

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

**Evidências de validade convergente do SON-R 6-40,
um teste não-verbal de inteligência**

Gabriel Olimpio Nascimento de Almeida

Orientador: Prof. Dr. Jacob Arie Laros

Brasília – DF

2013

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

**Evidências de validade convergente do SON-R 6-40,
um teste não-verbal de inteligência**

Gabriel Olimpio Nascimento de Almeida

Dissertação apresentada ao Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações, como requisito parcial para a obtenção do grau de Mestre em Psicologia Social, do Trabalho e das Organizações.

Orientador: Prof. Dr. Jacob Arie Laros

Brasília – DF

2013

“Julgue seu sucesso pelas coisas que você teve que renunciar para conseguir”

(Dalai Lama)

AGRADECIMENTOS

Em primeiro lugar aos meus pais Manoel e Iracema, por me darem todas as oportunidades e segurança pra que eu chegasse até aqui.

Em segundo lugar ao meu professor e orientador Dr. Jacob Arie Laros, que me ensinou a maioria das coisas que vou levar pra minha vida acadêmica, e que mesmo quando pareceu impossível não desistiu de me ajudar. Um obrigado especial pelo esforço.

Aos meus tios Olimpio e Luci, que me proporcionaram momentos maravilhosos na vida e as outras oportunidades que eu nem sonhava que iria alcançar.

Aos meus colegas Arthur, Talita, Renata, Felipe, Mary, Márcia e Wladimir, sem os quais as coletas jamais poderiam ser realizadas.

Aos meus professores da graduação, em especial Frederico Neves Condé, que me mostrou um pouco sobre psicometria, e Márcio Borges Moreira, que me deu chances que jamais eu teria.

A Deus, ou qualquer representação de uma força maior que me motivou a chegar até aqui, mesmo nos momentos em que quis desistir de tudo.

Aos professores do PSTO com quem tive contato, em especial Fábio Iglesias e Hartmut Günther, por mostrar que a vida não é tão fácil assim.

ÍNDICE

Lista de Figura e Tabelas.....	v
Resumo.....	vii
Abstract.....	viii
1. Introdução.....	1
1.1 Origens da testagem psicológica.....	1
1.2 Principais conceitos de inteligência.....	2
1.3 Os testes de inteligência individuais.....	6
As Escalas Stanford-Binet (S-B).....	6
As escalas de inteligência Wechsler.....	8
As baterias Kaufman.....	10
As Escalas de Habilidade Diferencial – DAS.....	11
1.4 Os testes SON.....	12
1.5 O teste BPR-5.....	17
1.6 Os conceitos de validade e validação.....	20
1.7 Referencial teórico.....	25
2. Objetivos.....	26
3. Método.....	27
3.1 Estudo 1 – SON-R 6-40 e SON-R 2½-7[a].....	28
Participantes.....	28
Instrumentos.....	28
Procedimento.....	30
3.2 Estudo2 – SON-R 6-40 e BPR-5.....	30
Participantes.....	30
Instrumentos.....	31
Procedimento.....	31
3.3 Análise de dados.....	32
4. Resultados.....	34
4.1 Estudo 1 – Aplicação do SON-R 6-40 junto com o SON-R 2½-7[a].....	34
Análises descritivas dos parâmetros psicométricos.....	34
Análises de validade convergente.....	42
4.2 Estudo 2 – Aplicação do SON-R 6-40 junto com o BPR-5.....	43
Análises descritivas dos parâmetros psicométricos.....	45
Análises de validade convergente.....	53
5. Considerações Finais	54

6. Limitações da pesquisa de validade convergente do SON-R 6-40.....	57
7. Referências.....	59

Lista de Tabelas e Figura

- Figura 1.1** Histórico dos testes SON, retirado de Jesus (2009), com adaptações.
- Tabela 1.1** Fatores do estrato II da teoria de Carroll
- Tabela 1.2** Alguns fatores do estrato I da teoria de Carroll
- Tabela 4.1** Parâmetros psicométricos dos quatro subtestes Analogias, Mosaicos, Categorias e Padrões da bateria SON-R 6-40 e da bateria completa – Estudo 1(N = 44).
- Tabela 4.2** Consistência interna, proporção de acertos (valor p) e correlação item-total corrigida (r_{it}) dos itens do subteste Analogias da bateria SON-R 6-40 - Estudo 1(N=44).
- Tabela 4.3** Consistência interna, proporção de acertos (valor p) e correlação item-total corrigida (r_{it}) dos itens para o subteste Mosaicos da bateria SON-R 6-40 – Estudo 1 (N=44).
- Tabela 4.4** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Categorias da bateria SON-R 6-40 – Estudo 1 (N=44).
- Tabela 4.5** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Padrões da bateria SON-R 6-40 – Estudo 1 (N=44).
- Tabela 4.6** Parâmetros psicométricos dos quatro subtestes Mosaicos, Categorias, Situações e Padrões da bateria SON-R 2½-7[a], da Escala de Execução (EE), da Escala de Raciocínio (ER) e da bateria completa – Estudo 1 (N = 44).
- Tabela 4.7** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para os subtestes Mosaicos e Padrões da bateria SON-R 2½-7[a] (N=44).
- Tabela 4.8** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para os subtestes Categorias e Situações da bateria SON-R 2½-7[a] (N=44).
- Tabela 4.9** Coeficientes de validade e de determinação entre os escores de QI Total do SON-R 6-40, QI Total do SON-R 2½-7[a] e a Escala de Execução e a Escala de Raciocínio do SON-R 2½-7[a] (N=44).
- Tabela 4.10** Parâmetros psicométricos dos subtestes Analogias, Mosaicos, Categorias e Padrões da bateria SON-R 6-40 – Estudo 2 (N = 68).

- Tabela 4.11** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Analogias da bateria SON-R 6-40 – Estudo 2 (N=68).
- Tabela 4.12** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Mosaicos da bateria SON-R 6-40 – Estudo 2 (N=68).
- Tabela 4.13** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Categorias da bateria SON-R 6-40 – Estudo 2 (N=68).
- Tabela 4.14** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Padrões da bateria SON-R 6-40 – Estudo 2 (N=68).
- Tabela 4.15** Parâmetros psicométricos dos subtestes Raciocínio Verbal (RV), Raciocínio Abstrato (RA), Raciocínio Espacial (RE) e Raciocínio Numérico (RN) do BPR-5 (N = 68).
- Tabela 4.16** Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Raciocínio Verbal da bateria BPR-5 (N=68).
- Tabela 4.17** Consistência interna, dificuldade dos itens (valor p), correlação item-total corrigida (r_{it}), e média de correlação entre os itens para o subteste Raciocínio Abstrato (RA) da bateria BPR-5 (N=68).
- Tabela 4.18** Consistência interna (λ_2), dificuldade dos itens (valor p), correlação item-total corrigida (r_{it}), consistência interna e média de correlação entre os itens para o subteste Raciocínio Espacial (RE) da bateria BPR-5 (N=68).
- Tabela 4.19** Consistência interna, dificuldade dos itens (valor p), correlação item-total corrigida (r_{it}), consistência interna e média de correlações entre os itens para o subteste Raciocínio Numérico (RN) da bateria BPR-5 (N = 68).
- Tabela 4.20** Coeficientes de validade e coeficientes de determinação para os escores de QI Total do SON-R 6-40, e as baterias de Raciocínio Abstrato, Verbal, Espacial e Numérico (RA, RV, RE e RN, respectivamente) do teste BPR-5(N=68).

RESUMO

Estudos que fornecem evidências de validade dos escores de um teste são essenciais para assegurar uma interpretação correta dos seus resultados. Assim, este estudo investigou a validade convergente do teste não-verbal de inteligência SON-R 6-40 com dois outros testes que medem inteligência: o SON-R 2½-7[a] e a forma A da Bateria de Provas de Raciocínio (BPR-5). O SON-R 6-40 é um teste não-verbal de inteligência para pessoas entre 6 e 40 anos de idade que foi publicado na Holanda em 2010. O teste foi normatizado e validado para Holanda e Alemanha. O SON-R 2½-7[a] é um teste não-verbal de inteligência para crianças entre 2 anos e 6 meses e 7 anos e 11 meses que foi normatizado para o Brasil em 2009. Ambos os testes SON-R foram desenvolvidos para medir a inteligência fluida, assim reduzindo a influência de fatores socioeconômicos no desempenho do teste. O teste BPR-5 é uma bateria de provas que avaliam cinco diferentes tipos de raciocínio: Raciocínio Verbal (RV), Raciocínio Abstrato (RA), Raciocínio Espacial (RE), Raciocínio Numérico (RN) e Raciocínio Mecânico (RM). No presente estudo, entretanto, foram aplicados apenas quatro subtestes: o subteste que avalia raciocínio mecânico não foi aplicado. A forma A da BPR-5 destina-se a alunos da 6ª, 7ª e 8ª séries do ensino fundamental e foi normatizado para o Brasil em 2000. No total, 112 alunos do Distrito Federal de escolas públicas participaram no presente estudo: 44 alunos com idades entre 6 e 7 anos no primeiro estudo de validade com o SON-R 2½-7[a] e 68 alunos entre 12 anos e 16 anos no segundo estudo de validade com o BPR-5. A correlação corrigida para atenuação encontrada entre o SON-R 6-40 e o SON-R 2½-7[a] foi igual a 0,75 e entre o SON-R 6-40 e a BPR-5 foi de 0,72. Como esperado, as correlações mais altas foram encontradas entre o SON-R 6-40 e os subtestes de Raciocínio Abstrato e Raciocínio Espacial do BPR-5 uma vez que estes subtestes também medem a inteligência fluida ($r = 0,68$ para ambos os subtestes). Correlações mais baixas foram encontradas entre o SON-R 6-40 e os subtestes de Raciocínio Verbal e Numérico da BPR-5 que medem a inteligência cristalizada ($r = 0,62$ e $r = 0,44$ respectivamente). Os resultados da presente pesquisa justificam a conclusão que evidências satisfatórias de validade convergente do teste SON-R 6-40 foram obtidas para a faixa de idade investigada.

Palavras-chave: testes de inteligência, validade convergente, tipos de raciocínio

ABSTRACT

Studies that provide validity evidence of tests scores are essential to assure a correct interpretation of the outcomes of a test. In this context, the present study investigated the convergent validity of the SON-R 6-40 administering it together with two other instruments that measure intelligence: the SON-R 2½-7[a] and form A of the Battery of Reasoning Tests (BPR-5). The SON-R 6-40 is a nonverbal test of intelligence intended for persons with ages ranging from 6 to 40 years and that was published in 2010. Normatization and validation studies of the SON-R 6-40 has been undertaken in The Netherlands and in Germany. The SON-R 2½-7[a] is a nonverbal test of intelligence for children with ages varying from 2 years and 6 months to 7 years and 11 months. The normatization and validation studies of the SON-R 2½-7[a] for Brazil were concluded in 2009. Both SON-R tests were developed to measure fluid intelligence to reduce the influence of socioeconomic factors on the test performance. The BPR-5 is a battery of reasoning subtests that assess five different reasoning types: Verbal Reasoning (VR), Abstract Reasoning (AR), Spatial Reasoning (SR), Numerical Reasoning (NR) and Mechanical Reasoning (MR). In the present study, however, the Mechanical Reasoning subtest was not administered. Brazilian norms of the BPR-5 were established in 2000 for students of the 6th, 7th and 8th grades of elementary school. Altogether, 112 students of public schools from the Distrito Federal (Brazil) participated in the present research: 44 students with ages between 6 and 7 years in the first validation study and 68 students between 12 and 16 years in the second validation study. The results of these two studies indicate a correlation corrected for attenuation between the SON-R 6-40 and the SON-R 2½-7[a] of .75 and of .74 between the SON-R 6-40 and form A of the BPR-5. As expected, the highest correlations were found between the SON-R 6-40 and the Abstract and Spatial Reasoning subtests of the BPR-5 since these also measure fluid intelligence ($r = .72$ and $r = .73$, respectively). Lower correlations were encountered between the SON-R 6-40 and the Verbal and Numerical Reasoning subtests of the BPR-5 which measure crystallized intelligence ($r = .65$ and $r = .47$, respectively). The results of the present research justify the conclusion that satisfactory evidence of the convergent validity of the SON-R 6-40 has been obtained for the investigated age group.

Key words: intelligence tests, convergent validity, types of reasoning

1. Introdução

1.1 Origens da testagem psicológica

O início da utilização de testes psicológicos ocorreu entre o final do século XIX e os primeiros anos do século XX. Alguns psicólogos foram importantes para a ampliação e divulgação da importância do uso de testes, dentre eles é possível citar os trabalhos de Lewis Terman (1877-1956), James McKeen Cattell (1860-1944), Alfred Binet (1857-1911), Francis Galton (1822-1911), David Wechsler (1896-1981), dentre outros expoentes. Alguns desses autores serão retratados no presente trabalho devido à sua importância para o campo a ser estudado, a saber, avaliação de inteligência.

A testagem psicológica não surgiu com o intuito de avaliar a inteligência como concebida atualmente, mas avaliando características observáveis humanas, tais como força, acuidade, capacidade de respirar, julgamento de cores, dentre outras. Essas características foram inicialmente testadas no “Laboratório Antropométrico” de Francis Galton. Ele acreditava que as medidas tiradas em seu laboratório estavam, de alguma forma, relacionadas à inteligência. Para ele essas medidas físicas seriam auxiliares para uma avaliação do desempenho acadêmico formal, pois naquele tempo a única forma de se avaliar a inteligência era com esse tipo de medidas. Galton auxiliou na construção de medidas físicas e também utilizou suas habilidades estatísticas para criar formas diferentes para agrupar os resultados individuais. Usando postos percentis, Galton possibilitou a comparação de grupos de diferentes idades e gêneros. (Kamphaus, Winsor, Rowe, & Kim, 2012)

É importante notar que Galton nunca alegou que seus testes mediam inteligência de uma forma direta. Ele alegava que as medidas sensoriais que eram avaliadas apenas auxiliavam na medição das capacidades que poderiam estar relacionadas ao conceito de inteligência. (Kamphaus, Winsor, Rowe, & Kim, 2012)

Enquanto Galton pretendia avaliar as medidas sensoriais humanas, e estabelecer relações entre essas medidas e a inteligência, Alfred Binet trabalhava no desenvolvimento do primeiro teste de inteligência propriamente dito. Por muitos considerado o pai da testagem psicológica, Alfred Binet se formou como advogado inicialmente, porém não quis exercer sua profissão. Por influência dos pais, tentou estudar medicina, mas não teve interesse em aprofundar seus estudos. Após as tentativas frustradas de estudar direito e medicina, Binet interessou-se pelos estudos em psicologia. Com dedicação exclusiva aos estudos psicológicos, escreveu seu primeiro artigo aos 23 anos de idade, e obteve seu doutorado em ciências naturais aos 37 anos. Segundo Binet, a inteligência seria um conjunto de habilidade cognitivas dependentes de habilidades chamadas superiores e que juntas formavam o conceito (Binet e Simon, 1911).

1.2 Principais conceitos de inteligência

O estudo da inteligência é um dos maiores empreendimentos dentro da ciência psicológica. Ao longo de décadas diversas definições de inteligência foram apresentadas. Spearman (1904; 1927), por exemplo, desenvolveu a teoria de inteligência geral, afirmando que toda atividade mental intelectual envolveria uma única capacidade. Segundo a teoria de Spearman o desempenho nos testes de inteligência dependeria principalmente de uma única capacidade mental geral, o fator *g*. A variância específica de cada teste é considerada como variância de erro na teoria de fator *g*. Outros autores não definiam a inteligência como uma única capacidade, mas como uma capacidade multifatorial (Thurstone, 1938; Hogan, 2006). Thurstone diferenciou oito fatores de inteligência, que ele chamou de Habilidades Mentais Primárias: Compreensão Verbal (V), Fluência com palavras (W), Facilidade com Números (N), Memória (M), Visualização ou

Pensamento Espacial (S), Velocidade Perceptual (P), Indução (I), e Velocidade de Julgamento (J). Essas definições de Thurstone serviram de berço para algumas outras teorias, conhecidas como teorias hierárquicas de inteligência. Segundo esse grupo de teorias, realmente existem diversos fatores de inteligência, porém cada fator é responsável por uma porcentagem na determinação dos escores nos testes. Quanto maior a porcentagem que o fator determina mais alto será seu “grau hierárquico”. Uma das teorias hierárquicas mais estudadas e reconhecidas no meio acadêmico é o modelo teórico de inteligência fluida e cristalizada de Carroll-Horn-Cattell (Schneider & McGrew, 2012). A teoria CHC é uma representação da união entre a teoria de três estratos de Carroll com a teoria Gf-Gc de Cattell (Para a presente dissertação será adotada a teoria hierárquica de inteligência como um conceito geral, mais especificamente será utilizada a teoria Carroll-Horn-Cattell (CHC).

A teoria CHC é uma junção entre duas teorias: teoria Gf-Gc de Horn-Cattell e a teoria de três estratos de Carroll. A primeira diz respeito à diferenciação entre inteligência fluida (Gf) e inteligência cristalizada (Gc). O conceito de Gf é resumido como “a habilidade para resolver problemas novos, que não podem ser realizados confiando-se exclusivamente em hábitos, esquemas e scripts aprendidos anteriormente” (Schneider & McGrew, 2012, pág.103). Já a inteligência cristalizada é definida como “a habilidade de aplicar definições, métodos e procedimentos de solução de problemas, aprendidos previamente, para lidar com situações problema”. (Schneider & McGrew, 2012, pág.104).

A teoria de Carroll surgiu após uma meta-análise de 460 bancos de dados em quais foi utilizada a técnica de análise fatorial de dados empíricos de escores em testes de inteligência. Esses bancos de dados foram analisados novamente para padronizar e utilizar as técnicas atualizadas de análise fatorial e averiguar possíveis erros ou inconsistências práticas. Os resultados dessas novas análises indicaram a existência de

fatores com altas correlações entre eles. Com base nessas correlações altas, Carroll decidiu colocar esses fatores no mesmo nível. Ele distinguiu três níveis hierárquicos diferentes: no nível mais alto, há apenas o fator geral (g), no nível intermediário existem os dez fatores amplos: Inteligência Fluida (Gf), Inteligência Cristalizada (Gc), Tempo de Reação (Gt), Raciocínio Quantitativo (Gq), Velocidade de Processamento (Gs), Recuperação de Longo Prazo (Glr), Processamento Auditivo (Ga), Processamento Visual (Gv), Leitura-escrita (Grw), Memória de Curto-Prazo (Gsm). Após esse nível, são encontrados os chamados fatores específicos. Cada um desses níveis foi chamado de estrato devido à estrutura hierárquica da teoria (Carroll, 1994). Assim, para que um teste possa ser considerado pertencendo a um fator do primeiro estrato (estrato I), ele deve possuir uma correlação alta com outros testes que pertencem a este fator.

A Tabela 1.1 mostra os fatores componentes do segundo estrato da teoria (estrato II), bem como suas definições. Nesse estrato estão os chamados “fatores amplos”. Os “fatores amplos” são aqueles que apresentam as maiores correlações com o fator geral. O terceiro estrato não é apresentado, pois nele está contido apenas o fator geral de inteligência, o fator g. (Carroll, 1993)

Tabela 1.1 Fatores do estrato II da teoria do Carroll

Símbolo	Nome do Fator	Descrição
Gf	Inteligência fluida	Raciocínio em situações novas
Gc	Inteligência cristalizada	Quantidade de conhecimentos adquiridos
Gt	Tempo de reação	Reagir a estímulos rapidamente
Gq	Raciocínio quantitativo	Raciocínio com estruturas matemáticas
Gs	Velocidade de processamento	Responder dentro de um intervalo definido
Glr	Recuperação de longo prazo	Recuperar informação de longo prazo
Ga	Processamento auditivo	Compreensão de estímulos auditivos
Gv	Processamento visual	Compreensão de estímulos visuais
Grw	Leitura-escrita	Expressão escrita bem elaborada
Gsm	Memória de curto prazo	Uso da informação por curto período

Informação retirada de Schelini (2006)

O estrato I apresenta os chamados “fatores estreitos”. Existem mais de 80 fatores estreitos listados pela teoria de Carroll. Cada fator desse estrato está relacionado com um fator do estrato II. Alguns desses fatores são apresentados na Tabela 1.2

Tabela 1.2 Alguns fatores do estrato I da teoria do Carroll

Símbolo	Nome do Fator	Descrição
RG	Raciocínio Sequencial Geral	Solucionar problemas com regras já estabelecidas
I	Indução	Descobrir a característica subjacente à solução
LS	Capacidade Auditiva	Ouvir e entender comunicações orais
P	Velocidade Perceptual	Identificar e comparar estímulos visuais
K2	Informação sobre Cultura	Amplitude do conhecimento cultural adquirido
LA	Aptidão para língua	Taxa de aprendizagem de uma nova língua
KF	Conhecimento de símbolos	Compreensão de símbolos falados ou gesticulados
Vz	Visualização	Identificação de estímulos visuais rotacionados
LE	Estimação de distância	Capacidade de estimar distâncias sem instrumentos
UM	Memória para padrões	Habilidade de reter eventos audíveis

Informação retirada de Schelini (2006)

1.3 Os testes de inteligência individuais

Dos testes de inteligência individuais as escalas Stanford-Binet juntamente com as escalas Wechsler são os testes individuais de inteligência mais amplamente utilizados no mundo (Oakland, 2005).

As escalas Stanford–Binet (SB)

O primeiro teste Stanford–Binet (SB) foi publicado em 1916. A primeira revisão do Stanford-Binet apareceu em 1937 (SB-1). A segunda revisão (SB-2) foi publicada em 1960: esta versão ainda usou as normas de 1937. A terceira revisão (SB-3) foi publicada em 1972. Em comparação com as normas de 1937, as normas de 1972 baseavam-se em uma amostra mais representativa, além de serem mais atualizadas.

Entre a revisão de 1937 e a revisão de 1972 se passaram 35 anos sem atualização das normas. Os autores do Stanford-Binet 3 (SB-3) notaram uma melhoria substancial nos escores brutos nas idades pré-escolares, equivalente ao um ganho de 10 pontos na escala de QI, ou seja, por volta de 0,3 ponto por ano. Flynn (2000) encontrou um efeito de aumento semelhante em escores brutos de outros testes de inteligência. O efeito Flynn implica no caso do Stanford-Binet um acréscimo 10,5 pontos de QI em 35 anos ($35 \times 0,3 = 10,5$). Uma consequência prática importante deste efeito é que pessoas testadas com uma versão anterior de teste de QI parecem mais inteligentes, enquanto pessoas testadas com uma versão atualizada de teste parecem menos inteligentes, uma vez que são comparadas com um padrão mais elevado.

A penúltima revisão do Stanford-Binet, a versão SB-4 (1987), sofreu uma revisão mais ampla em comparação com as revisões anteriores. A versão atual, o SB-5, foi publicada em 2006, nos EUA. O teste foi construído, como todas as versões anteriores, para pessoas entre 2 e 85 anos. Uma característica importante do SB-5 é o procedimento de testagem adaptativo. Nesse tipo de testagem cada indivíduo faz apenas aqueles itens cuja dificuldade é apropriada para o seu nível de desempenho demonstrado. Uma complicação de testagem adaptativa é que os itens precisam ser pontuados na hora, uma vez que isso define qual seria o próximo item a ser aplicado. Um procedimento adaptativo semelhante é utilizado nos testes não-verbais SON-R, o SON-R 2½-7 e o SON-R 6-40. (Tellegen & Laros, 2011).

Os testes Stanford-Binet são utilizados também como entrevistas clínicas uma vez que existem amplas oportunidades de interação entre o examinador e examinando. O valor dessas observações depende muito da habilidade e experiência do examinador.

O teste SB-5 mede cinco fatores (áreas cognitivas): (1) Inteligência Fluida; (2) Conhecimento; (3) Raciocínio Quantitativo; (4) Raciocínio Espacial / Visual e (5)

Memória de curto prazo. Foram realizados três tipos de estudos de validade do SB-5: (1) Análise Fatorial, (2) Correlações com outros testes de inteligência (validade convergente) e (3) Comparação do desempenho de grupos excepcionais (Hogan, 2006).

Hogan (2006) sugere que o SB-5 é um instrumento especificamente indicado para avaliar a inteligência de crianças talentosas, em virtude do alto teto proporcionado pelo intervalo de idade do teste. Por outro lado os revisores comentam que o teste não possui itens de solo suficientemente baixos para diagnosticar o retardo mental entre os testandos mais jovens. Em outras palavras, o SB-5 contém itens suficientemente difíceis mais não suficientemente fáceis. O teste Stanford-Binet não possui normas brasileiras.

As escalas de inteligência Wechsler

David Wechsler criou seu primeiro teste de inteligência, o *Wechsler-Bellevue Intelligence Scale*, em 1939 para medir a inteligência de adultos. Ele criou o Wechsler-Bellevue por que ele estava insatisfeito com o fato que o teste mais utilizado nessa época, o Stanford-Binet, fornecia apenas um escore global da inteligência. Na visão do Wechsler inteligência envolvia mais de uma única capacidade intelectual. Wechsler definiu a inteligência como sendo a capacidade agregada ou global que o indivíduo tem de agir com um propósito, de pensar racionalmente e de lidar com eficácia com seu ambiente (Hogan, 2006).

O *Wechsler Bellevue Intelligence Scale* fornece medidas separadas de QI Verbal, QI de Execução e QI Total. Em 1955, uma primeira revisão do *Wechsler-Bellevue Intelligence Scale* foi lançada sob o nome de *Wechsler Adult Intelligence Scale* (WAIS). A segunda revisão, o WAIS-R, foi publicado em 1981 e a terceira revisão, o WAIS-III, do teste apareceu em 1997. A última versão do *Wechsler Adult Intelligence Scale*, o WAIS-IV, foi publicada em 2008.

Wechsler criou também uma versão para crianças e adolescentes de 6 a 16 anos, chamado de *Wechsler Intelligence Scale for Children* (WISC). A primeira edição do teste foi publicada em 1949, a primeira revisão do teste foi publicada em 1974 sob o nome de WISC-R, e a segunda revisão do teste, o WISC-III, foi publicado em 1991 e o WISC-IV, publicado em 2003. Além dos testes WAIS e WISC, foi desenvolvido o *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI) um teste para as crianças de 2,5 a 7 anos. Em 1989 a primeira revisão desse teste foi publicada com o nome WPPSI-R. Em 2002 a segunda revisão do WPPSI foi publicada sob o nome de WPPSI-III. A última versão do WPPSI, o WPPSI-IV, foi publicada em 2012 (Wechsler, 2012).

Adicionalmente, foi elaborado uma versão não-verbal do teste Wechsler, o WNV (*Wechsler Nonverbal Scale of Ability*). Assim, existem atualmente os seguintes testes Wechsler: o WPPSI-IV, O WISC-IV, o WAIS-IV e o WNV. Na quarta edição do WPPSI, WISC e WAIS as seguintes medidas são fornecidas: Escala Geral, Compreensão Verbal, Raciocínio Perceptual, Memória de Trabalho e Velocidade de Processamento.

O coeficiente de fidedignidade Alfa de Cronbach do QI Geral (*Full Scale*) de todas as escalas Wechsler varia entre 0,90 as 0,98. Para as Escalas Verbais o coeficiente varia entre 0,86 e 0,97 e para as Escalas de Desempenho entre 0,85 e 0,94. Os manuais das escalas Wechsler relatam Erros Padrão de Mensuração (EPM) entre 2,5 e 5,0 para a Escala Verbal e entre 3,7 e 5,0 para a Escala de Desempenho. O EPM é fundamental para o cálculo de intervalos de confiança e na análise de perfil dos escores (Urbina, 2006).

Os revisores dos testes Wechsler criticam a maneira em que a validade é tratada nos primeiros testes: a validade nestes primeiros testes é tratada essencialmente como a validade de conteúdo (Hogan, 2006). Relativamente pouca atenção foi dada aos aspectos de validade de construto e validade critério. Isso já mudou nos manuais das versões III e

IV dos testes Wechsler. Foram elaboradas normas para o Brasil e fornecidas evidências da validade nesse país do WISC-IV (Wechsler, 2013) e do WAIS-III (Wechsler, 2004).

As Baterias Kaufman

Existem basicamente dois tipos de testes Kaufman: um para crianças, o K-ABC (*Kaufman Assessment Battery for Children*) e um para adultos, o KAIT (*Kaufman Adult Intelligence Test*). Os testes representam tentativas por parte dos seus autores de criar instrumentos ancorados nas teorias de inteligência.

O foco na construção do K-ABC foi a psicologia cognitiva, especificamente no processamento da informação. O K-ABC tem três subtestes que representam o Processamento Sequencial e sete subtestes que representam o Processamento Simultâneo. A bateria K-ABC também inclui uma Escala de Realização, compreendendo seis subtestes. Anastasi e Urbina (2000) criticam o nome de que os autores escolheram para a Escala de Realização porque na extremidade aptidão – realização esses testes estão muito mais próximos da extremidade aptidão. Os testes não medem o conhecimento factual ensinado na escola. O K-ABC foi padronizado nos Estados Unidos em uma amostra de 2.000 crianças de 2½ até 12½ anos.

Os revisores da escala sugerem que o K-ABC é um instrumento especificamente indicado para avaliar a inteligência de crianças com proficiência limitada nas habilidades verbais, em virtude do K-ABC depender menos das habilidades verbais.

O KAIT (*Kaufman Adult Intelligence Test*) foi planejado como uma medida de inteligência para as idades de 11 a 85 anos. O teste representa uma tentativa de integrar a teoria da inteligência fluida e cristalizada de Horn & Cattell com noções de inteligência de Piaget e Luria. A bateria KAIT é composta por uma Escala Cristalizada e uma Escala Fluida. O KAIT também tem uma Escala Expandida para ser usada com indivíduos com

suspeita de dano neurológico. As qualidades técnicas do KAIT têm características psicométricas satisfatórias tão bem como qualquer um dos testes da atual geração de testes. Lichtenberger e Broadbooks (2000, págs. 208-211) mostra que os índices de fidedignidade corrigidos por atenuação dupla variam entre 0,89 e 0,95 para os subtestes. O que realmente distingue os KAIT é o cuidado tomado para desenvolver e testar seu conjunto de mais de 2.500 itens. (Lichtenberger e Broadbooks, 2000)

Além do K-ABC e KAIT existe também o K-BIT (*Kaufman Brief Intelligence Test*), um teste abreviado. O K-BIT foi planejado como um instrumento de triagem rápida para estimar o nível de funcionamento intelectual. O K-BIT abrange o intervalo de idade de 4 a 90 anos. O K-BIT consiste de três escalas: a Escala Verbal, a Escala Não-Verbal e Escala Composta (Lichtenberger e Broadbooks, 2000). Não existem normas brasileiras para as baterias Kaufman.

As Escalas de Habilidade Diferencial - DAS

As escalas DAS (*Differential Ability Scales*) (Elliot, 1990) foi criado nos Estados Unidos como uma extensão e revisão do BAS (*British Ability Scales*). O instrumento foi planejado para pessoas entre 2½ e 18 anos de idade. O DAS é um instrumento novo com muitas características modernas, mas refletindo o conhecimento e a experiência obtidos durante várias décadas com o BAS. O que distingue as escalas DAS de outros testes de inteligência é que no manual os autores afirmam que os termos QI e inteligência não fazem parte do vocabulário das escalas. Anastasi & Urbina (2000) afirmam a este respeito que essa declaração dos autores do DAS deve ajudar a desfazer os estereótipos e as concepções errôneas associadas ao uso popular destes termos. Atualmente, existe a versão DAS-II (Elliot, 2007) que está sendo comercializada, possui 20 subtestes que avaliam a

inteligência segundo o modelo CHC em pessoas entre 2 anos e 6 meses e 17 anos e 11 meses de idade,

O DAS é incomum no sentido que o teste corporifica muitos avanços na teoria e na prática psicométrica inexistentes em outras escalas. O DAS emprega uma estrutura hierárquica que é compatível com os achados empíricos sobre o desenvolvimento das habilidades cognitivas e usa um procedimento de testagem adaptativa que é baseado em um modelo de Teoria de Resposta ao Item (TRI) de um parâmetro para permitir a calibração dos valores de dificuldade para cada item. A principal vantagem da testagem adaptativa é a flexibilidade que permite ao examinador selecionar subconjuntos de itens apropriados a cada examinando. Nas idades entre 6 anos e 17 anos e 11 meses, um modelo de três fatores se ajustou melhor aos dados: Habilidade Conceitual Geral, Habilidade de Raciocínio Não-verbal e Habilidade Espacial.

A complexidade dos procedimentos de aplicação e pontuação do DAS pode atrapalhar sua disseminação e seu uso em ambientes de aplicação (Hogan, 2006). Uma outra limitação é no grupo de idade mais baixo os itens não são suficiente fáceis, enquanto no grupo de idade mais alto os itens não são suficiente difíceis. Não obstante o DAS é um instrumento de vanguarda. Anastasi & Urbina elogiam muito a qualidade do manual do DAS. Não existem normas brasileiras para as Escalas de Habilidade Diferencial.

1.5 Os testes SON

Os testes SON surgiram na Holanda na década de 40, e as letras que compõem seu nome são uma homenagem à sua criadora: a prof.^a Nan Snijders-Oomen. A primeira versão do teste surgiu no ano de 1943 e tinha como população-alvo crianças surdas cujas idades variavam entre 4 e 14 anos. Quase duas décadas depois da sua construção, surgiu uma

segunda versão com normas mais atualizadas e com a possibilidade de aplicação para crianças ouvintes e com idades de até 16 anos. Essa versão chamou-se SON-'58, pois surgiu no ano de 1958. Após essas duas versões, o teste diversificou-se em duas vertentes: aqueles com participantes cujas idades iriam até a primeira infância e aqueles em que a idade dos participantes poderia variar desde a primeira infância até a adolescência ou adultez. A última versão do SON-R, o SON-R 6-40, será utilizada na presente pesquisa e já foi normatizada no país de origem (Tellegen & Laros, 2011), e está sendo normatizada no Brasil. O diagrama na página seguinte abaixo mostra como os testes SON se desenvolveram cronologicamente, bem como as versões que foram criadas.

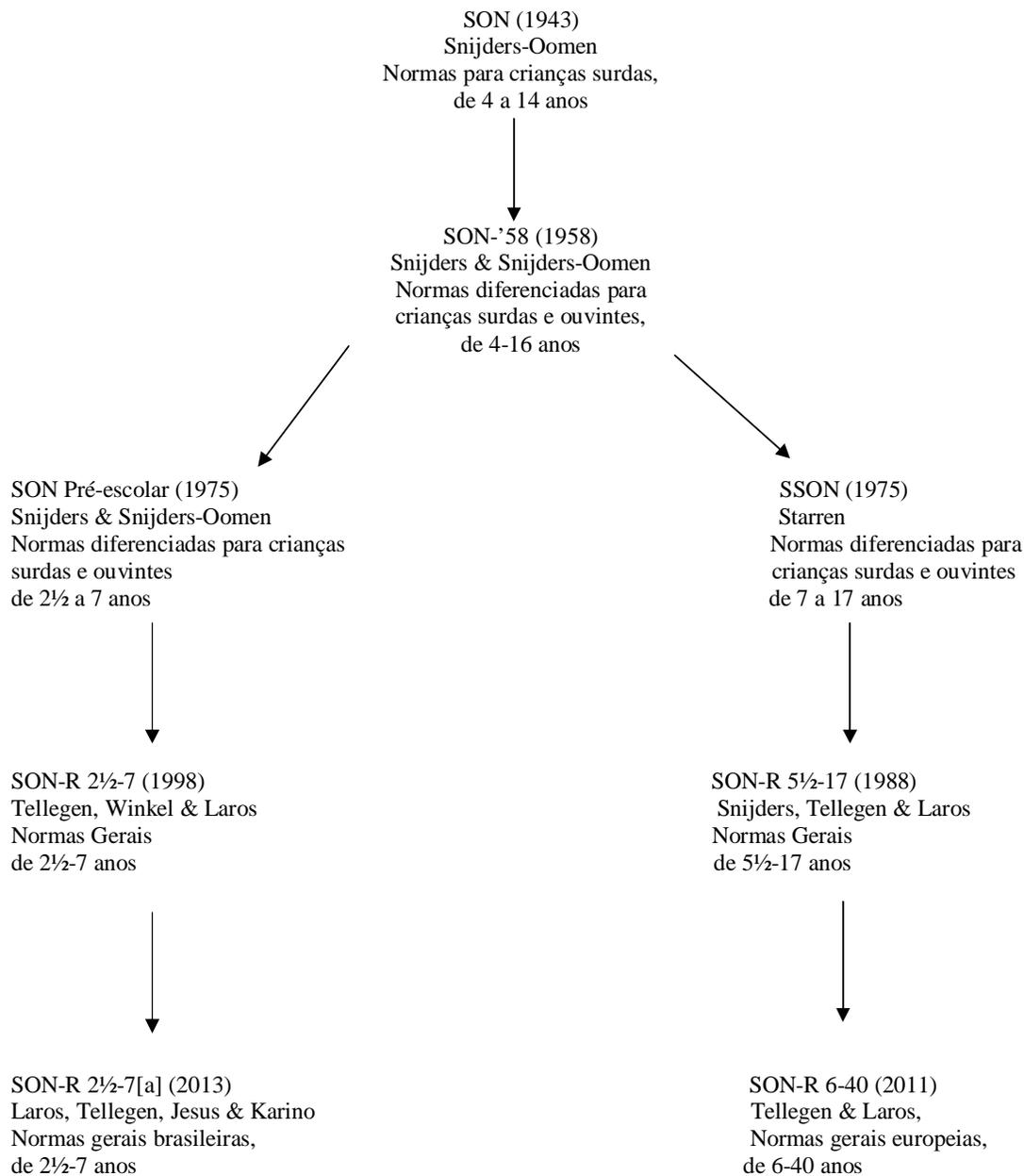


Figura 1.1 Histórico dos testes SON, retirado de Jesus (2009), com adaptações.

Em seu país de origem, foram realizados alguns estudos, dentre eles estudos que buscavam obter evidências de validade convergente com outros testes tais como WAIS-III e WISC-III. Os resultados encontrados mostram correlações altas de 0,79 entre os escores gerais do SON-R 6-40 e o WAIS-III. A correlação entre o escore geral do SON-R

6-40 e a Escala de Desempenho e Escala Verbal do WAIS-III foi 0,85 e 0,64 respectivamente. A correlação de 0,64 chama a atenção para o fato que um teste não-verbal de inteligência é capaz de explicar uma boa parte da variância de uma escala que pretende medir a inteligência verbal. Quanto ao teste WISC-III, o escore geral do SON-R 6-40 obteve um coeficiente de correlação de 0,64. Enquanto para os escores obtidos na escala de desempenho e na escala verbal do WISC-III, os resultados das correlações obtidas foram de 0,71 e 0,46, respectivamente.

No Brasil, há uma versão do SON já normatizada (Jesus, 2009), o SON-R 2½-7[a]. O SON-R 2½-7[a] é a versão abreviada do SON-R 2½-7, a última versão dos testes SON para crianças entre 2 anos e meio e 7 anos. Esse instrumento de origem holandesa foi normatizado e validado em vários países de Europa consiste em seis subtestes, a saber: Categorias, Analogias, Situações, Histórias, Mosaicos e Padrões. A versão abreviada do teste composta pelos subtestes Categorias, Situações, Mosaicos e Padrões.

No estudo de normatização do SON-R 2½-7[a] para o Brasil, os pesquisadores utilizaram uma amostra representativa de 1.200 crianças cujas idades variaram entre 3 anos e 3 meses e 7 anos e 9 meses de idade. A amostra para o estudo de normatização do SON-R 2½-7[a] para o Brasil foi composta por 1.200 crianças, divididas equitativamente quanto à idade e ao sexo. Foram estabelecidos 10 grupos de idade entre 3 anos e 3 meses a 7 anos e 9 meses, cada grupo constituído por 120 crianças, sendo tolerável uma variação de três meses em relação à idade pretendida. Não foram incluídas na amostra crianças com deficiência auditiva ou motora. Ao todo, a pesquisa foi realizada nas cinco regiões brasileiras, em 13 estados diferentes, contemplando 36 cidades. O SON-R 2½-7[a] foi aprovado pelo Conselho Federal de Psicologia em 2012 como um teste psicológico em condições de uso profissional.

Os seguintes testes foram aplicados para obter evidências da validade convergente do SON-R 2½-7[a] no Brasil: o WISC-III (Escala de inteligência Wechsler para Crianças), WPPSI-III (*Wechsler Preschool and Primary Scale of Intelligence – Third Edition*), Matrizes Progressivas Coloridas de Raven e a Escala de Maturidade Mental Columbia. (Karino, Laros, e Jesus, 2011; Laros, Reis, e Tellegen 2010).

Um estudo avaliando o Funcionamento Diferencial dos Itens (DIF) da bateria SON-R 2½-7[a] foi realizado por Karino, Laros e Jesus (2012). Esse trabalho contou com a análise dos 1.200 escores da amostra de normatização do SON-R 2½-7[a]. Os resultados demonstraram que apenas 5 itens demonstraram DIF para as regiões em que o teste foi aplicado, enquanto que 13 itens apresentaram DIF entre os sexos.

Os resultados da pesquisa de normatização e validação do teste SON-R 2½-7[a] demonstram índices de consistência interna de 0,92 para a Escala Geral e de 0,87 para as Escala de Execução (os subtestes Mosaicos e Padrões) e Escala de Raciocínio (Categorias e Situações). O subteste Categorias demonstrou o nível de fidedignidade mais alto, 0,81, enquanto o nível mais baixo foi encontrado no de Situações, 0,77. No primeiro estudo de validade convergente, o SON-R 2½-7[a] e o WPPSI-III obtiveram uma correlação de 0,75 entre seus escores de QI total, enquanto a menor correlação do estudo ocorreu entre a escala verbal do WPPSI-III e o escore de QI total do SON-R 2½-7[a] ($r = 0,66$). Ressalta-se a relação dos dois testes com o nível socioeconômico (NSE). A correlação entre o SON-R 2½-7[a] e o NSE foi de 0,48 e entre o WPPSI-III e o NSE foi de 0,67. Assim, o SON-R 2½-7[a] compartilhou 23% da variância com o NSE e o WPPSI-III, 45%. Esses resultados sugerem que em comparação com o WPPSI-III, o SON-R 2½-7[a] sofreu menos influência do nível socioeconômico.

O estudo de validade convergente envolvendo o teste WISC-III mostrou uma correlação de 0,67 entre o SON-R 2½-7[a] e o WISC-III. A correlação entre o SON-R

2½-7[a] e a Escala de Execução do WISC-III foi de 0,64 e com a Escala Verbal, 0,51. O teste Matrizes Progressivas Coloridas de Raven e o SON-R 2½-7[a] mostraram uma correlação entre eles de 0,68. Por fim, o SON-R 2½-7[a] também foi aplicado em conjunto com a Escala de Maturidade Mental Columbia, e a correlação encontrada de 0,56 sugere que os testes podem estar avaliando aspectos diferentes de inteligência.

Um teste não-verbal de inteligência possui algumas vantagens sobre aqueles em que a linguagem do testando é um fator importante para sua aplicação. É possível citar algumas, sem saturar as opções: (1) é possível realizar estudos transnacionais, ou seja, aplicar o mesmo teste em diferentes países sem necessidade de tradução, usando apenas os estímulos contidos no instrumento, (2) a aplicação pode ser realizada em populações com necessidades especiais, como pessoas com deficiência auditiva, tendo em vista que o mesmo é independente de instruções verbais, (3) pessoas que apresentam dificuldades de comunicação ou que não falam o mesmo idioma do aplicador também podem realizar o teste.

1.6 O teste BPR-5

A Bateria de Provas de Raciocínio BPR-5 (Primi & Almeida, 2000) pretende avaliar habilidade cognitivas por meio da utilização de 5 subtestes: Prova de Raciocínio Verbal (RV), Prova de Raciocínio Abstrato (RA), Prova de Raciocínio Numérico (RN), Prova de Raciocínio Mecânico (RM) e a Prova de Raciocínio Espacial (RE). Esse teste é uma versão da Bateria de Provas de Raciocínio Diferencial (Almeida, 1986, 1988), sendo este último idealizado a partir dos Testes de Raciocínio Diferencial de Meuris (1969). (Baumgartl & Nascimento, 2004).

O primeiro estudo de validação da escala (Primi & Almeida, 2000) contou com a participação de 1.243 alunos brasileiros e portugueses, sendo que a maioria dos

estudantes foi brasileira (62%) e do sexo feminino (64%). Quanto a precisão, os subtestes apresentaram altos índices que variaram entre 0,77 (Prova RM, forma B) e 0,94 (Prova RN, Forma A). Quanto aos coeficientes de validade, os subtestes que apresentaram maiores correlações entre si foram a prova RV e RA, para a Forma A ($r = 0,50$), enquanto para a forma B, os subtestes RE e RN apresentaram a correlação mais alta ($r = 0,60$). Ainda, obteve-se índices de validade de critério através da comparação entre os escores no BPR-5 e as notas escolares dos participantes. Os maiores índices ocorreram entre o subteste RV e as notas de Matemática ($r = 0,53$) para a forma A, enquanto na forma B, o maior índice surgiu entre a prova RE e as notas em Desenho Técnico Científico ($r = 0,66$).

O segundo estudo de validação contou com 411 alunos portugueses, sendo em sua maioria (52,8%) do sexo feminino, cursando o 10º ano (equivalente ao 1º ano do ensino médio) (34,5%), e que responderam à forma B do teste. Foram obtidas correlações entre a Forma A do BPR-5 e as notas escolares e o maior índice obtido ocorreu entre a prova RV e as notas em Português ($r = 0,55$). Para a forma B, o maior índice esteve entre os escores da prova RN e as notas de Matemática ($r = 0,36$).

Para a padronização do teste, foram coletados 1.722 dados (603 para a forma A e 1119 para a forma B), em sua maioria do sexo feminino (55 %), com 16 anos de idade, cursando o 2º ano do ensino médio (19%). Os maiores coeficientes de fidedignidade ocorreram para a prova de Raciocínio Numérico tanto para a Forma A (0,91), quanto para a Forma B (0,89).

Muitos estudos utilizando o BPR-5 foram realizados desde o seu surgimento. Baumgartl e Nascimento (2004) realizaram um estudo com a bateria em um contexto organizacional. Nesse estudo, participaram 57 indivíduos cujas idades variavam entre 19 e 45 anos com escolaridade mínima de Ensino Médio completo. Além disso, dividiram a

amostra em 3 grupos por idade, 2 grupos por atividade exercida e 4 grupos por escolaridade. Aplicou-se a bateria na forma B e percebeu-se que em relação às funções exercidas os resultados não variavam excessivamente, o grau de escolaridade era diretamente proporcional ao escore apresentado e a idade possuía uma relação inversa, na qual quanto maior a idade do sujeito maior a tendência a tirar uma nota menor.

Em outro estudo (Cobêro, Primi & Muniz, 2006), buscou-se avaliar as correlações existentes entre inteligência emocional e desempenho no trabalho. Para esse estudo, foi utilizada uma amostra de 119 indivíduos trabalhadores de empresas do interior de São Paulo, em sua maioria do sexo masculino e com idade média de 30 anos. Juntamente ao BPR-5 foram utilizados os instrumentos MSCEIT (*Mayer, Salovey, Caruso Emotional Intelligence Test*) e o 16PF (*16 Personality Factors*). Ao realizar uma análise de regressão para avaliar a validade incremental da inteligência emocional em relação ao desempenho no trabalho, foi encontrada uma correlação baixa entre o MSCEIT e o BPR-5, variando entre -0,10 até 0,40 e entre o 16PF e o BPR-5 foi encontrada correlação significativa apenas para o fator inteligência (Fator B), como era esperado.

Um estudo também foi realizado para avaliar possíveis relações entre escolaridade, idade e inteligência (Primi, Couto, Guisande, Almeida & Miguel, 2012). Nesse trabalho, investigou-se a influência que Gf e Gc sofrem devido à escolaridade e à idade dos indivíduos. Contou-se com a participação dos 1.722 estudantes que constituíram a padronização brasileira do teste. Em uma análise descritiva, constatou-se que a média era, em geral, maior para os alunos que estudavam em escolas particulares do que para aqueles que estudavam em escolas públicas, em todos os subtestes da escala.

Um estudo atual com a escala BPR-5 (Almeida & cols., 2010) teve como objetivo principal a realização de um estudo exploratório da escala. Nesse estudo, utilizou-se uma amostra de 330 estudantes universitários cursando o segundo ano de Psicologia, sendo a

maioria do sexo feminino (63,3%) com idade aproximada de 21 anos. Para esse estudo aplicou-se junto com o BPR-5, as escalas Matrizes Progressivas Coloridas de Raven e o G-36. Análises descritivas mostraram que o teste de Matrizes Progressivas apresentou maior média de acertos entre os testes respondidos, enquanto o subteste Raciocínio Mecânico apresentou o menor índice. Ainda, os resultados de um estudo correlacional mostrou que as Matrizes Progressivas demonstram maior índice de correlação com o subteste Raciocínio Abstrato ($r = 0,30$), enquanto o G-36 teve maior correlação com o subteste Raciocínio Espacial ($r = 0,30$). Uma análise fatorial exploratória também foi realizada e demonstrou a existência de um fator único para a escala, sugerindo que a mesma avaliava inteligência fluida (Gf), principalmente quando se analisavam os subtestes Raciocínio Mecânico, Espacial e Abstrato. Por outro lado, os testes Raciocínio Verbal e Numérico tendem a constituir um segundo fator, sugerindo que eles medem a inteligência cristalizada (Gc).

1.7 Os conceitos de validade e validação

As questões concernentes à validade e validação na avaliação psicológica são, e ainda serão por algum tempo, geradoras de grandes controvérsias para a psicologia. Alguns teóricos aceitam apenas um tipo de validade como sendo correta (Anastasi, 1999), enquanto outros admitem a existência de outros tipos de validade, em geral três tipos: validade de conteúdo, validade de construto e validade de critério. Os três conceitos serão discutidos de forma breve nessa dissertação, para maiores informações ver Lissitz (2009) e Anastasi & Urbina (2000).

O primeiro tipo de validade é a validade de conteúdo. Segundo essa perspectiva, os itens de um teste devem ter a capacidade de avaliar, em determinado grau, o domínio que o mesmo pretende avaliar. Testes escolares são bons exemplos de instrumentos com os

quais se pretende possuir boas evidências de validade de conteúdo. Para que sejam encontradas tais evidências, o pesquisador deveria construir itens juntamente com outros experts no assunto a ser tratado, e esses pesquisadores deveriam chegar a um consenso sobre a parcela específica do universo de domínio que tais itens conseguem mensurar.

Quando se fala em validade de critério, busca-se compreender se os resultados apresentados em um determinado teste seriam replicáveis em situações do mundo real, que não apenas as do momento da testagem. Através de um critério, tal como desempenho em uma situação de trabalho real, é possível comparar os escores do teste com os resultados obtidos por um indivíduo nessa situação. Para que possa ser realizado um estudo de validade de critério, alguns critérios devem ser seguidos (Crocker & Algina, 2006). Inicialmente, é importante a identificação de um comportamento como critério de avaliação e uma maneira de medi-lo. Tais comportamentos podem ser relacionados às tarefas que seriam desempenhadas no ambiente de trabalho e poderiam ser medidos através de observação, por exemplo. O segundo passo seria identificar uma amostra com características semelhantes para as pessoas em que o teste seria utilizado em situações cotidianas. Tal amostra poderia ser constituída de colegas de trabalho ou estudantes de uma mesma sala de aula. Logo após a identificação da amostra, é necessário aplicar o teste e registrar os escores de cada participante. Quando os testes forem corrigidos, e os resultados estiverem disponíveis é necessário comparar o desempenho de cada examinando. Finalmente, é importante que o pesquisador determine a força das relações entre os escores do teste e o desempenho na tarefa critério. As análises da validade de critério são, em grande parte, determinada pelo pesquisador e seu interesse no construto a ser avaliado. Por exemplo, o pesquisador pode estar interessado em avaliar raciocínio lógico em determinada tarefa, utilizada como critério de uma situação de dinâmica de

grupo e em seguida aplicar um teste que mensure o mesmo constructo. Para o pesquisador, cada critério terá um peso na sua avaliação.

A validade de construto é o método mais aplicado em estudos de validade atualmente, em especial o procedimento de análise fatorial, devido às suas sofisticadas metodológicas e a facilidade com a qual pode ser realizada. Tal método é inicialmente realizado através da formulação de hipóteses, por parte do pesquisador, que envolvam dados baseados em uma teoria previamente conhecida e com um constructo bem definido para mensurar. O segundo passo para a realização da validade constructo é a seleção de um instrumento que envolva itens que representem comportamentos representativos do processo cognitivo a ser estudado (p. ex., inteligência). Após essa formulação, avaliar se os dados empíricos podem ser comparados com os achados teóricos. E finalmente, testar se a hipótese inicial do pesquisador pode ser confirmada ou não pelos dados obtidos. Existem quatro formas para avaliar a validade de construto de um instrumento: Análise fatorial, matriz multitraço-multimétodo, diferenciação entre grupos e as correlações existentes entre testes que se propõem a avaliar o mesmo constructo.

A análise fatorial é o método mais conhecido e aceito atualmente pela comunidade acadêmica. Sua complexidade metodológica e interpretativa são consideradas barreiras para os psicólogos para sua utilização (Pasquali, 2012). A realização de uma análise fatorial envolve conhecimentos de matemática matricial, métodos multivariados de análise de dados e tem como principal objetivo a identificação dos fatores que o teste em questão mensura. Em resumo, a análise fatorial pode ser explicada nos seguintes passos: é necessário encontrar as correlações entre os itens do teste inicialmente, para que os itens que se correlacionem melhor com outros possam fazer parte de um mesmo fator, enquanto outros possam constituir outros fatores. Logo após encontrar as correlações é necessário estabelecer critérios para que sejam extraídos fatores. Para esses critérios,

deve-se criar gráficos, chamados *scatter plots*, nos quais as correlações entre cada item será expressa por um ponto específico, e a distribuição dessas correlações auxiliará o pesquisador a realizar suas primeiras interpretações com base na teoria estipulada. Após a realização dessa análise gráfica, o pesquisador deve analisar os fatores quanto a teoria, e avaliar se os fatores encontrados possuem lógica baseados nos preceitos teóricos.

A matriz multitraço-multimétodo é um método sofisticado para medir a validade de construto. Proposta por Campbell e Fiske (1959), a matriz multitraço-multimétodo fornece dados tanto sobre um mesmo constructo avaliado de formas diferentes, quanto de constructos diferentes avaliados de uma forma semelhante. Utiliza-se uma mesma amostra de sujeitos para avaliar diferentes constructos e de diferentes formas, assim as correlações entre cada par de medidas são computadas. Por exemplo, o pesquisador poderia estar interessado em avaliar proficiência em língua estrangeira (inglês, francês e espanhol) de um grupo de cinco estudantes através de dois métodos diferentes (redação e ditado). A construção dessa matriz envolve a demonstração de como diferentes métodos podem se correlacionar com as variáveis a serem medidas, assim mostrando que, ao menos teoricamente, um métodos pode ser mais influente que outro na aquisição de escores.

O terceiro método corresponde à busca por diferenças perceptíveis entre grupos quanto aos escores encontrados. Um exemplo de como ocorre esse procedimento é a demonstração do quanto meninos e meninas diferem em seus resultados ao considerar a teoria na qual a hipótese está baseada. Geralmente, esses estudos são realizados de forma experimental, para demonstrar se determinadas mudanças, seja no controle de alguma característica do estudo ou na aplicação de diferentes métodos, podem influenciar de maneira significativa nos resultados.

Finalmente, o último método ocorre quando se busca por correlações entre um teste que já possui evidências de validade para avaliar determinado constructo e um que ainda não possui provas de que avalie tal constructo. Esse tipo de validade também é chamado de validade convergente e é realizado de forma simples, através da demonstração da formulação de hipóteses em que se propõe que ambos os testes possuam altos índices de correlação, e assim possam avaliar o mesmo constructo com mais fidedignidade para o pesquisador. O presente trabalho utilizou esse procedimento para avaliar as evidências de validade convergente entre os testes SON-R 6-40, SON-R 2½-7[a] e o BPR-5.

Porém, a partir de um ponto de vista mais complexo, é possível notar que o conceito de validade está inserido dentro de um processo. E esse processo é o que chamamos de validação. O conceito de validade de construto está relacionado à questão se o teste mede os fatores que ele se propõe a medir (i.e., inteligência, atenção, depressão, etc.) e em que nível o faz (i.e., baixas ou altas cargas fatoriais em cada item ou entre os itens de um mesmo teste). Esse conceito é um dos mais estudados atualmente e também o que apresenta maiores complicações para os pesquisadores devido à sua sofisticação matemática (Lissitz, 2009).

É importante salientar a diferença entre validade e validação. O primeiro termo refere-se às características dos escores de determinados testes e mostra em que grau o teste é apropriada para ser aplicado em um determinado contexto. Já o conceito de validação engloba todo o processo envolvido para que se encontrem bons indícios de validade.

Ao falar-se de validade, deve-se falar em evidências ou indícios de validade, isso se deve ao fato de que bons escores para um teste em determinado contexto podem representar com muita ou pouca fidedignidade as características desejadas da amostra selecionada. A validação, segundo Zumbo (2007), é um processo que busca evidências de

validade, baseando-se na adequação, significado e utilidade dos escores de um teste em determinadas situações e contextos. O conceito de validação deve se pautar e três características principais ou centrais. A primeira diz respeito às conclusões que são tiradas da aplicação e dos resultados dos testes. Os escores de um teste são o principal fator que mostra a qualidade das suas evidências de validade. Logo, a possibilidade de diferentes escores, em diferentes momentos ou em diferentes contextos de aplicação, pode servir de alicerce para diferentes formas de interpretação.

A segunda característica central da validação diz que testes e avaliações de qualquer natureza devem passar pelo processo, uma vez que para os dados referentes ao processo podem ser atribuídos diferentes graus de significância entre pesquisadores, a depender dos pontos de vista defendidos.

E, por fim, os pesquisadores na área de validade e validação devem se preocupar com o grau em que as inferências retiradas de um teste podem ser valoradas para um determinado contexto. Por exemplo, em uma amostra homogênea, como a população de alunos de uma mesma escola, é possível que sejam encontrados bons índices de validade devido à baixa variabilidade de idades ou diferenças socioeconômicas.

1.7 Referencial Teórico

A presente dissertação utiliza como referencial teórico a teoria hierárquica de inteligência CHC (Carroll-Horn-Cattell) (Flanagan & Harrison, 2012). Essa teoria é uma revisão integrada de duas teorias anteriores: a teoria de três estratos (Carroll, 1993) e a teoria de inteligência fluida e cristalizada (Gf-Gc) (Cattell, 1943).

A primeira teoria afirma que existem três estratos nos quais diversos fatores estão dispostos segundo suas correlações com um fator amplo ou geral. No topo da hierarquia está o fator geral (g) e logo abaixo dele estão os fatores amplos, que são altamente

correlacionados com o fator geral. E, por fim, estão presentes os fatores estreitos, que são fatores que não se correlacionam diretamente com o fator geral, mas sim com os fatores do estrato dois.

Segundo a teoria de inteligência fluida e cristalizada existem dois fatores amplos e correlacionados. O primeiro estaria relacionado à capacidade de lidar com situações novas para o examinando, através da resolução de problemas que não envolvam vieses culturais. Em geral, testes que avaliam a inteligência fluida possuem natureza não-verbal, e podem ser tarefas que envolvem lidar com estímulos desconhecidos aos participantes. A inteligência cristalizada estaria envolvida na realização de tarefas nas quais os indivíduos devem evocar conhecimentos previamente adquiridos com o intuito de responder a itens de um determinado domínio do conhecimento. Os testes SON foram construídos para avaliar principalmente a inteligência fluida, fato este que justifica a utilização da teoria citada anteriormente.

2. Objetivos

A presente dissertação tem por objetivo investigar evidências de validade convergente da bateria não-verbal de inteligência SON-R 6-40. Para poder alcançar o objetivo geral, os seguintes objetivos específicos serão realizados:

- 1 – Analisar os parâmetros psicométricos das baterias SON-R 6-40, SON-R 2½-7[a], forma A da BPR-5 e dos subtestes dessas baterias.
- 2 – Avaliar a validade convergente entre as baterias SON-R 6-40 e SON-R 2½-7[a], para crianças entre 6 e 7 anos.
- 3 – Avaliar a validade convergente entre as baterias SON-R 6-40 e a forma A da BPR-5, para adolescentes entre 12 e 16 anos;

3. Método

Em estudos de validade de um teste psicológico é recomendável aplicar mais de uma medida que avalia o mesmo construto teórico. Por esse motivo, optou-se pela realização de dois estudos: no primeiro estudo o SON-R 6-40 (Tellegen & Laros, 2011) foi aplicado junto com SON-R 2½-7[a] (Laros , Tellegen, Jesus & Karino, 2012), enquanto que no segundo estudo aplicou-se o SON-R 6-40 junto com o BPR-5 (Primi & Almeida, 2000).

O primeiro estudo de validação foi realizado com o teste SON-R 2½-7[a] devido a sua similaridade com o teste SON-R 6-40 em relação aos itens e forma de aplicação.

Existem duas diferenças principais entre os testes SON: a faixa etária e os subtestes. O SON-R 2½-7[a] avalia crianças com idades variando entre 2 anos e 6 meses e 7 anos e 11 meses, enquanto o SON-R 6-40 avalia pessoas entre 6 anos e 40 anos. A segunda diferença na composição das duas baterias: o SON-R 6-40 é composto dos subtestes Analogias, Mosaicos, Categorias e Padrões enquanto e o SON-R 2½-7[a] dos subtestes Mosaicos, Categorias, Situações e Padrões.

O segundo estudo de validação realizou-se com a forma A do teste Bateria de Provas de Raciocínio (BPR-5). O BPR-5 foi escolhido porque avalia pessoas com idades entre 13 e 18 anos, uma faixa etária que também é avaliada pelo SON-R 6-40. O BPR-5 é composto por cinco subtestes, a saber, Raciocínio Verbal, Raciocínio Espacial, Raciocínio Abstrato, Raciocínio Mecânico e Raciocínio Numérico. Entretanto, o subteste de Raciocínio Mecânico não foi incluído no estudo de validação uma vez que a sua aplicação não foi possível dentro de tempo disponível fornecido pelas escolas.

3.1 Estudo 1 – SON-R 6-40 e SON-R 2½-7[a]

Participantes

A amostra pesquisada foi constituída de 44 alunos vindos de diferentes escolas públicas do DF, sendo 23 do sexo masculino (52,3%) e 21 do sexo feminino (47,7%). Esses alunos possuíam idades variando entre 5,9 e 7,8 anos de idade ($M = 6,9$; $DP = 0,47$) no momento da aplicação do teste. Um fator relevante neste contexto é a duração de aplicação dos testes: para administrar o SON-R 6-40 junto com SON-R 2½-7[a] são necessários aproximadamente duas horas para cada participante. Finalmente, existiram dificuldades com um número elevado de pessoas na aplicação de ambos os testes. Nesses casos a segunda aplicação não foi realizada, e os participantes não puderam ser incluídos na amostra final. Todas as aplicações do primeiro estudo de validação ocorreram em escolas públicas da região da Asa Norte no Distrito Federal.

Instrumentos

O primeiro estudo foi realizado com os instrumentos descritos em seguida:

- O teste SON-R 6-40 (Laros & Tellegen, 2010): uma bateria não-verbal de aplicação individual com duração média de 60 minutos. A população-alvo dos respondentes do teste consistiu de pessoas entre 6 e 40 anos. O SON-R 6-40 busca avaliar a inteligência fluida como descrita pela teoria de Carroll (2005). O teste é composto de quatro subtestes: Analogias e Categorias, que avaliam raciocínio abstrato e Mosaicos e Padrões que avaliam raciocínio espacial. No SON-R 6-40, cada subteste possui 26 (Padrões e Mosaicos) ou 36 itens (Analogias e Categorias), divididos em 2 (Padrões e Mosaicos) ou 3 séries (Analogias e Categorias). Essa divisão surgiu com base no conceito de testagem adaptativa. Segundo a testagem adaptativa, o respondente não necessariamente faz o teste inteiro, mas sim apenas os itens que ele consegue realizar segundo um padrão de acertos e erros específicos. No SON-R 6-40, por exemplo, cada

série termina após o sujeito ter 2 erros consecutivos ou em qualquer ordem. Os subtestes Mosaicos e Padrões tem um tempo limite na aplicação. No subteste Analogias a tarefa consiste em descobrir qual transformação ocorreu na primeira figura e aplicar a mesma transformação na segunda figura. No subteste Categorias o sujeito precisa descobrir o conceito subjacente aos três desenhos apresentados. No subteste Mosaicos, diversas figuras de mosaicos precisam ser copiados numa moldura utilizando quadrados vermelhos, brancos e vermelho-brancos. No subteste Padrões o sujeito deve completar um padrão formado por uma ou duas linhas em que uma parte está faltando. Os escores do SON-R 6-40 são normatizados numa escala de QI com média 100 e desvio-padrão de 15. Existe um programa de computador que pode ser utilizado para calcular os escores normatizados do SON-R 6-40 sem uso das tabelas de normas do manual. As principais vantagens do programa de computador são: (a) os erros no cálculo dos escores normatizados são evitados; (b) há um ganho considerável em tempo; (c) são oferecidas informações extras, que não se encontram nas tabelas e (d) os resultados são mais precisos porque são baseados na idade exata. Com o uso de programa de computador é fornecida informação sobre a idade de referência de uma pessoa. A idade de referencia representa o desempenho médio da faixa etária normativa à qual corresponde. Maiores informações sobre o teste SON-R 6-40 e sobre o programa de computador podem ser encontradas no site criado pelos autores (www.testresearch.nl).

- O teste SON-R 2½-7[a]: uma bateria composta por quatro subtestes, a saber, Mosaicos, Categorias, Situações e Padrões. A população-alvo do teste são crianças entre 2 anos e 6 meses e 7 anos e 11 meses de idade. O SON-R 2½-7[a] foi normatizado e validado para o Brasil (Laros, Tellegen Jesus & Karino, 2012) e tem duração de aplicação aproximada de uma hora. É possível avaliar os resultados através de três escores

diferentes: escore de QI da Escala de Execução, escore de QI da Escala de Raciocínio e escore de QI total. Os três escores estão numa escala de QI (100,15).

Procedimento

Para a aplicação dos testes, utilizou-se do apoio de uma equipe de alunos bolsistas de iniciação científica e estagiários de bacharel. As principais atribuições da equipe foram: entrar em contato com as escolas, auxiliar na aplicação dos testes individuais e ajudar na construção do banco de dados.

Os testes foram aplicados nas escolas cujos diretores permitiram a entrada da equipe para a realização dos mesmos. Como os testes eram de aplicação individual, os alunos eram retirados de sala de aula apenas se não estivessem fazendo atividades curriculares ou que tivessem terminado as atividades solicitadas pelo professor.

Os diretores das escolas disponibilizaram salas de aula, ou salas de recursos que não estavam sendo utilizadas no momento das aplicações. Se as escolas tinham uma sala de informática, as aplicações foram realizadas neste local. Os alunos realizavam os testes solicitados com duração média de uma hora, geralmente com dois ou mais aplicadores na mesma sala, para que fosse possível realizar ao menos duas aplicações diárias.

3.2 Estudo2 – SON-R 6-40 e BPR-5

Participantes

Para a realização do segundo estudo envolvendo o SON-R 6-40 e o BPR-5, ambos os testes foram aplicados em 68 pessoas. Desses, 36 (52,9%) eram do sexo masculino, enquanto 32 (47,1%) do sexo feminino. A idade dos participantes variou entre 11,7 anos e 16,1 anos ($M = 14,2$; $DP = 1,17$). Os participantes foram oriundos de escolas da rede pública do Distrito Federal, da região da Asa Norte, em sua totalidade.

Instrumentos

Os instrumentos para a realização desse segundo estudo são listados e descritos nos parágrafos seguintes:

- O teste SON-R 6-40 (Laros & Tellegen, 2010): a descrição deste instrumento foi feita no Estudo 1 (pág. 28-29).

- O BPR-5 (Almeida & Primi, 2000): O teste BPR-5 é uma bateria de provas que avaliam cinco diferentes tipos de raciocínio: raciocínio verbal, raciocínio abstrato, raciocínio espacial, raciocínio mecânico e raciocínio numérico. No presente estudo, optou-se na retirada do um dos cinco subtestes, Raciocínio Mecânico, devido ao alto grau de dificuldade de seus itens para a amostra estudada. Sendo assim foram administradas quatro provas do BPR-5. A bateria completa tem duração média de 75 minutos e pode ser aplicada em grupo. O teste possui duas formas: forma A e forma B. A forma A é utilizada para estudantes cursando as séries finais do ensino fundamental e cuja idade varia entre 13 e 18 anos. Já a forma B é aplicada em participantes cursando o ensino médio. Optou-se pela utilização da forma A da bateria, pois ela se adequava à amostra estudada. Os escores normatizados são fornecidos como EPN (Escore Padrão Normatizado: M=100; DP=15).

Procedimentos

Os testes foram aplicados em escolas públicas do DF entre os meses finais de 2012 e os primeiros meses de 2013. Foram selecionadas escolas que possuíam os anos finais do fundamental, a saber a 7^a, 8^a e 9^a anos, pois a aplicação do BPR-5 exigia participantes nessa escolaridade.

Para possibilitar a aplicação dos testes, os alunos levaram um Termo de Consentimento Livre e Esclarecido para os seus pais. Assim, os pais podiam optar pela participação ou não-participação de seus filhos na pesquisa.

Como primeiro teste foi aplicado o BPR-5, pois sua aplicação poderia ser realizada em grupo. Em uma escola, o teste BPR-5 foi dividido em duas aplicações devido ao tempo disponibilizado. O teste SON-R 6-40 foi aplicado individualmente por um membro da equipe de pesquisa. Como o teste é individual e possui tempo médio de duração de 70 minutos foram aplicados quatro testes por dia em média, sendo dois por cada membro da equipe. O grupo de aplicadores constituído de cinco alunos bolsistas da Universidade de Brasília (UnB) passou por treinamento para aplicação dos dois testes.

3.3 Análise de dados

Antes de realizar as análises principais foram realizadas análises exploratórias de frequência para verificar ocorrência de erros de digitação e outros tipos de erros. A seguir, foram calculadas as características psicométricas dos testes aplicados. Na análise de consistência interna dos testes foram utilizados três índices: o coeficiente Lambda 2 de Guttman, a correlação item-resto (item total corrigido) e a correlação média entre os itens. Utilizou-se como índice de consistência interna o coeficiente Lambda 2 de Guttman, uma vez que estudos apontam que o Lambda 2 de Guttman é uma estimativa melhor da fidedignidade, principalmente quando os instrumentos contêm poucos itens ou quando a amostra é pequena (Sijtsma, 2012; Karino, Laros & Jesus, 2011; Tellegen & Laros, 2004; Ten Berge & Zegers, 1978). Para a realização das análises de validade convergente dos dados, utilizou-se da correlação bivariada de Pearson (r). Teoricamente, a medida de correlação varia entre -1 e 1 e quanto mais próxima de zero, significa a ausência de relação entre as variáveis (Moore, 2005). O cálculo das correlações entre testes foi baseado em escores normatizados. Foram calculadas além das correlações brutas, correlações depois da aplicação da correção dupla para atenuação (Miles & Shevlin, 2001; Osborne, 2003). A fórmula da correção dupla é: $r_{xy*} = r_{xy} \div \sqrt{r_{xx} \cdot r_{yy}}$, onde, r_{xy*} é

a correlação depois da correção dupla para atenuação, r_{xy} = correlação bruta e r_{xx} e r_{yy} são os coeficientes de fidedignidade dos dois testes. Osborne (2003) argumenta que a mensuração com testes que parcialmente medem variância de erro causa a subestimação das relações reais entre as variáveis. Segundo esse autor, pesquisadores precisam corrigir para atenuação para obter uma estimativa melhor da relação verdadeira entre as variáveis na população. Outro índice calculado entre os testes foi o índice de determinação, que a correlação elevada ao quadrado. Esse índice é utilizado para mostrar a variância em comum ou a variância compartilhada entre duas variáveis. Para viabilização das análises, foi utilizado *SPSS (Statistical Package for the Social Sciences)*, versão 18.

4. Resultados

4.1 Estudo 1 – Aplicação do SON-R 6-40 junto com o SON-R 2½-7[a]

Análises descritivas dos parâmetros psicométricos

Análises dos parâmetros psicométricos foram realizadas para o SON-R 6-40 e o SON-R 2½-7[a]. Primeiramente serão apresentados os parâmetros psicométricos dos subtestes do SON-R 6-40 e em seguida os resultados dos subtestes do SON-R 2½-7[a].

Um resumo dos parâmetros psicométricos para a bateria SON-R 6-40 é apresenta na Tabela 4.1.

Tabela 4.1 – Parâmetros psicométricos dos quatro subtestes Analogias, Mosaicos, Categorias e Padrões da bateria SON-R 6-40 e da bateria completa – Estudo 1(N = 44).

Parâmetros psicométricos	Subtestes			
	Analogias	Mosaicos	Categorias	Padrões
Correlação item-total (r_{ii})	0,47	0,45	0,42	0,51
Consistência Interna (λ_2)	0,92	0,83	0,89	0,90
Proporção de acertos	0,42	0,35	0,35	0,32
Correlação entre os itens	0,24	0,26	0,21	0,29
Número de itens	36	26	36	26

A bateria SON-R 6-40: Consistência interna: $\lambda_2 = 0,95$; correlação média entre os itens = 0,16; média proporção de acertos = 0,36; n itens = 124.

A Tabela 4.1 mostra bons coeficientes de consistência interna ($\lambda_2 > 0,80$) para todos os subtestes da bateria SON-R 6-40. Uma inspeção detalhada da tabela mostra que o subteste Analogias apresentou maior coeficiente de consistência interna ($\lambda_2 = 0,92$) e o subteste Mosaicos menor coeficiente ($\lambda_2 = 0,83$). A bateria inteira mostrou um coeficiente de consistência interna de 0,95. Observando o nível de dificuldade dos subtestes, podemos concluir que os testes tiveram níveis de dificuldade abaixo de 0,50, demonstrando o nível de dificuldade foi alto para a amostra estudada. O subteste Padrões foi o subteste mais difícil (valor p médio = 0,32). A correlação entre os itens foi maior

para o subteste Padrões ($r = 0,29$). Utilizando o teste t foi observado que não existem diferenças significativas entre meninos e meninas no desempenho no SON-R 6-40 e nos seus subtestes. A Tabela 4.2 apresentada a seguir fornece os parâmetros psicométricos para o subteste Analogias do SON-R 6-40 no nível de item.

Tabela 4.2 – Consistência interna, proporção de acertos (valor p) e correlação item-total corrigida (r_{it}) dos itens do subteste Analogias da bateria SON-R 6-40 - Estudo 1(N=44).

Subteste Analogias (n itens = 36)						
Item	Série a		Série b		Série c	
	valor p	r_{it}	valor p	r_{it}	valor p	r_{it}
1	1,00	-----	1,00	-----	0,98	0,29
2	0,91	0,23	0,84	0,41	0,91	0,35
3	0,86	0,31	0,86	0,34	0,89	0,51
4	0,80	0,58	0,70	0,52	0,77	0,55
5	0,59	0,41	0,57	0,62	0,59	0,58
6	0,20	0,63	0,32	0,64	0,43	0,60
7	0,14	0,60	0,25	0,67	0,25	0,61
8	0,16	0,65	0,07	0,46	0,14	0,57
9	0,09	0,44	0,07	0,53	0,14	0,48
10	0,05	0,33	0,02	0,20	0,02	0,34
11	0,02	0,31	0,02	0,34	0,00	-----
12	0,00	-----	0,02	0,34	0,00	-----
Médias	0,40	0,45	0,40	0,46	0,43	0,49

Consistência interna (λ_2 de Guttman) = 0,92; Média de correlações entre itens = 0,24

Nota: valor p = proporção de acerto dos itens; r_{it} = correlação item-total corrigida.

Inspeção dos dados da Tabela 4.2 indica que das três séries do subteste Analogias a série c foi a série mais fácil (valor p médio = 0,43). As séries a e b mostraram o mesmo grau de dificuldade (valor p médio = 0,40). Os primeiros três itens das series a, b e c foram bastante fáceis para esta amostra (valores $p \geq 0,84$). Como pode ser verificado na Tabela 4.2, em geral o nível de dificuldade dos itens dentre cada série aumenta sucessivamente, conforme o esperado. Apenas dois itens (itens 8a e 3b) apresentaram

dificuldade inferior à do item anterior. Analisando o poder discriminativo dos itens podemos afirmar que a série c tinha em média um poder discriminativo maior (média $r_{it} = 0,49$), enquanto a série a mostrou o valor mais baixo (média $r_{it} = 0,45$). O subteste Analogias mostrou um coeficiente bem adequado de consistência interna, com um valor acima de 0,90 ($\lambda_2 = 0,92$). A próxima tabela, Tabela 4.3, mostra os parâmetros psicométricos no nível de item para o subteste Mosaicos.

Tabela 4.3 – Consistência interna, proporção de acertos (valor p) e correlação item-total corrigida (r_{it}) dos itens para o subteste Mosaicos da bateria SON-R 6-40 – Estudo 1 (N=44).

Subteste Mosaicos (n itens = 26)				
Item	Série a		Série b	
	valor p	r_{it}	valor p	r_{it}
1	1,00	-----	1,00	-----
2	0,98	0,19	1,00	-----
3	0,68	0,45	0,82	0,50
4	0,82	0,43	0,84	0,44
5	0,61	0,66	0,50	0,45
6	0,16	0,46	0,45	0,63
7	0,07	0,40	0,20	0,47
8	0,05	0,36	0,05	0,32
9	0,00	-----	0,00	-----
10	0,00	-----	0,00	-----
11	0,00	-----	0,00	-----
12	0,00	-----	0,00	-----
13	0,00	-----	0,00	-----
Médias	0,33	0,42	0,37	0,47
Consistência interna (λ_2 de Guttman) = 0,83; Média de correlações entre itens = 0,26				

Nota: valor p = proporção de acerto dos itens; r_{it} = correlação item-total corrigida.

Os dados apresentados na Tabela 4.3 mostram que a média de dificuldades dos itens ficou entre 0,33 (série a) e 0,37 (série b). Os dois primeiros itens de cada série foram extremamente fáceis para a amostra estudada. Todos os participantes acertaram itens 1a, 1b e 2b; o item 2a mostrou uma percentagem de acerto de 98%. Os últimos seis itens de

cada série foram bastante difíceis: a percentagem de acerto destes itens ficou menor que 10%. Em geral, o nível de dificuldade dos itens dentre cada série aumentou sucessivamente, conforme o esperado. Apenas dois itens (itens 4a e 4b) apresentaram dificuldade inferior à do item anterior. Quanto ao poder discriminativo do subteste Mosaicos pode ser observado que, em geral, os itens da série b tinham um poder discriminativo mais elevado (média $r_{it} = 0,47$). A Tabela 4.4 apresenta os parâmetros psicométricos relativos ao subteste Categorias.

Tabela 4.4 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Categorias da bateria SON-R 6-40 – Estudo 1 (N=44).

Subteste Categorias (n itens = 36)						
Item	Série a		Série b		Série c	
	valor p	r_{it}	valor p	r_{it}	valor p	r_{it}
1	0,84	0,39	0,77	0,54	0,93	0,50
2	0,86	0,36	0,93	0,30	0,93	0,44
3	0,80	0,58	0,84	0,53	0,84	0,60
4	0,73	0,43	0,73	0,47	0,59	0,41
5	0,39	0,43	0,55	0,51	0,70	0,56
6	0,23	0,38	0,20	0,29	0,20	0,47
7	0,02	0,10	0,20	0,49	0,16	0,53
8	0,02	0,04	0,11	0,50	0,14	0,41
9	0,00	-----	0,02	0,33	0,07	0,44
10	0,00	-----	0,05	0,45	0,00	-----
11	0,00	-----	0,00	-----	0,00	-----
12	0,00	-----	0,00	-----	0,00	-----
Médias	0,32	0,34	0,35	0,44	0,38	0,48
Consistência interna (λ_2 de Guttman) = 0,89; Média de correlações entre itens = 0,21						

Nota: valor p = proporção de acerto dos itens; r_{it} = correlação item-total corrigida.

A tabela relativa aos dados do subteste Categorias mostra que a dificuldade média dos itens variou entre 0,32 (série a) e 0,38 (série c). Os três últimos itens de cada série foram extremamente difíceis para essa amostra (valores $p \leq 0,05$). Em geral, o nível de dificuldade dos itens dentre cada série aumentou sucessivamente, conforme o esperado.

Apenas dois itens (itens 2a e 2b) apresentaram dificuldade inferior à do item anterior. Em relação do poder discriminativo do subteste Categorias chama a atenção o fato de a série c mostrar um poder discriminativo mais elevado (média $r_{it} = 0,48$). A Tabela 4.5 mostra os parâmetros psicométricos do subteste Padrões, o último subteste do SON-R 6-40.

Tabela 4.5 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Padrões da bateria SON-R 6-40 – Estudo 1 (N=44).

Subteste Padrões (n itens = 26)				
Item	Série a		Série b	
	valor p	r_{it}	valor p	r_{it}
1	0,86	0,50	0,77	0,63
2	0,89	0,45	0,98	0,29
3	0,84	0,48	0,73	0,65
4	0,57	0,61	0,77	0,64
5	0,27	0,61	0,48	0,60
6	0,32	0,65	0,41	0,63
7	0,09	0,44	0,18	0,54
8	0,07	0,43	0,07	0,48
9	0,02	0,27	0,00	-----
10	0,00	-----	0,02	0,27
11	0,00	-----	0,00	-----
12	0,00	-----	0,00	-----
13	0,00	-----	0,00	-----
Médias	0,30	0,49	0,34	0,53

Consistência interna (λ_2 de Guttman) = 0,90; Média de correlações entre itens = 0,29

Nota: valor p = proporção de acerto dos itens; r_{it} = correlação item-total corrigida.

Inspeção da Tabela 4.5 mostra que os últimos seis itens das duas séries foram extremamente difíceis para a amostra estudada (valores $p \leq 0,10$). Em geral, o nível de dificuldade dos itens dentre cada série aumentou sucessivamente, conforme o esperado. Do total, quatro itens (itens 2a, 6a, 2b e 4b) apresentaram dificuldade inferior à do item anterior. O poder discriminativo da série b (média $r_{it} = 0,53$) foi maior que o da série a (média $r_{it} = 0,49$). Em seguida, a Tabela 4.6 apresenta um resumo dos parâmetros psicométricos para a bateria SON-R 2½-7[a].

Tabela 4.6 – Parâmetros psicométricos dos quatro subtestes Mosaicos, Categorias, Situações e Padrões da bateria SON-R 2½-7[a], da Escala de Execução (EE), da Escala de Raciocínio (ER) e da bateria completa – Estudo 1 (N = 44).

Parâmetros psicométricos	Subtestes			
	Mosaicos	Categorias	Situações	Padrões
Correlação item-total (r_{it})	0,49	0,43	0,27	0,30
Consistência Interna (λ_2)	0,78	0,78	0,60	0,62
Proporção de acertos	0,75	0,80	0,78	0,83
Correlação entre os itens	0,24	0,26	0,21	0,29
Número de itens	15	15	14	16

A bateria SON-R 2½-7[a]: Consistência interna: $\lambda_2 = 0,83$; correlação média entre os itens = 0,11; média proporção de acertos = 0,79; n itens = 60.

Escala de Execução: Consistência interna: $\lambda_2 = 0,80$; correlação média entre os itens = 0,20; média proporção de acertos = 0,79; n itens = 31.

Escala de Raciocínio: Consistência interna: $\lambda_2 = 0,71$; correlação média entre os itens = 0,10; média proporção de acertos = 0,79; n itens = 29.

A Tabela 4.6 mostra índices de consistência interna adequados para os subtestes Mosaicos e Categorias ($\lambda_2 > 0,75$). Os coeficientes para Situações e Padrões, entretanto, foram minimamente aceitáveis. A fidedignidade relativamente baixa desses dois subtestes pode ser devido ao fato que os participantes tinham uma idade limite para a realização do SON-R 2½-7[a]. A média da idade da amostra foi de quase sete anos (6,9 anos) e o SON-R 2½-7[a] tem normas até 7 anos e 11 meses. Assim, pode ser que um efeito de teto ficou responsável pela baixa fidedignidade de Situações e Padrões. A bateria inteira mostrou um coeficiente de consistência interna de 0,83 e as Escalas de Execução e de Raciocínio de 0,80 e 0,71 respectivamente. Observando o nível de dificuldade dos subtestes, podemos concluir que o teste em geral foi fácil para a amostra estudada (valor p médio de todos os subtestes acima de 0,75). O subteste Padrões foi o subteste mais fácil (valor p médio de 0,83). O subteste Padrões mostrou-se também o subteste com a correlação mais alta entre os itens (média $r = 0,29$). A Tabela 4.7 apresenta os parâmetros psicométricos no nível de item para os subtestes Mosaicos e Padrões da Escala de Execução da bateria SON-R 2½-7[a].

Tabela 4.7 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para os subtestes Mosaicos e Padrões da bateria SON-R 2½-7[a] (N=44).

Subteste Mosaicos (n itens = 15)			Subteste Padrões (n itens = 16)	
Item	valor p	r_{it}	valor p	r_{it}
1	1,00	----	1,00	----
2	1,00	----	1,00	----
3	1,00	----	1,00	----
4	1,00	----	1,00	----
5	1,00	----	1,00	----
6	1,00	----	1,00	----
7	1,00	----	1,00	----
8	1,00	----	1,00	----
9	0,89	0,41	1,00	----
10	0,77	0,36	0,93	0,07
11	0,61	0,59	0,84	0,10
12	0,34	0,44	0,80	0,37
13	0,32	0,58	0,66	0,37
14	0,27	0,62	0,68	0,52
15	0,11	0,45	0,27	0,35
16	----	----	0,11	0,32
Médias	0,75	0,49	0,83	0,30
Consistência interna (λ_2 de Guttman) = 0,78			Consistência interna (λ_2 de Guttman) = 0,62	
Média de correlações entre itens = 0,32			Média de correlações entre itens = 0,15	

Nota: valor p = proporção de acertos, r_{it} = correlação corrigida item-total.

Os dados da Tabela 4.7 mostram que os primeiros oito itens foram acertados por todas as crianças da amostra. Em relação aos testes em si, nota-se que o subteste Padrões possui uma maior proporção de acertos do que o subteste Mosaicos. Conforme o esperado, o nível de dificuldade dos itens dentre cada série aumentou sucessivamente, com exceção do item 14 do subteste Padrões que apresentou dificuldade inferior à do item anterior. Inspeção dos índices de discriminação permite afirmar que os itens do subteste Mosaicos tem um poder discriminativo maior do que os itens do subteste Padrões. A Tabela 4.8 mostra em seguida os parâmetros psicométricos no nível de item para os subtestes Categorias e Situações da Escala de Raciocínio da bateria SON-R 2½-7[a].

Tabela 4.8 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para os subtestes Categorias e Situações da bateria SON-R 2½-7[a] (N=44).

Subteste Categorias (n itens = 15)			Subteste Situações (n itens= 14)	
Item	valor p	r_{it}	valor p	r_{it}
1	1,00	----	1,00	----
2	1,00	----	1,00	----
3	1,00	----	1,00	----
4	1,00	----	1,00	----
5	0,93	0,33	0,93	0,38
6	0,98	0,48	0,95	0,22
7	0,91	0,37	0,98	0,20
8	0,95	0,40	1,00	----
9	0,89	0,35	0,91	0,35
10	0,86	0,43	0,73	0,26
11	0,61	0,46	0,61	0,21
12	0,59	0,42	0,43	0,42
13	0,55	0,49	0,18	0,11
14	0,39	0,50	0,25	0,31
15	0,30	0,48	----	----
Médias	0,80	0,43	0,78	0,27
Consistência interna (λ_2 de Guttman) = 0,78			Consistência interna (λ_2 de Guttman) = 0,60	
Média de correlações entre itens = 0,25			Média de correlações entre itens = 0,14	

Notas: valor p = proporção de acertos, r_{it} = correlação corrigida item-total.

A análise dos parâmetros psicométricos dos subtestes Categorias e Situações mostra que o primeiro apresenta um índice aceitável de consistência interna ($\lambda_2 = 0,78$), já o último, entretanto, apresenta um baixo índice ($\lambda_2 = 0,60$). O baixo coeficiente de fidedignidade está provavelmente relacionado com o efeito de teto do subteste nesta amostra. A proporção média de acerto para ambos os subtestes foi por volta de 0,80 indicando que os subtestes estavam muito fáceis para a amostra estudada. Pode ser observado que os primeiros quatro itens de cada subteste foram acertados por todas as crianças da amostra. O nível de dificuldade dos itens dentre cada série aumentou sucessivamente, com exceção dos itens 6 e 8 do Categorias e 6, 7 e 8 do Situações. Os

dados da Tabela 4.8 mostram que o subteste Categorias tem um poder discriminativo maior em comparação do subteste Situações (média $r_{it} = 0,43$ versus média $r_{it} = 0,27$).

Análises de validade convergente

Para conseguir evidências de validade convergente da bateria SON-R 6-40 foram realizadas análises de correlação entre esse teste e a bateria SON-R 2½-7[a]. A Tabela 4.9 mostra os resultados das análises realizadas.

Tabela 4.9 – Coeficientes de validade e de determinação entre os escores de QI Total do SON-R 6-40, QI Total do SON-R 2½-7[a] e a Escala de Execução e a Escala de Raciocínio do SON-R 2½-7[a] (N=44).

Coeficientes de validade e determinação				
	SON-R 6-40	EE SON-R	ER SON-R	SON-R 2½-7[a]
SON-R 6-40	-	0,38 (0,28)	0,52 (0,37)	0,56 (0,46)
EE SON-R	0,62 (0,54)	-	0,28 (0,20)	0,87 (0,69)
ER SON-R	0,72 (0,60)	0,53 (0,45)	-	0,92 (0,76)
SON-R 2½-7[a]	0,75 (0,67)	0,93 (0,83)	0,96 (0,87)	-

IC 90%: $r = 0,62$ - entre 0,44 e 0,75; $r = 0,72$ - entre 0,57 e 0,82; $r = 0,75$ - entre 0,61 e 0,84.

Notas: EE SON-R = Escala de Execução do SON-R 2½-7[a]; ER SON-R = Escala de Raciocínio do SON-R 2½-7[a]. Os valores entre parênteses são as correlações brutas e os primeiros valores foram corrigidos para atenuação. Os coeficientes de validade encontram-se no triângulo inferior da tabela, e os coeficientes de determinação (os coeficientes de correlação elevados ao quadrado) encontram-se no triângulo superior. IC 90% = intervalo de confiança de 90%.

Os resultados da Tabela 4.9 apontam que a correlação bruta do SON-R 6-40 e o SON-R 2½-7[a] foi 0,67. Para obter a correlação corrigida, a correlação bruta precisa ser dividida pela raiz quadrada do produto dos coeficientes de fidedignidade de ambos os testes ($r_{xy}^* = r_{xy} \div \sqrt{r_{xx} \cdot r_{yy}} = 0,67 \div \sqrt{0,95 \cdot 0,83} = 0,75$). O valor corrigido de 0,75 é a melhor estimativa do coeficiente de validade convergente do SON-R 6-40 com o SON-R 2½-7[a]. O valor de 0,75 da correlação entre o SON-R 6-40 e o SON-R 2½-7[a] é bem próximo ao valor de 0,77, encontrado em um estudo realizado na Alemanha (N =35) entre

o SON-R 6-40 e o WISC-IV (Tellegen & Laros, 2011). O intervalo de confiança de 90% da correlação de 0,75 para uma amostra de 44 pessoas ficou entre 0,61 e 0,84 (limite inferior e limite superior do intervalo) mostrada na Tabela 4.9 justifica a conclusão que existe uma correlação forte entre a baterias SON-R 6-40 e SON-R 2½-7[a].

4.2 Estudo 2 – Aplicação do SON-R 6-40 junto com o BPR-5

Análises descritivas dos parâmetros psicométricos

No estudo 2 o SON-R 6-40 foi aplicado junto com a forma A da BPR-5 em 68 estudantes. Para essa amostra são apresentados primeiramente os parâmetros psicométricos da bateria SON-R 6-40 e seus subtestes e depois os parâmetros psicométricos da forma A da BPR-5 e seus subtestes. Um resumo dos parâmetros psicométricos da bateria SON-R 6-40 é apresenta na Tabela 4.10.

Tabela 4.10 – Parâmetros psicométricos dos subtestes Analogias, Mosaicos, Categorias e Padrões da bateria SON-R 6-40 – Estudo 2 (N = 68).

Parâmetros psicométricos	Subtestes			
	Analogias	Mosaicos	Categorias	Padrões
Correlação item-total	0,39	0,44	0,39	0,49
Consistência Interna (λ_2)	0,87	0,89	0,90	0,91
Proporção de acertos (valor p)	0,61	0,62	0,55	0,64
Correlação entre os itens	0,15	0,22	0,18	0,24
Número de itens	36	26	36	26

A bateria SON-R 6-40: Consistência interna: $\lambda_2 = 0,95$; correlação média entre os itens = 0,15; média proporção de acertos = 0,60; n itens = 124.

Verificando os coeficientes de consistência interna torna-se evidente que todos os subtestes do SON-R 6-40 têm uma fidedignidade alta ($\lambda_2 > 0,85$), sendo que o subteste Padrões apresentou o valor mais alto ($\lambda_2 = 0,91$). A bateria inteira mostrou um coeficiente de consistência interna de 0,95. Observando o nível de dificuldade dos subtestes, podemos afirmar que todos tiveram níveis de dificuldade média acima de 0,50,

demonstrando um nível de dificuldade moderado. O subteste mais fácil foi Padrões (valor p médio = 0,64) e o subteste mais difícil foi Categorias (valor p médio = 0,55). A correlação entre os itens foi maior para o subteste Padrões ($r = 0,24$). Utilizando o teste t ficou evidente que não existem diferenças significativas entre meninos e meninas no desempenho no SON-R 6-40 e nos seus subtestes. A Tabela 4.11 os parâmetros psicométricos no nível de item para o subteste Analogias.

Tabela 4.11 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Analogias da bateria SON-R 6-40 – Estudo 2 (N=68).

Subteste Analogias (n itens = 36)						
Item	Série a		Série b		Série c	
	valor p	r_{it}	valor p	r_{it}	valor p	r_{it}
1	1,00	----	1,00	----	1,00	----
2	1,00	----	1,00	----	1,00	----
3	0,99	0,11	1,00	----	1,00	----
4	0,99	0,11	0,97	0,21	0,99	0,05
5	0,99	0,11	0,84	0,23	0,93	0,37
6	0,54	0,52	0,76	0,44	0,72	0,58
7	0,59	0,38	0,51	0,53	0,68	0,55
8	0,54	0,54	0,40	0,56	0,56	0,30
9	0,25	0,38	0,22	0,50	0,54	0,46
10	0,10	0,39	0,15	0,48	0,28	0,41
11	0,07	0,35	0,15	0,50	0,10	0,32
12	0,00	0,16	0,03	0,10	0,09	0,36
Médias	0,59	0,31	0,59	0,39	0,66	0,48

Consistência interna (λ_2 de Guttman) = 0,87; Média de correlações entre itens = 0,15

Nota: valor p = proporção de acertos, r_{it} = correlação corrigida item-total.

Ao realizar-se uma verificação dos parâmetros psicométricos apresentados pela Tabela 4.11, pode-se notar que a dificuldade média de cada série foi de aproximadamente 0,60 (valor p médio = 0,61), mostrando que o teste não foi muito difícil para a amostra estudada. Das três séries, a série c apresentou maior proporção de acertos (valor p médio = 0,66). Como pode ser verificado na Tabela 4.11, em geral o nível de dificuldade dos

itens dentre cada série aumentou sucessivamente, conforme o esperado. Apenas um item (item 7a) apresentou dificuldade inferior à do item anterior. Em uma análise do poder discriminativo dos itens, pode-se perceber que a série c foi em que os itens tiveram maior poder discriminativo ($r_{it} = 0,48$). A Tabela 4.12 apresenta os parâmetros psicométricos do subteste Mosaicos.

Tabela 4.12 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Mosaicos da bateria SON-R 6-40 – Estudo 2 (N=68).

Subteste Mosaicos (n itens = 26)				
Item	Série a		Série b	
	valor p	r_{it}	valor p	r_{it}
1	1,00	-----	1,00	-----
2	1,00	-----	1,00	-----
3	0,96	0,26	0,94	0,35
4	0,99	0,25	0,96	0,29
5	0,87	0,42	0,87	0,44
6	0,62	0,48	0,85	0,44
7	0,50	0,61	0,84	0,54
8	0,47	0,61	0,71	0,57
9	0,24	0,59	0,29	0,58
10	0,10	0,54	0,32	0,65
11	0,06	0,49	0,12	0,58
12	0,03	0,30	0,07	0,53
13	0,01	0,26	0,03	-0,14
Médias	0,57	0,44	0,67	0,44

Consistência interna (λ_2 de Guttman) = 0,89; Média de correlações entre itens = 0,22

Nota: valor p = proporção de acertos, r_{it} = correlação corrigida item-total.

Na análise da dificuldade dos itens, é possível notar que os quatro primeiros itens de cada série foram extremamente fáceis para a amostra estudada (valor p > 0,90). Em média, os itens da série b foram mais fáceis que os itens da série a (valor p médio = 0,67 versus valor p médio = 0,57). Em geral, o nível de dificuldade dos itens dentre cada série aumentou sucessivamente, conforme o esperado. Entretanto, três itens (itens 4a, 4b e 10b)

apresentaram dificuldade inferior à do item anterior. As correlações item-total corrigidas mostram que os itens de ambos os testes possuem poder discriminativo igual. Na Tabela 4.13 são apresentados os parâmetros psicométricos no nível de item para o subteste Categorias do SON-R 6-40.

Tabela 4.13 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Categorias da bateria SON-R 6-40 – Estudo 2 (N=68).

Subteste Categorias (n itens = 36)						
Item	Série a		Série b		Série c	
	valor p	r_{it}	valor p	r_{it}	valor p	r_{it}
1	0,97	0,36	0,99	0,26	1,00	-----
2	0,99	0,20	0,96	0,39	0,97	0,22
3	0,99	0,20	0,93	0,41	0,99	0,13
4	0,96	0,34	0,91	0,40	0,79	0,32
5	0,76	0,21	0,72	0,34	0,96	0,32
6	0,49	0,53	0,54	0,56	0,60	0,54
7	0,56	0,48	0,43	0,55	0,59	0,64
8	0,26	0,40	0,37	0,64	0,50	0,53
9	0,18	0,44	0,29	0,62	0,35	0,44
10	0,12	0,30	0,18	0,54	0,15	0,39
11	0,04	0,26	0,10	0,41	0,10	0,39
12	0,00	-----	0,04	0,28	0,07	0,26
Médias	0,53	0,34	0,54	0,45	0,59	0,38

Consistência interna (λ_2 de Guttman) = 0,90; Média de correlações entre itens = 0,18

Nota: valor p = proporção de acertos, r_{it} = correlação corrigida item-total.

Quanto à dificuldade dos itens, os três primeiros itens de cada série foram muito fáceis para a amostra estudada (valor p > 0,90), enquanto que os dois últimos itens de cada série foram os mais difíceis para os respondentes (valor p ≤ 0,10). Também é possível verificar que a série c foi a mais fácil para a amostra estudada (valor p médio = 0,59), enquanto a série a foi ligeiramente mais difícil (valor p médio = 0,53). Em geral, o nível de dificuldade dos itens dentro de cada série aumentou sucessivamente, com a exceção dos itens 7a, 3c e 5c. Analisando o poder discriminativo podemos observar que a série b

possui itens com maior valor discriminativo (r_{it} médio = 0,45). A Tabela 4.14 descreve os parâmetros psicométricos no nível de item do subteste Padrões do SON-R 6-40.

Tabela 4.14 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subteste Padrões da bateria SON-R 6-40 – Estudo 2 (N=68).

Subteste Padrões (n itens = 26)				
Item	Série a		Série b	
	valor p	r_{it}	valor p	r_{it}
1	0,97	0,42	1,00	----
2	1,00	----	1,00	----
3	0,97	0,24	1,00	----
4	0,97	0,15	1,00	----
5	0,94	0,20	0,94	0,31
6	0,81	0,39	0,93	0,35
7	0,81	0,43	0,81	0,44
8	0,40	0,61	0,72	0,52
9	0,37	0,77	0,51	0,69
10	0,28	0,69	0,35	0,70
11	0,18	0,60	0,25	0,70
12	0,07	0,54	0,09	0,51
13	0,06	0,51	0,04	0,46
Médias	0,60	0,46	0,67	0,52

Consistência interna (λ_2 de Guttman) = 0,91; Média de correlações entre itens = 0,24

Nota: valor p = proporção de acertos, r_{it} = correlação corrigida item-total.

A inspeção da Tabela 4.14 mostra que ambas as séries foram moderadamente fáceis para a amostra estudada (valor p médio = 0,64), porém a série a apresentou itens mais difíceis que a série b para os participantes (valor p médio = 0,67 versus valor p médio = 0,60). Em geral, o nível de dificuldade dos itens dentre cada série aumentou sucessivamente, com a exceção dos itens 7a, 3c e 5c. A média de correlações item-total corrigida da série b é maior, mostrando que essa série possui o maior poder discriminativo (r_{it} médio = 0,52 versus r_{it} médio = 0,46).

A seguir é apresentado um resumo dos parâmetros psicométricos para a bateria BPR-5 aplicada no estudo 2. Posteriormente, são mostradas as características psicométricas dos quatro subtestes do BPR-5 no nível do item. A Tabela 4.15 apresenta um resumo dos parâmetros psicométricos para a bateria BPR-5.

Tabela 4.15 – Parâmetros psicométricos dos subtestes Raciocínio Verbal (RV), Raciocínio Abstrato (RA), Raciocínio Espacial (RE) e Raciocínio Numérico (RN) do BPR-5 (N = 68).

Parâmetros psicométricos	Subtestes			
	RV	RA	RE	RN
Correlação item-total	0,60	0,62	0,64	0,64
Consistência Interna (λ_2)	0,95	0,95	0,92	0,94
Proporção de acertos	0,37	0,39	0,29	0,24
Correlação entre os itens	0,38	0,40	0,35	0,38
Número de itens	25	25	20	20

A forma A do BPR-5: Consistência interna: $\lambda_2 = 0,98$; correlação média entre os itens = 0,33; média proporção de acertos = 0,33; n itens = 90.

Ao analisar os coeficientes de consistência interna, percebe-se que todos foram considerados altos ($\lambda_2 > 0,90$). Ao analisar cada subteste, é notável que o maior coeficiente de consistência interna é obtido pelos subtestes Raciocínio Verbal e Raciocínio Abstrato ($\lambda_2 = 0,95$). A bateria inteira mostrou um coeficiente de consistência interna de 0,98. Verificando a proporção de acertos dos itens torna-se evidente que todas as médias foram abaixo de 0,40, o que demonstra um grau de dificuldade elevado para a amostra estudada. O subteste Raciocínio Numérico demonstrou a menor proporção de acertos (valor p médio = 0,24), enquanto o teste Raciocínio Abstrato mostrou a maior porcentagem de acertos (valor p médio = 0,39). A Tabela 4.16 mostra os parâmetros psicométricos no nível do item para o subteste Raciocínio Verbal.

Tabela 4.16 – Consistência interna, proporção de acertos (valor p), correlação item-total corrigida (r_{it}) para o subtteste Raciocínio Verbal da bateria BPR-5 (N=68).

Subteste Raciocínio Verbal (n itens =25)					
Itens 1 até 13			Itens 14 até 25		
Item	valor p	r_{it}	Item	valor p	r_{it}
1	0,57	0,83	14	0,34	0,52
2	0,46	0,71	15	0,34	0,64
3	0,38	0,52	16	0,44	0,80
4	0,57	0,79	17	0,49	0,69
5	0,47	0,78	18	0,07	0,31
6	0,57	0,79	19	0,26	0,52
7	0,56	0,65	20	0,22	0,44
8	0,44	0,54	21	0,29	0,43
9	0,47	0,68	22	0,38	0,48
10	0,32	0,69	23	0,22	0,42
11	0,32	0,65	24	0,18	0,42
12	0,25	0,56	25	0,10	0,35
13	0,51	0,82	----	----	----

Média valor p (25 itens) = 0,37; média correlação item-total corrigida (25 itens) = 0,60

Consistência interna (λ_2 de Guttman) = 0,95; Média de correlações entre itens = 0,38

Nota: valor p = proporção de acertos, r_{it} = correlação corrigida item-total.

Segundo os dados obtidos na tabela 4.16, pode-se perceber que os itens do subtteste Raciocínio Verbal foram relativamente difíceis para a amostra pesquisada, com uma média de dificuldade de 0,37. A percentagem de acerto variou entre 7% e 57%. Os resultados na Tabela 4.16 mostram que os itens têm alto poder discriminativo (r_{it} médio = 0,60). A Tabela 4.17 descreve os parâmetros psicométricos para o subtteste Raciocínio Abstrato da escala BPR-5.

Tabela 4.17 – Consistência interna, dificuldade dos itens (valor p), correlação item-total corrigida (r_{it}), e média de correlação entre os itens para o subtteste Raciocínio Abstrato (RA) da bateria BPR-5 (N=68).

Subteste Raciocínio Abstrato (n itens = 25)					
Itens 1 até 13			Itens 14 até 25		
Item	valor p	r_{it}	Item	valor p	r_{it}
1	0,57	0,85	14	0,56	0,78
2	0,60	0,82	15	0,47	0,72
3	0,53	0,84	16	0,31	0,54
4	0,54	0,76	17	0,40	0,61
5	0,47	0,82	18	0,37	0,60
6	0,47	0,79	19	0,25	0,44
7	0,53	0,75	20	0,24	0,40
8	0,50	0,72	21	0,18	0,33
9	0,41	0,64	22	0,31	0,70
10	0,44	0,52	23	0,18	0,42
11	0,44	0,62	24	0,19	0,41
12	0,35	0,60	25	0,07	0,17
13	0,44	0,63	-----	-----	-----
Média valor p (25 itens) = 0,39; Média correlação item-total corrigida (25 itens) = 0,62					
Consistência interna (λ_2 de Guttman) = 0,95; Média de correlações entre itens = 0,40					

Nota: valor p = proporção de acertos, r_{it} = correlação corrigida item-total.

Inspeção da Tabela 4.17 revela que a porcentagem de acerto dos itens do subtteste Raciocínio Abstrato variou entre 7% e 60%. As correlações item-teste corrigidas mostram que todos os itens têm um alto um poder discriminatório (r_{it} médio = 0,62). A Tabela 4.18 mostra os índices relativos ao subtteste Raciocínio Espacial da escala BPR-5.

Tabela 4.18 – Consistência interna (λ_2), dificuldade dos itens (valor p), correlação item-total corrigida (r_{it}), consistência interna e média de correlação entre os itens para o subteste Raciocínio Espacial (RE) da bateria BPR-5 (N=68).

Subteste Raciocínio Espacial (n itens = 20)					
Itens 1 até 10			Itens 11 até 20		
Item	valor p	r_{it}	Item	valor p	r_{it}
1	0,41	0,85	11	0,24	0,78
2	0,43	0,82	12	0,34	0,72
3	0,40	0,84	13	0,25	0,54
4	0,38	0,76	14	0,25	0,61
5	0,26	0,82	15	0,32	0,60
6	0,38	0,79	16	0,19	0,44
7	0,32	0,75	17	0,19	0,40
8	0,37	0,72	18	0,24	0,33
9	0,19	0,64	19	0,21	0,70
10	0,22	0,52	20	0,16	0,42
Média valor p (25 itens) = 0,29; Média correlação item-total corrigida (25 itens) = 0,64					
Consistência interna (λ_2 de Guttman) = 0,92; Média de correlações entre itens = 0,35					

Nota: p = proporção de acertos, r_{it} = correlação corrigida item-total.

Inspeção da Tabela 4.18 revela que a porcentagem de acerto dos itens do subteste Raciocínio Espacial variou entre 16% e 40%. As correlações item-teste corrigidas mostram que todos os itens têm um alto poder discriminatório (r_{it} médio = 0,64). Os dados referentes às análises dos parâmetros psicométricos do subteste Raciocínio Numérico são resumidos na Tabela 4.19.

Tabela 4.19 – Consistência interna, dificuldade dos itens (valor p), correlação item-total corrigida (r_{it}), consistência interna e média de correlações entre os itens para o subteste Raciocínio Numérico (RN) da bateria BPR-5 (N = 68).

Subteste Raciocínio Numérico (n itens = 20)					
Itens 1 até 10			Itens 11 até 20		
Item	valor p	r_{it}	Item	valor p	r_{it}
1	0,68	0,58	11	0,18	0,73
2	0,25	0,82	12	0,15	0,54
3	0,63	0,59	13	0,16	0,66
4	0,41	0,51	14	0,15	0,76
5	0,28	0,71	15	0,12	0,71
6	0,37	0,56	16	0,10	0,65
7	0,24	0,74	17	0,06	0,32
8	0,26	0,73	18	0,06	0,34
9	0,21	0,72	19	0,09	0,28
10	0,32	0,46	20	0,04	0,25

Média valor p (25 itens) = 0,24; Média correlação item-total corrigida (25 itens) = 0,64

Consistência interna (λ_2 de Guttman) = 0,94; Média de correlações entre itens = 0,38

Nota: valor p = proporção de acertos., r_{it} = correlação corrigida item-total.

Segundo os dados apresentados na Tabela 4.16, pode-se observar que os itens do subteste Raciocínio Verbal foram muito difíceis para a amostra pesquisada, com uma média de dificuldade de 0,24. A percentagem de acerto variou entre 4% e 68%. Os resultados na Tabela 4.16 mostram que os itens têm alto poder discriminativo (r_{it} médio = 0,64). A seguir são apresentados os resultados do estudo 2 em relação da análise da validade convergente do SON-R 6-40 com o BPR-5.

Análises de validade convergente

Para conseguir evidências de validade convergente da bateria SON-R 6-40 foram realizadas análises de correlação entre esse teste a bateria BPR-5. A Tabela 4.20 mostra os resultados das análises realizadas.

Tabela 4.20 – Coeficientes de validade e coeficientes de determinação para os escores de QI Total do SON-R 6-40, e as baterias de Raciocínio Abstrato, Verbal, Espacial e Numérico (RA, RV, RE e RN, respectivamente) do teste BPR-5(N=68).

Coeficientes de validade e de determinação						
	SON-R 6-40	RA	RV	RE	RN	BPR-5
SON-R 6-40	-	0,46 (0,42)	0,38 (0,35)	0,46 (0,41)	0,19 (0,18)	0,52 (0,48)
RA	0,68 (0,65)	-	0,44 (0,40)	0,64 (0,56)	0,24 (0,21)	0,81 (0,76)
RV	0,62 (0,59)	0,66 (0,63)	-	0,44 (0,38)	0,49 (0,44)	0,79 (0,74)
RE	0,68 (0,64)	0,80 (0,75)	0,66 (0,62)	-	0,28 (0,24)	0,79 (0,72)
RN	0,44 (0,42)	0,49 (0,46)	0,70 (0,66)	0,53 (0,49)	-	0,71 (0,61)
BPR-5	0,72 (0,69)	0,90 (0,87)	0,89 (0,86)	0,89 (0,85)	0,84 (0,78)	-

IC 90%: $r = 0,68$ - entre 0,58 e 0,76; $r = 0,62$ - entre 0,48 e 0,73; $r = 0,44$ - entre 0,26 e 0,59; $r = 0,72$ - entre 0,61 e 0,81.

Notas: RA = Raciocínio Abstrato, RV = Raciocínio Verbal, RE = Raciocínio Espacial. RN = Raciocínio Numérico. Os valores entre parênteses são as correlações brutas e os primeiros valores foram corrigidos para atenuação. Os coeficientes de validade encontram-se no triângulo inferior da tabela, e os coeficientes de determinação (os coeficientes de correlação elevados ao quadrado) encontram-se no triângulo superior. IC 90% = intervalo de confiança de 90%.

Um exame do triângulo inferior da Tabela 4.20 mostra coeficientes de correlação altos entre o SON-R 6-40 e o BPR-5 e seus subtestes, com apenas uma correlação abaixo de 0,50 (a correlação com o subteste Raciocínio Numérico). O coeficiente mais importante no triângulo inferior para avaliar a validade convergente do SON-R 6-40 é a correlação corrigida entre os escores totais do SON-R 6-40 e do BPR-5 ($r = 0,72$). Uma correlação de 0,72 significa que os dois testes têm mais de 50% da variância em comum. Como esperado, as correlações mais altas foram encontradas entre o SON-R 6-40 e os

subtestes de Raciocínio Abstrato e Raciocínio Espacial do BPR-5 ($r = 0,68$ para ambos os subtestes) uma vez que estes subtestes medem o mesmo construto de inteligência que o SON-R 6-40 visa medir, a saber, a inteligência fluida. Correlações mais baixas foram encontradas entre o SON-R 6-40 e os subtestes de Raciocínio Verbal e Numérico da BPR-5 que medem a inteligência cristalizada ($r = 0,62$ e $r = 0,44$ respectivamente). Um estudo realizado com uma amostra de 330 universitários que foram testados com o Raven o G-36 e o BPR-5 (Almeida, Nascimento, Lima, Vasconcelos, Akama & Santos, 2010) indicou que os subtestes Raciocínio Verbal e Raciocínio Numérico medem inteligência cristalizada e os outros subtestes do BPR-5 medem inteligência fluida. O valor de $0,72$ da correlação entre o SON-R 6-40 e o BPR-5[a] é bem próximo ao valor de $0,77$, encontrado em um estudo realizado na Alemanha ($N = 35$) entre o SON-R 6-40 e o WISC-IV (Tellegen & Laros, 2011). O intervalo de confiança de 90% da correlação de $0,72$ para uma amostra de 68 pessoas entre $0,61$ e $0,81$ (limite inferior e limite superior do intervalo) mostrada na Tabela 4.20 justifica a conclusão que existe uma correlação forte entre as baterias SON-R 6-40 e BPR-5.

5. Considerações Finais

O propósito principal da atual pesquisa foi fornecer evidências de validade convergente da bateria SON-R 6-40 através da aplicação da bateria com dois outros instrumentos que medem inteligência: o SON-R 2½-7[a] e a forma A da BPR-5. Um dos principais resultados da pesquisa é que a aplicação do SON-R 6-40 em crianças com idades entre 6 e 7 anos fornece resultados mais fidedignos do que a do SON-R 2½-7[a]. Atualmente, a pesquisa para desenvolver normas nacionais do SON-R 6-40 para o Brasil está em andamento. Até a data nenhum outro estudo sobre a validade do SON-R 6-40 foi efetivado nesse país. Assim, ressalta-se a importância do estudo atual uma vez que é o

primeiro estudo sobre a validade do SON-R 6-40 realizado no Brasil. O primeiro passo na obtenção de evidências da validade convergente do SON-R 6-40 foi o cálculo dos parâmetros psicométricos dos testes aplicados. Foi importante obter dados sobre a consistência interna dos testes administrados uma vez que estes dados precisavam ser utilizadas na correção para atenuação dos coeficientes de correlação entre os testes. No Estudo 1, o SON-R 6-40 foi aplicado junto com o SON-R 2½-7[a] em 44 crianças. Neste estudo os parâmetros psicométricos do SON-R 6-40 foram satisfatórias: o coeficiente de consistência interna da bateria inteira foi de 0,95 e os coeficientes de consistência interna dos subtestes variam entre 0,83 e 0,92. Para o outro teste aplicado neste estudo, o SON-R 2½-7[a], os coeficientes encontrados de consistência interna foram menores. A fidedignidade da bateria completa foi de 0,83 e das Escalas de Execução e de Raciocínio foi de 0,80 e 0,71 respectivamente. A fidedignidade dos subtestes do SON-R 2½-7[a] variou entre 0,60 e 0,78. Assim, pode-se notar que os índices de fidedignidade do SON-R 2½-7[a] são consideravelmente menores que os índices para o SON-R 6-40. No desenvolvimento dos testes SON-R, foi levado em consideração a possibilidade de testar as crianças entre 6 e 7 com os dois testes. Os resultados dessa pesquisa sugerem que o SON-R 2½-7[a] mostra um efeito de teto para a amostra pesquisada, a saber, crianças do DF que possuem nível socioeconômico elevado em relação ao resto do Brasil. Aparentemente, esse efeito de teto causou essas fidedignidades relativamente baixas. Um efeito de teto é uma explicação sensata considerando que a idade média da amostra de Estudo 1 foi de quase 7 anos, e a faixa etária do SON-R 2½-7[a] é entre 2 anos e 6 meses e 7 anos e 11 meses. Assim, para essa amostra os itens do SON-R 2½-7[a] ficaram muito fáceis, com uma porcentagem média de acerto de 79%.

No Estudo 2 o SON-R 6-40 foi aplicado junto com a forma A da BPR-5 em 68 alunos. Os índices obtidos de fidedignidade neste estudo do SON-R 6-40 foram

satisfatórios e muito semelhantes aos índices encontrados no Estudo 1. O coeficiente de consistência interna da bateria foi de 0,95 e os dos subtestes variam entre 0,87 e 0,91. Os parâmetros psicométricos do segundo instrumento que tem o objetivo de medir o construto de inteligência aplicado neste estudo, a forma A da BPR-5, foram todos muito satisfatórios mostrando coeficientes de consistência interna da bateria completa de 0,98 e dos subtestes variando entre 0,92 e 0,95. A BPR-5 foi um teste difícil para os adolescentes da amostra: a média percentagem de acerto dos itens da bateria foi de 33%.

O próximo passo depois de calcular os parâmetros psicométricos dos testes aplicados nos dois estudos foi o cálculo da correlação entre o SON-R 6-40 e os outros dois instrumentos aplicados. Foram calculados tanto as correlações brutas como as correlações corrigidas para atenuação. A correção para atenuação é necessária uma vez que sem essa correção a relação entre os testes será subestimada. Também foram calculados os intervalos de 90% de confiança das correlações corrigidas do SON-R 6-40 com os outros testes de inteligência. O SON-R 6-40 mostrou uma correlação corrigida de 0,75 (IC 90%: $0,61 \leq 0,84$) com o SON-R 2½-7[a] e uma correlação corrigida de 0,72 (IC 90%: $0,61 \leq 0,81$) com a forma A da BPR-5. Ambas as correlações fornecem evidências fortes da validade convergente do SON-R 6-40. As correlações entre os testes completos nos dois estudos podem ser consideradas satisfatórias ($r = 0,75$ entre o SON-R 6-40 e o SON-R 2½-7[a] e $r = 0,72$ entre o SON-R 6-40 e a BPR-5).

Em comparações específicas entre os subtestes do SON-R 6-40 no estudo 1, a maior correlação ocorreu entre os subtestes Padrões e Mosaicos ($r = 0,84$), enquanto a menor correlação ocorreu entre os subtestes Categorias e Mosaicos ($r = 0,31$). No estudo 2 a maior correlação entre os subtestes do SON-R 6-40 também ocorreu entre os subtestes Padrões e Mosaicos ($r = 0,76$) e a correlação menor ocorreu entre os subtestes Padrões e Categorias ($r = 0,48$). Essas correlações eram esperadas, tendo em vista que os subtestes

Padrões e Mosaicos fazem parte da mesma escala (a Escala de Execução) e que o subteste Categorias faz parte de uma escala diferente (a Escala de Raciocínio).

Para o estudo 2, a correlação encontrada entre os escores totais do teste BPR-5 e o subteste Raciocínio Abstrato (RA) foi a maior dentre todas ($r = 0,87$), e o menor correlação foi encontrada o BPR-5 e o subteste Raciocínio Numérico (RN) ($r = 0,78$). Tal fato pode ter ocorrido devido ao elevado nível de dificuldade do subteste RN em comparação com os demais subtestes. Outra possível explicação para a ocorrência desse fato é que o subteste RN foi o último apresentado para os participantes, o que pode ter desmotivado a realização do mesmo, com a ausência ou “chute” de respostas.

É importante ressaltar aqui que foram encontradas correlações mais elevadas entre o SON-R 6-40 e os subtestes Raciocínio Abstrato e Raciocínio Espacial da BPR-5 ($r=0,68$ para ambos os subtestes) do que com os subtestes Raciocínio Verbal e Raciocínio Numérico ($r=0,62$ e $r=0,44$, respectivamente). Os dois primeiros subtestes da BPR-5 medem principalmente a inteligência fluida e os dois últimos a inteligência cristalizada. O fato que as correlações do SON-R 6-40 com os subtestes da BPR-5 que medem inteligência fluida são mais altas, é uma evidência a mais da validade convergente do SON-R 6-40 que também visa medir a inteligência fluida. Concluindo podemos afirmar que os dois estudos de validade forneceram evidências fortes da validade do SON-R 6-40 no contexto brasileiro.

6. Limitações da pesquisa de validade convergente do SON-R 6-40

A seguir são apresentadas as limitações da atual pesquisa. Em primeiro lugar temos que mencionar como uma limitação da pesquisa o número reduzido de participantes em os dois estudos. Com um número maior de participantes seria possível avaliar com maior precisão os parâmetros psicométricos dos testes e as correlações entre eles. Com o

número maior de aplicações seria possível utilizar a técnica de análise fatorial para obter evidências da validade de construto do SON-R 6-40. Uma outra limitação é que não foram coletados dados de contexto dos alunos e de suas famílias. Variáveis como o nível socioeconômico e dados sobre o desempenho escolar teriam fornecidas mais evidências da validade do SON-R 6-40. Por última, existe a limitação da representatividade da amostra: foram apenas incluídas participantes do Distrito Federal.

Referências

- Almeida, L. S. (1986). *Bateria de Provas de Raciocínio Diferencial (BPRD)*. Porto: Faculdade de Psicologia e Ciências da Educação.
- Almeida, L. S. (1988). *O raciocínio diferencial de jovens*. Porto: INIC.
- Almeida, L. S., & Primi, R. (2000). *BPR-5: Bateria de provas de raciocínio: manual técnico*. São Paulo: Casa do Psicólogo.
- Almeida, L.S., & Primi, R. (2004). Perfis de capacidades cognitivas na Bateria de Provas de Raciocínio (BPR-5). *Psicologia Escolar e Educacional* 8(2), 135-144.
- Almeida, L., Nascimento, E. do, Lima, A. O. F., Vasconcelos, A. G., Akama, C. T., & Santos, M. T. (2010). Bateria de Provas de Raciocínio (BPR-5): Estudo Exploratório em Alunos Universitários. *Avaliação Psicológica*, 9(2), 155-162.
- Anastasi, A. & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artmed.
- Baumgartl, V. de, O., & Nascimento, E. (2004). A Bateria de Provas de Raciocínio (BPR-5) aplicada a um contexto organizacional. *Psico-USF* 9(1), 1-10.
- Bernstein, B. (1960) Language and Social Class. *British Journal of Sociology*, 11(3), 261-276.
- Binet, A., & Simon, T. (1911). *A method of measuring the development of the intelligence of young children*. Lincoln, IL: Courier. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 3-19). Cambridge: Cambridge University Press.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 69-76). New York: The Guilford Press.
- Cobêro, C., Primi, R., & Muniz, M. (2006). Inteligência emocional e desempenho no trabalho: um estudo com MSCEIT, BPR-5 e 16PF. *Paidéia (Ribeirão Preto)*, 16(35), 337-348.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory* (2nd ed.). Ohio: Cengage Learning.
- Davidson, J. E., & Kemp, I. A. (2011). Contemporary models of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 58-82). Cambridge: Cambridge University Press.
- Elliott, C. D. (1990). *Differential Ability Scales*. San Antonio, TX: The Psychological Corporation.
- Elliott, C. D. (2007). *Differential Ability Scales II*. San Antonio, TX: The Psychological Corporation.
- Flynn. (2000) IQ trends over time: Intelligence, race, and meritocracy. In K. Arrow, S. Bowles, & S. M. Durlauf (Eds.), *Meritocracy and economic inequality* (pp. 35-61). Princeton, NJ: Princeton University Press.
- Hogan, T. P. (2006). *Introdução à prática de testes psicológicos*. Rio de Janeiro: LTC – Livros e Técnicos e Científicos Editores S.A.
- Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443-451). New York: Macmillan.

- Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 41-76). New York: The Guilford Press.
- Jesus, G. R., de (2009). *Normatização e validação do teste não-verbal de inteligência SON-R 2½-7[a] para o Brasil*. Tese de Doutorado, Universidade de Brasília, Brasília.
- Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2012). A history of intelligence test interpretation. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 56-70). New York: The Guilford Press.
- Karino, C. A., Laros, J. A., & Jesus, G. R. (2011). Evidências de validade convergente do SON-R 2½-7[a] com o WISC-III e WPPSI-III. *Psicologia: Reflexão e Crítica*, 24, 621-629.
- Keith, T. Z., & Reynolds, M. R. (2012). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 758-799). New York: The Guilford Press.
- Laros, J. A., Reis, R. F., & Tellegen, P. J. (2010). Indicações da validade convergente do teste não-verbal de inteligência SON-R 2½-7[a]. *Avaliação Psicológica*, 9, 43-52.
- Laros, J. A., Tellegen, P. J., Jesus, G. R., de & Karino, C. A. (2012). *Teste não-verbal de inteligência. SON-R 2½-7[a]: Manual técnico*. Porto Alegre: Casa do Psicólogo.
- Lichtenberger, O. & Broadbooks, D. (2000). *Essentials of Cognitive Assessment with KAIT and other Kaufman Measures*. John Wiley.
- Lissitz, R. W. (2009). *The concept of validity: Revisions, New Directions and Applications*. Information Age Publishing.
- Moore, D.S. (2005). *A estatística básica e sua prática*. LTC.
- Oakland T. (2005). Selected ethical issues relevant to test adaptations. In R. K. Hambleton, Meranda, P. F., & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 65-92). Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Pasquali, L. (2012). *Análise Fatorial para Pesquisadores*. LABPam.
- Primi, R., & Almeida, L. S. (2000). Estudo de validação da Bateria de Provas de Raciocínio (BPR-5). *Psicologia: Teoria e Pesquisa*, 16(2), 165-173.
- Primi, R., & Almeida, L. S. (2002). Inteligência Fluida: Definição Fatorial, Cognitiva e Neuropsicológica. *Paideia*, 12(23), 57-75
- Primi, R., Couto, G., Almeida, L. S., Guisande, M. A., & Miguel, F.K. (2012). Intelligence, Age and Schooling: Data from the Battery of Reasoning Tests (BPR-5). *Psicologia: Reflexão e Crítica*, 25(1), 79-88.
- Primi, R., Silva, M. C. R. de, Rodrigues, P., Muniz, M., & Almeida, L. (2013). The use of the bi-factor model to test the uni-dimensionality of a battery of reasoning tests. *Psicothema*, 25(1), 115-122.

- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 136-181). New York: The Guilford Press.
- Nascimento, E. (2000). *Adaptação e validação do teste WAIS-III para um contexto brasileiro*. Tese de Doutorado, Universidade de Brasília, Brasília.
- Osborne J. W. (2003) Effect sizes and the disattenuation of correlation and regression coefficients: lessons from educational psychology. *Practical Assessment, Research & Evaluation*, 8(11),
- Schelini, P.W. (2006). Teoria das Inteligências Fluida e Cristalizada: Início e Evolução. *Estudo de Psicologia (Natal)*, 11(3), 323-332.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 99-144). New York: The Guilford Press.
- Snijders-Oomen, N. (1943). *Intelligentie-onderzoek van doofstomme kinderen* [The examination of intelligence of deaf-mute children]. Nijmegen: Berkhout.
- Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Tellegen, P. J., & Laros, J. A. (2011). *Research report of the nonverbal intelligence test SON-R 6-40*. Göttingen, Germany: Hogrefe GMBH (178 p.).
- Tellegen, P. J., & Laros, J. A. (2004). Cultural Bias in the SON-R Test: Comparative Study of Brazilian and Dutch Children. *Psicologia: Teoria e Pesquisa*,
- Urbina, S. (2006). *Fundamentos da Testagem Psicológica*. Porto Alegre; Artmed.
- Wechsler, D. (2004). *WAIS-III: manual para administração e avaliação*. Adaptação brasileira: E. Nascimento. São Paulo: Casa do Psicólogo.
- Wechsler, D. (2013). *WISC-IV: manual técnico*. Adaptação brasileira: F. J. M. Rueda, A. P. P. Noronha, F. F. Sisto, A. A. A. dos Santos e N. R. de Castro. São Paulo: Casa do Psicólogo.
- Wechsler, D. (2012). *WPPSI-IV: Technical And Interpretative Manual Supplement: Special Group Validity Studies And Other Measures*. PsychCorp.
- Wilis, J. O., Dumont, R., Kaufman, A.S. (2011). Factor-analytic models of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 39-57). Cambridge: Cambridge University Press.
- Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. S. Rao & S. Sinharay (Eds.), *Handbook of statistics*, Vol. 26 (pp. 45-80). Amsterdam: Elsevier B.V.