entre as habilidades da teoria CHC, é razoável esperar que haja correlações positivas de maior ou menor magnitude entre todos os testes. No entanto, algumas correlações específicas podem ser melhor especificadas. Estas relações estão na Tabela 2.

Tabela 2.

*Relações esperadas na matriz de correlações.*

Teste construído	Tarefa análoga	Outras correlações importantes
BMH-VP	AA	Não possui
BMH-MV	TMR	Testes do Processamento Visual (BMH-Vz e BPR-5 RE)
BMH-K0	BPR-5: RV	Não possui
BMH-RGS	BPR-5: RA	Não possui
BMH-Vz	BPR: RE	Testes do Processamento Visual (BMH-MV e TMR)

**Matriz de correlações**

Houve correlações positivas significativas entre diversos testes, conforme é esperado entre as habilidades amplas que compõem o fator *g*. As correlações significativas a nível de  $p < 0,05$  tiveram magnitude entre 0,235 e 0,777. As correlações negativas não foram significativas. Os dados são apresentados na tabela 3.

A correlação esperada entre o BMH-VP e o AA era positiva de alta magnitude. No entanto, a correlação encontrada foi de 0,47, considerada moderada (Evans, 1996). O teste BMH-MV deveria se correlacionar positivamente e fortemente com o TMR. No entanto, a correlação teve magnitude fraca: 0,273. Outra correlação importante seria com o BMH-Vz e o BPR-5: RE. Estas correlações foram positivas de 0,34 e 0,32, respectivamente, sendo consideradas fracas.

Esperava-se correlação de alta magnitude entre o teste BMH-K0 e o teste BPR-5: RV. A correlação hipotetizada entre os dois foi confirmada, de 0,74, considerada forte. Entre o teste BMH-RGS e o teste BPR-5: RA, a correlação hipotetizada, positiva e forte, foi confirmada, de 0,78.

Por fim, o teste BMH-Vz deveria se correlacionar positivamente com o BMH-MV, TMR e o BPR-5: RE, sendo esta última correlação de alta magnitude,  $r = 0,64$ . A correlação entre o BMH-Vz e o BMH-MV foi significativa e positiva de baixa magnitude, de 0,336. Não houve correlação significativa entre os escores da BMH-Vz e o TMR,  $p = 0,214$ . A hipótese

de correlação positiva de alta magnitude entre o BMH-Vz e o BPR-5: RE foi confirmada,  $r = 0,635$ .

Tabela 3.

*Correlação entre os testes de habilidades da BHM, BPR-5, AA e TMR.*

Testes	BMH-VP	AA	TMR	BMH-Vz	BMH-K0	BMH-RGS	BPR5-RA	BPR5-RE	BMH-MV
BMH-VP	$r$ 1								
	$p$ <0,001								
AA	$r$ 0,474								
	$p$ <0,001								
TMR	$r$ 0,276	0,260							
	$p$ 0,004	0,004							
BMH-Vz	$r$ 0,381	0,312	-0,078						
	$p$ <0,001	<0,001	0,214						
BMH-K0	$r$ 0,153	0,101	-0,046	0,525					
	$p$ 0,073	0,152	0,322	<0,001					
BMH-RGS	$r$ 0,281	0,139	-0,067	0,606	0,724				
	$p$ 0,003	0,078	0,025	<0,001	<0,001				
BPR5-RA	$r$ 0,212	0,129	-0,133	0,603	0,707	0,777			
	$p$ 0,021	0,095	0,088	<0,001	<0,001	<0,001			
BPR5-RV	$r$ 0,136	0,039	0,044	0,520	0,740	0,731	0,710		
	$p$ 0,097	0,346	0,330	<0,001	<0,001	<0,001	<0,001		
BPR5-RE	$r$ 0,242	0,139	0,016	0,635	0,627	0,677	0,702	0,594	
	$p$ 0,010	0,078	0,436	<0,001	<0,001	<0,001	<0,001	<0,001	
BMH-MV	$r$ 0,235	0,137	0,273	0,336	0,238	0,345	0,296	0,359	0,321
	$p$ 0,012	0,082	0,003	<0,001	0,007	<0,001	0,001	<0,001	0,001

### Ajuste do modelo de equações estruturais

Inicialmente, foi desenvolvido um modelo fatorial que teve como princípio o agrupamento dos testes nas categorias propostas pelo modelo teórico CHC. Este modelo inclui, no nível hierárquico mais alto, o fator  $g$ . As habilidades amplas Processamento Visual ( $Gv$ ), Raciocínio Fluido ( $Gf$ ), Compreensão-Conhecimento ( $Gc$ ) e Velocidade de Processamento ( $Gs$ ) foram posicionadas no segundo nível hierárquico. O terceiro nível hierárquico foi composto pelos testes representando as habilidades específicas, por serem representantes de uma única habilidade específica, exceto no caso do BMH-Vz e BPR-5, que foram agrupados na habilidade específica Visualização (Vz), e o TMR e o BMH-MV, que foram agrupados em Memória Visual (MV). Os testes da Bateria de Múltiplas Habilidades já haviam sido criados com base nas descrições realizadas por Schneider e McGrew (2012), e

foram posicionados no terceiro nível hierárquico conforme sua construção. Os demais testes foram agrupados conforme a categoria de onde o seu teste análogo na bateria estava posicionado. O modelo final pode ser observado na Figura 1.

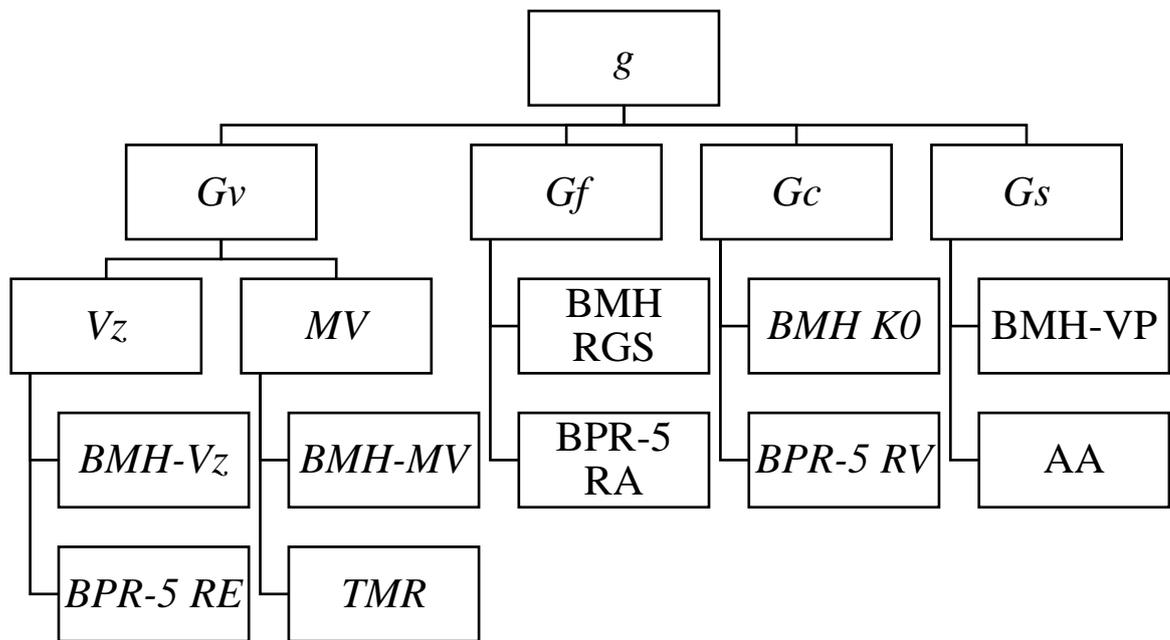


Figura 1.

*Representação gráfica do modelo utilizado na análise.*

A amostra necessária para o ajuste do modelo foi calculada com base na função *findRMSEA* do pacote *semTools*. O tamanho da amostra, com base no valor de probabilidade de erro tipo 1 de 5%, poder estatístico 80%, graus de liberdade e limites para o RMSEA foi de 111. O modelo confirmatório final foi especificado e calculado pela função *cfa* do pacote *lavaan*, utilizando o método da máxima verossimilhança com estimadores robustos.

Após o ajuste do primeiro modelo (Modelo 1), descrito conforme a teoria CHC, foi verificado que alguns coeficientes não alcançaram os valores adequados. Inicialmente, foi corrigido um problema de variâncias negativas por meio da definição de uma restrição. Além disso, foi encaixado um segundo modelo, no qual uma grande fonte de variância não explicada foi removida. Este segundo modelo é idêntico ao primeiro, exceto pela ausência do

teste TMR na habilidade ampla  $G_v$ . Os índices de ajuste produzidos podem ser verificados na Tabela 1.

Tabela 4.

*Valores das medidas de ajuste dos modelos confirmatórios especificados para análise.*

	$\chi^2$	gl	p	CFI	TLI	AGFI	NNFI	RMSEA
Modelo 1	55,18	31	0,005	0,938	0,909	0,971	0,916	0,096 IC90 [ 0,053 ; 0,136 ]
Modelo 2: sem TMR	27,731	23	0,226	0,987	0,980	0,980	0,968	0,049 IC 90 [ 0 ; 0,101 ]

Foi produzido um gráfico hierárquico com os coeficientes produzidos pelo modelo 1. Os coeficientes podem ser observados na Figura 1. Pela ausência de coeficientes, é possível observar que não foi possível formar o construto MV, devido à pequena correlação entre o TMR e o BMH-MV.

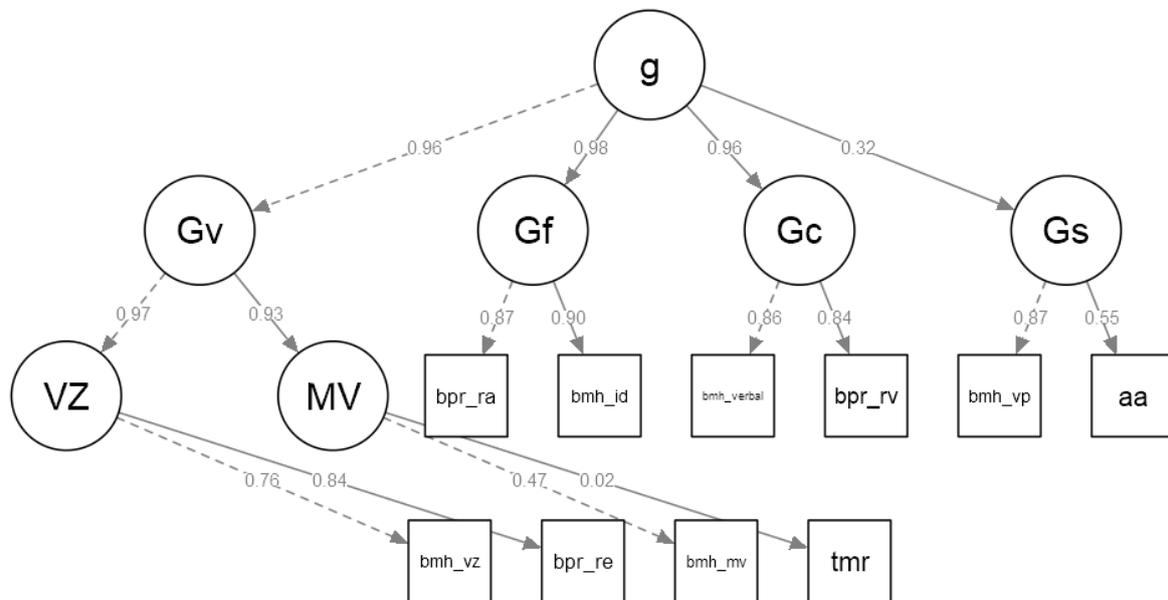


Figura 2.

*Coefficientes da análise por equações estruturais do modelo 1.*

A fim de alcançar coeficientes adequados, foi necessário construir um segundo modelo que levasse em consideração a anomalia introduzida pelas baixas correlações do teste TMR. Conforme os resultados expostos na Tabela 1, este modelo obteve coeficientes adequados, segundo recomendações de Hu e Bentler (1999). O teste do qui-quadrado não

rejeitou a hipótese nula, os índices de ajustamento foram superiores a 0,9, e o RMSEA, inferior a 0,06. O modelo resultante, bem como seus coeficientes, pode ser observado na Figura 3.

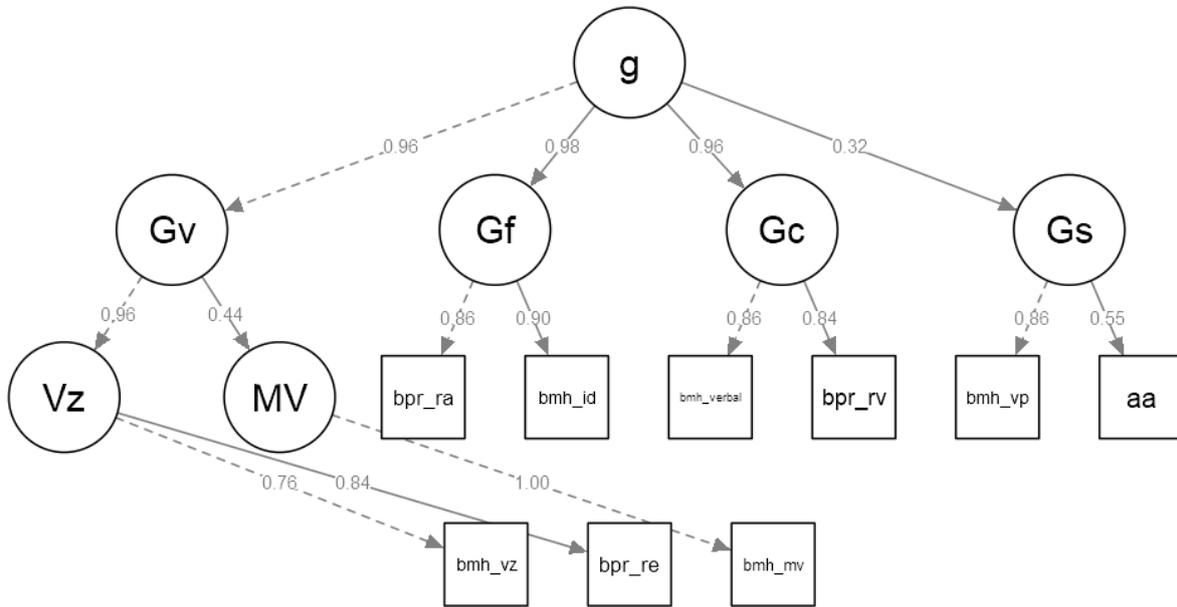


Figura 3.

*Coeficientes da análise por equações estruturais do modelo 2, sem o TMR.*

### Discussão

A construção de instrumentos psicológicos é um processo ainda pouco documentado. Dada a natureza abstrata dos construtos, é de se esperar que a mensuração seja dificultada. No caso dos instrumentos psicológicos, é inadequado considerar que a resposta a qualquer dado item seja relacionada causalmente a apenas um construto (Pasquali, 2011). Além de outros construtos que podem estar influenciando na resposta de uma dada pessoa, temos ainda as diversas fontes de erro decorrentes da aplicação e correção dos instrumentos (Hogan, 1999).

Duas características da forma como a medição de construtos psicológicos é feita até hoje podem justificar a dificuldade de obter um ajuste adequado dos modelos estatísticos. Por um lado, pode ser causada pela dificuldade de se isolar o construto abordado por meio de sua

operacionalização (Pasquali, 2011). Por outro, esta tarefa pode ser impossível, visto que a divisão dos construtos é frequentemente epistemológica: serve para explicação em um nível mais abstrato, e não necessariamente representa ontologicamente a forma como o item é realmente processado cognitivamente pelo sujeito que o responde, fato que Borsboom (2004) critica em trabalho seminal. De qualquer forma, construtos entendidos como distintos podem estar misturados em qualquer comportamento observável do indivíduo. No estudo de Primi et al. (2012), por exemplo, diversos testes tiveram melhor ajuste dentro do construto de inteligência fluida, que é extremamente amplo e representa processos cognitivos que, de uma forma ou de outra, estão envolvidos na maior parte das habilidades previstas na teoria Cattell–Horn–Carroll.

Esta dificuldade de ajuste dos modelos é verificada pelas medidas de adequação do ajuste estatístico. O problema de índices inadequados é tal que é comum a recomendação de que a estatística computada do qui-quadrado seja levada em consideração para avaliar a adequação do ajuste do modelo apenas em conjunto com outros índices, visto que pode rejeitar bons modelos ou aceitar maus modelos (Barrett, 2007; Bentler, 1980; Gefen, Straub & Bordreau, 2000; Hu & Bentler, 2000). Apesar disso, o consenso acadêmico continua sendo no sentido de preferir modelos que não rejeitem a hipótese nula no teste do qui-quadrado. Outra prática frequente é a modificação do modelo com base nas recomendações pelos índices de modificação que os softwares podem sugerir. No entanto, a utilização destas modificações produz um novo modelo, não vinculado a uma hipótese, e não parece haver justificativa teórica para que isto seja feito para aceitação de evidências de validade de um instrumento, exceto se para diagnosticar as fraquezas no modelo atual.

No presente estudo, foram produzidas uma série de medidas de adequação do modelo. Dentre os índices apresentados, que representam a relação da variância explicada pelo modelo e a variância não explicada no caso de índices absolutos, e a relação da variância

explicada pelo modelo e a variância explicada por um modelo nulo, no caso de índices relativos, considera-se que o valor ideal para aceite de um modelo seja acima de 0,9 (Baumgartner, & Hombur, 1996). No caso do RMSEA, que é uma medida representando o erro médio, considera-se que o valor ideal seja abaixo de 0,06. O modelo proposto se adequou sob os índices CFI, TLI e GFI. No entanto, o modelo não se adequou considerando a rejeição da hipótese nula sob o qui-quadrado e o RMSEA, apesar de conter valores abaixo de 0,06 no intervalo de confiança, é mais bem estimado pelo valor de 0,09, acima do valor desejado.

Apesar de diversas correlações entre os instrumentos confirmarem o modelo pretendido, uma em especial comprometeu o ajuste do modelo 1. Trata-se da relação entre o teste de memória visual TMR e o fator de raciocínio visual. O ajuste do modelo revelou que a relação entre os dois é de apenas -0,02, sem significância estatística. No modelo CHC, proposto por Schneider e McGrew (2012), a habilidade ampla de raciocínio visual é composta por, entre outras habilidades específicas, a Visualização (mensurada pelo teste BMH-Vz) e a Memória Visual (mensurada pelos testes TMR e BMH-MV). Neste sentido, o modelo testado incluía estes testes sob a habilidade de raciocínio visual. No entanto, o teste TMR não pontuou significativamente, e obteve correlação significativa, porém baixa, com o teste BMH-Vz. Este resultado é surpreendente, visto que os dois testes possuem atividades que elencam largamente as mesmas capacidades mentais. Uma hipótese para este resultado pode ser o desinteresse dos participantes em realizar dois testes distintos com o mesmo propósito. No entanto, para testar esta hipótese, os testes deveriam ter sido aplicados em grupos diferentes em ordens distintas, o que não foi possível neste estudo, considerando as questões logísticas de aplicação de mais de uma centena de testes simultaneamente.

O modelo 2, no qual o teste TMR está ausente, revela índices consideravelmente melhores e alcança todas as recomendações propostas por Baumgartner, & Hombur (1996) e

Hu e Bentler (1999). O qui-quadrado não revelou diferenças significativas, os índices absolutos e relativos foram acima de 0,9, e o RMSEA, abaixo de 0,06. Estes resultados permitem concluir que, removida a anomalia introduzida pelo TMR, o modelo se encaixa adequadamente aos dados.

Apesar das evidências aqui apresentadas serem satisfatórias, este estudo não constitui uma fonte definitiva de evidência de validade. Conforme a recomendação dos Standards (2014), o processo de validação é contínuo e as evidências de validade devem estar sendo coletadas durante toda a vida útil de um instrumento. Adicionalmente, o ajuste de modelos de equações estruturais ou análise fatorial confirmatória é apenas uma indicação estatística de que o modelo proposto é um modelo possível. No entanto, este ajuste não permite a extrapolação para a ideia de que este modelo é definitivamente correto e descreve adequadamente o fenômeno estudado.

Por fim, conforme recomendação por Newton (2016), o julgamento de validade depende da apresentação de uma série de evidências de maior ou menor importância. No presente estudo, apresentou-se evidências de validade de grande importância, conforme descrição nos Standards (2014). No entanto, novas evidências, de diferentes bases, devem ser produzidas para garantir a qualidade da avaliação por meio dos testes aqui apresentados.

### **Referências**

AERA, APA, & NMCE. (2014). *Standards for Educational and Psychological Testing*.

Washington: AERA Publications.

Barrett, P. (2007). Structural equation modeling: Judging model fit. *Personality and*

*Individual Differences*, 42(5), 815-824.

Baumgartner, H., & Hombur, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13, 139-161.

- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588-606.
- Byrne, B. M. (2017). Using Multitrait–Multimethod Analyses in Testing for Evidence of Construct Validity. In K. Schwizer, DiStefano C. (eds), *Principles and Methods of Test Construction: Standards and Recent Advances*. Hogrefe Publishing.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40, 153–193.
- DeVellis, R.F. (2012). *Scale development: Theory and applications*. Los Angeles: Sage.
- Finney, S. J., DiStefano, C., Kopp, J. P. (2017). Overview of Estimation Methods and Preconditions for Their Application With Structural Equation Modeling. Em K. Schweizer, C. DiStephano (eds.). *Principles and Methods of Test Construction: Standards and Recent Advances*. Hogrefe.
- Flynn, J. R. (2009). *What Is Intelligence?*. Cambridge University Press. Kindle Edition.
- Gefen, D., Straub, D. W., & Boudreau, M. C. (2000). Structural equation modeling and regression: Guideline for research practice. *Communications of the Association for Information Systems*, 4(1), 1-77.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253–270.
- Hu, L., & Bentler, M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.

- James, L., Jacobs, K. E., & Roodenburg, J. (2015). Adoption of the Cattell–Horn–Carroll model of cognitive abilities by Australian psychologists. *Australian Psychologist*, 50(3), 194–202. <https://doi.org/10.1111/ap.12110>
- Jewsbury, P. A., Bowden, S. C., Duff, K. (2017). The Cattell–Horn–Carroll Model of Cognition for Clinical Assessment. *Journal of Psychoeducational Assessment*, 35(6), 1-21.
- Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2012). A History of Intelligence Test Interpretation. In, D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues (3rd ed.)* (p. 99-144). New York: Guilford.
- Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley- Interscience.
- Kolenikov, S., Bollen, K. A. (2012). Testing Negative Error Variances: Is a Heywood Case a Symptom of Misspecification? *Sociological Methods and Research*, 41(1), 124–167.
- Newton, P. E. (2016). Macro- and Micro-Validation: Beyond the Five Sources Framework for Classifying Validation Evidence and Analysis. *Practical Assessment, Research and Evaluation*, 21(12), 1–13.
- Pasquali, L. (2007). Validade dos Testes Psicológicos: Será Possível Reencontrar o Caminho? *Psicologia: Teoria e Pesquisa*, 23(especial), 99-107.
- Pasquali, L. (2011). *Psicometria: Teoria dos Testes na Psicologia e na Educação*. Petrólis: Vozes.
- 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.

- Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In, D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues (3rd ed.)* (p. 99-144). New York: Guilford.
- Schneider, W. J., & McGrew, K. (2013). The Cattell–Horn–Carroll (CHC) Model of Intelligence v. 2.2. *Institute of Applied Psychometrics*. Acessado em 08 de dezembro de 2017. Disponível em <http://www.iapsych.com/chcv2.pdf>.
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15 (2), 201–292. doi:10.2307/1412107. JSTOR 1412107.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

## **Considerações Finais**

Esta dissertação foi composta por três estudos. No primeiro estudo, foi realizada uma revisão da literatura com o objetivo de identificar a maturação da pesquisa em avaliação da inteligência. No segundo estudo, o nicho de instrumentos de inteligência decorrente da pequena quantidade de testes inspirou a construção de cinco testes de inteligência que tinham como objetivo começar a cobrir algumas das dimensões descritas na teoria CHC. No terceiro estudo, as evidências de validade que começam a ser demonstradas no segundo estudo são levadas mais a fundo, com uma matriz de correlação entre testes e seus critérios, e um modelo de equações estruturais que relaciona os testes construídos com base no modelo previsto pela teoria Cattell–Horn–Carroll.

No primeiro estudo, foi possível observar que a literatura internacional de inteligência tem utilizado com maior precisão o conceito de inteligência e de suas dimensões na formulação de seus estudos. Considerando que há uma escassez de testes de inteligência no Brasil, especificamente de testes construídos com base na teoria Cattell–Horn–Carroll, haveria uma dissonância entre a literatura brasileira e a internacional. Apesar do estudo revelar que há uso considerável da teoria CHC nos artigos publicados no Brasil, dentre os instrumentos utilizados nos artigos, nenhum havia sido criado com base nesta teoria. Este dado é preocupante, visto que os testes de inteligência não podem ser reutilizados constantemente. Conforme as tarefas e os estímulos passam a ser conhecidos pelo respondente, as condições de testagem passam a ser consideravelmente diferenciadas em relação àquelas para as quais as evidências de validade foram coletadas. Em outras palavras, os escores passam a ser inválidos.

Neste estudo, foi realizada apenas uma avaliação dos artigos científicos. Neste sentido, a prevalência da teoria CHC na avaliação da inteligência só pôde ser conhecida na academia. No entanto, de nada valem os avanços na avaliação psicológica, se estes avanços

não estão alcançando os consultórios dos psicólogos. Há um nicho de pesquisa sobre a integração da teoria com a prática da avaliação psicológica realizada por psicólogos e neuropsicólogos que deveria ser preenchido.

No segundo estudo, foram construídos cinco testes de inteligência. Além disso, algumas evidências de validade com base na qualidade dos itens foram demonstradas. Neste estudo, grande parte dos itens foi classificado como fácil ou medianamente fácil. Considerando que o objetivo do estudo era a construção de itens mais difíceis, foi levado em conta no processo de construção a necessidade da inclusão de mais elementos nos itens. No entanto, não há uma teoria substancial acerca das dificuldades que são enfrentadas pelos sujeitos durante a resposta dos itens que norteie este procedimento. Novos estudos podem ser realizados para este fim.

No terceiro estudo, os testes previamente construídos foram aplicados juntamente com testes já disponíveis no mercado. Famosamente, há um paradoxo no cálculo da qualidade ou capacidade discriminativa dos itens na teoria clássica dos testes, por esta envolver a correlação ponto-bisserial entre o acerto ou erro do candidato e sua nota. A crítica é que se ainda não se conhece a discriminação dos demais itens, como se pode afirmar que o teste como um todo é um bom parâmetro para a qualidade do item?

Evidências de validade com base na correlação do teste com outros testes já existentes sofrem deste mesmo paradoxo. Se o objetivo da dissertação era a construção de instrumentos que sejam melhores do que aqueles já existentes no mercado, como a qualidade deles pode ser avaliada com base na correlação com estes instrumentos que se tentava melhorar? Certamente, se espera que os instrumentos se correlacionem em alguma medida, mas não tanto que o escore de um e de outro sejam intercambiáveis. No entanto, não existem evidências de que esta variância não compartilhada seja uma melhora do instrumento, e não uma piora, ou variação normal relacionada a diferenças na atividade.

Por fim, o processo de construção de testes é custoso. A aplicação de um número grande de testes é um procedimento caro, e não há como aplicar um número suficientemente grande de itens sem cansar os participantes do estudo, comprometendo a validade dos dados obtidos. Neste sentido, a construção de testes nunca será capaz de cumprir a demanda de oferecer novos itens para sempre que um psicólogo necessitar reavaliar a inteligência. No contexto de concursos públicos, no qual candidatos participam de vários processos diferentes, o risco da aprendizagem ameaçando a validade da aplicação é real. Os construtos testados nos concursos são comumente os mesmos, e um pequeno número de testes é disponível para estas mensurações. Para a produção de itens que avaliem o mesmo construto, o ideal seria que, em condições de testagem já estabelecidas, como as dos próprios concursos, novos itens pudessem ser testados e ter seus parâmetros calculados. Este cenário, no entanto, só é possível na testagem computadorizada. Novos estudos devem investigar a viabilidade de construção de instrumentos com itens modulares, permitindo que os parâmetros da teoria de resposta ao item de novos itens sejam constantemente estimados, e como utilizar estes testes com base nas resoluções do Conselho Federal de Psicologia que regem a testagem psicológica até agora, que exigem formas estáticas de teste para aprovação de testes psicológicos no Sistema de Avaliação de Testes Psicológicos.

## Lista Geral de Referências

- AERA, APA, & NMCE. (2014). *Standards for Educational and Psychological Testing*. Washington: AERA Publications.
- Aken, L. van, Heijden, P. T. van der, Oomens, W., Kessels, R. P. C., & Egger, J. I. M. (2017). Predictive Value of Traditional Measures of Executive Function on Broad Abilities of the Cattell–Horn–Carroll Theory of Cognitive Abilities. *Assessment*, 1073191117731814.
- Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The Impact of the Cattell–Horn–Carroll Theory on Test Development and Interpretation of Cognitive and Academic Abilities. In D. Flanagan, P. Harrison. (eds.) *Contemporary Intellectual Assessment, Second Edition: Theories, Tests, and Issues*. New York: Guilford Press.
- Almeida, L. S., Guisande, M. A., Primi, R., Ferreira, A. (2008). Construto e medida da inteligência: contributos da abordagem fatorial. Em: A. Candeias, L. Almeida, A. Roazzi & R. Primi (orgs.). *Inteligência: Definição e medida na confluência de múltiplas concepções*. São Paulo: Casa do Psicólogo.
- Baker, F. & Kim, S-H. (2004). *Item Response Theory (2nd ed)*. New York: Marcel Dekker.
- Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland College Park, MD.
- Barrett, P. (2007). Structural equation modeling: Judging model fit. *Personality and Individual Differences*, 42(5), 815-824.
- Baumgartner, H., & Hombur, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13, 139-161.

- Becker, K.A (2003). History of the Stanford–Binet Intelligence scales: Content and psychometrics. *Stanford–Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No. 1*.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588-606.
- Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux [A new method for the diagnosis of the intellectual level of abnormal persons]. *L'Année Psychologique*, 11, 191–244.
- Bock, R.D.; Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*., 46(4), 443–459.
- Byrne, B. M. (2017). Using Multitrait–Multimethod Analyses in Testing for Evidence of Construct Validity. In K. Schwizer, DiStefano C. (eds), *Principles and Methods of Test Construction: Standards and Recent Advances*. Hogrefe Publishing.
- Campos, H. D. (2005). *Análise de conteúdo e sua relação com a dificuldade dos itens da bateria de provas de raciocínio (BPR-5)*. Dissertação de Mestrado. Universidade São Francisco.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, R. B. (1943). *The measurement of adult intelligence*. *Psychological Bulletin*, 40, 153–193.
- Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology*, 23, 285–299.
- Decker, S. L., Fillmore, P. T., & Roberts, A. M. (2017). Cognitive Mediators of Reading Comprehension in Early Development. *Contemporary School Psychology*, 4(1), 3–13.  
<https://doi.org/10.15540/nr.4.1.3>

- Finney, S. J., DiStefano, C., Kopp, J. P. (2017). Overview of Estimation Methods and Preconditions for Their Application With Structural Equation Modeling. Em K. Schweizer, C. DiStephano (eds.). *Principles and Methods of Test Construction: Standards and Recent Advances*. Hogrefe.
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of Cross-Battery Assessment (2nd edition)*. Michigan: Wiley & Sons.
- Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell- Horn-Carroll cognitive abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly*, 22, 200–233.
- Floyd, R. G., Meisinger, E., Gregg, N., & Keith, T. (2012). An explanation of reading comprehension across development using models form Cattell-Horn-Carroll theory: support for integrative models of reading. *Psychology in the Schools*, 49(8), 725–743.
- Flynn, J. R. (2009). *What Is Intelligence?*. Cambridge University Press. Kindle Edition.
- Gefen, D., Straub, D. W., & Boudreau, M. C. (2000). Structural equation modeling and regression: Guideline for research practice. *Communications of the Association for Information Systems*, 4(1), 1-77.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282.
- Hajovsky, D., Reynolds, M. R., Floyd, R. G., Turek, J. J., & Keith, T. Z. (2014). A multigroup investigation of latent cognitive abilities and reading achievement relations. *School Psychology Review*, 43(4), 385-406.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185. <http://dx.doi.org/10.1007/BF02289447>.

- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic and developmental study of the structure among primary mental abilities*. Unpublished doctoral dissertation, University of Illinois, Champaign.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253–270.
- Hu, L., & Bentler, M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55.
- James, L., Jacobs, K. E., & Roodenburg, J. (2015). Adoption of the Cattell–Horn–Carroll model of cognitive abilities by Australian psychologists. *Australian Psychologist*, *50*(3), 194–202. <https://doi.org/10.1111/ap.12110>
- Jewsbury, P. A., Bowden, S. C., & Duff, K. (2016). The Cattell–Horn–Carroll Model of Cognition for Clinical Assessment. *Journal of Psychoeducational Assessment*, *1*(June), 1–21. <https://doi.org/10.1177/0734282916651360>
- Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2012). A History of Intelligence Test Interpretation. In D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3rd ed.) (p. 99-144). New York: Guilford.
- Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley- Interscience.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children (2nd ed.)*. Circle Pines, MN: American Guidance Service.
- Khundrakpan et al. (2017). Imaging structural covariance in the development of intelligence. *Neuroimage*, *144*, 227–240.
- Koriakin, T., White, E., Breaux, K. C., DeBiase, E., O’Brien, R., Howell, M., ... Courville, T. (2017). Patterns of Cognitive Strengths and Weaknesses and Relationships to Math

Errors. *Journal of Psychoeducational Assessment*, 35(1–2).

<https://doi.org/10.1177/0734282916669909>

Kolenikov, S., Bollen, K. A. (2012). Testing Negative Error Variances: Is a Heywood Case a Symptom of Misspecification? *Sociological Methods and Research*, 41(1), 124–167.

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016a).

Classification agreement analysis of Cross-Battery Assessment in the identification of specific learning disorders in children and youth. *International Journal of School & Educational Psychology*, 4(3), 124–136.

<https://doi.org/10.1080/21683603.2016.1155515>

Lei 6.119, de 27 de agosto de 1962 (1962). Diário Oficial da União. Seção 1. 05/09/1962. p. 9253.

Lima, R. M. F. L. (s.d.). *Normatização e Validação do Teste não-verbal de Inteligência SON-R 6-40 para o Brasil*. Qualificação de doutorado não-publicada. Universidade de Brasília.

Loughman, A., Bowden, S. C., & D’Souza, W. (2014). Cognitive functioning in idiopathic generalised epilepsies: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, 43, 20-34. doi:10.1016/j.neubiorev.2014.02.012

McGill, R. J. (2015). Interpretation of KABC-II Scores: An Evaluation of the Incremental Validity of Cattell-Horn-Carroll (CHC) Factor Scores in Predicting Achievement. *Psychological Assessment*, 27(4), 1417–1426. <https://doi.org/10.1037/pas0000127>

Nettelbeck, T., & Burns, N. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences*, 48, 379–384. doi:10.1016/j.paid.2009.10.032

- Newton, P. E. (2016). Macro- and Micro-Validation: Beyond the Five Sources Framework for Classifying Validation Evidence and Analysis. *Practical Assessment, Research and Evaluation, 21*(12), 1–13.
- Nicolas, S., Andrieu, B., Croizet, J.-C., Sanitioso, R. B., & Burman, J. T. (2013). Sick? Or slow? On the origins of intelligence as a psychological object. *Intelligence, 41*(5), 699–711.
- Pasquali, L. (2007). Validade dos Testes Psicológicos: Será Possível Reencontrar o Caminho?. *Psicologia: Teoria e Pesquisa, 23*(especial), 99-107.
- Pasquali, L. (2011). *Psicometria: Teoria dos Testes na Psicologia e na Educação*. Petrólis: Vozes.
- 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.
- Rapaport, D., Gill, M. M., & Schafer, R. (1945–1946). *Diagnostic psychological testing (Vols. 1–2)*. Chicago: Yearbook Publishers.
- Resolução CFP nº 002. (2003). *Define e regulamenta o uso, a elaboração e a comercialização de testes psicológicos e revoga a Resolução CFP nº 025/2001*. Conselho Federal de Psicologia.
- Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In, D. Flanagan & P. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3rd ed.) (p. 99-144). New York: Guilford.
- Schneider, W. J., & McGrew, K. (2013). *The Cattell–Horn–Carroll (CHC) Model of Intelligence v. 2.2*. Institute of Applied Psychometrics. Acessado em 08 de dezembro de 2017. Disponível em <http://www.iapsych.com/chcv2.pdf>.

- Silva, M. C. V. M. (2011). *História dos Testes Psicológicos: Origens e Transformações*. 1ª Edição, São Paulo: Vetor.
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15 (2), 201–292. doi:10.2307/1412107. JSTOR 1412107.
- Spearman, C. (1927). *The abilities of man*. London: MacMillan & Co.
- Spearman, C. (1939). Thurston's work reworked. *Journal of Educational Psychology*, 30, 1–16.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Valentini, F., Gomes, C. M. A, Muniz, M., Mecca, T. P., Laros, J. A., Andrade, J. M. de (2015). Confiabilidade dos índices fatoriais da Wais-III adaptada para a população brasileira. *Revista Psicologia: Teoria e Prática*, 17(2), 123–139.
- van der Linder, W. J., & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.
- Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- Wechsler, D. (1944). *The measurement of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1991). *The Wechsler intelligence scale for children—third edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). *The Wechsler intelligence scale for children—fourth edition*. London: Pearson.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WAIS-IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment, 31*, 94–113. doi:10.1177/0734282913478030

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8*, 231–258.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army Mental Tests*. New York: H. Holt.

Zajac, I., Burns, N., & Nettelbeck, T. (2012). Do purpose-designed auditory tasks measure general speediness? *International Journal of Intelligence Science, 2*, 23–31.  
doi:10.4236/ijis.2012.22004