

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

**NORMATIZAÇÃO E VALIDAÇÃO DO TESTE NÃO-VERBAL
DE INTELIGÊNCIA SON-R 2½-7[a] PARA O BRASIL**

Girlene Ribeiro de Jesus

Orientador: Prof. Dr. Jacob Arie Laros

Brasília – DF
Fevereiro de 2009

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

**NORMATIZAÇÃO E VALIDAÇÃO DO TESTE NÃO-VERBAL
DE INTELIGÊNCIA SON-R 2½-7[a] PARA O BRASIL**

Girlene Ribeiro de Jesus

Tese de doutorado apresentada ao Programa de Pós-graduação em Psicologia Social, do Trabalho e das Organizações, como requisito parcial para a obtenção do grau de doutor em Psicologia Social, do Trabalho e das Organizações.

Orientador: Prof. Dr. Jacob Arie Laros

Brasília – DF
Fevereiro de 2009

**NORMATIZAÇÃO E VALIDAÇÃO DO TESTE NÃO-VERBAL
DE INTELIGÊNCIA SON-R 2½-7[a] PARA O BRASIL**

Tese de doutorado defendida diante e avaliada pela banca constituída por:

Prof. Ph.D. Jacob Arie Laros (Presidente)
Universidade de Brasília

Prof. Ph.D. Dalton Francisco Andrade (Membro)
Universidade Federal de Santa Catarina

Prof^a. Dr^a. Rosana Maria T. Tristão (Membro)
Universidade de Brasília

Prof. Dr. Valdiney Veloso Gouveia (Membro)
Universidade Federal da Paraíba

Prof. Docteur. Luiz Pasquali (Membro)
Universidade de Brasília

Prof^a. Dr^a. Isolda de Araújo Günther (Suplente)
Universidade de Brasília

Agradecimentos

Ao Deus Onisciente, Onipotente, Onipresente, por todas as coisas.....

Às minhas duas famílias, uma de sangue: Maria Creuza, José Clementino, Luiz, José João, Reis, Gilson, Gardênia, Neide, Gleiciane e Vânia, por todo o apoio, incentivo, carinho, compreensão e pelo espírito de doação; outra do coração: Josué e Maria da Guia, por serem dois anjos em minha vida, por tudo que fizeram e fazem por mim de coração, por serem os responsáveis diretos por eu estar hoje na Psicologia, não tenho como agradecer-los por tudo.

Ao meu amigo e orientador, Professor Jacob Arie Laros, responsável direto pelo êxito desse trabalho. Sempre calmo, paciente e com muita disposição para ensinar e corrigir. Presente em todos os momentos. Tenho muito que agradecer!

Ao amigo e professor Valdiney, por ter me mostrado o mundo da pesquisa e influenciado diretamente minha escolha dentro da Psicologia.

À professora Rosana, pela orientação em muitos momentos, pela preciosa ajuda durante os estudos com o teste no Distrito Federal e pelas proveitosas discussões.

Aos demais membros da banca, por terem prontamente aceito o convite para colaborarem com o presente trabalho.

Ao professor Neto, por ser um grande amigo e incentivador.

Ao professor Peter J. Tellegen, da Universidade de Groningen, pelo programa de normatização e pela contribuição fundamental nas análises.

À minha amiga Maja, por tudo que fez por mim ao chegarmos em Brasília, seu apoio e ajuda foram essenciais.

À minha amiga Fabiana, sempre presente, colaborando comigo em todas as coisas, não tenho como agradecer.

Ao CESPE/UnB e aos amigos da Coordenadoria de Pesquisa em Avaliação pelo apoio. Ao CNPq e à Universidade de Groningen, Holanda, pelo apoio financeiro.

À minha amiga Camila Akemi, meu braço direito na pesquisa, responsável pelo acompanhamento de toda a coleta e finalização da base de dados.

A todos que contribuíram com o nosso grupo de pesquisa: Cleide, Juliana, Mica, Lana, Nilza e Rafaela. A todos os estudantes de graduação e pós-graduação que colaboraram com esse projeto.

A todos os pais e responsáveis pelas crianças que participaram desse estudo. Muito obrigada!

Agradecimento às instituições

Nossos agradecimentos sinceros às seguintes instituições escolares,
fundamentais para a realização da presente pesquisa:

- Manaus-AM**
CMEI Nossa Senhora da Paz
Escola Municipal José Tavares de Macedo
- Araguaína-TO**
Creche Dona Regina
Escola Lar Adventista
- Tocantinópolis-TO**
Centro Educacional Nossa Escolinha
Escolinha Jardim da Mônica
Escola Paroquial Cristo Rei
- Augustinópolis-TO**
Escola Municipal Presidente Kennedy
- Salvador-BA**
Escola Paroquial de Santana
Escola Cid Passos
Escola Municipal Vale das Pedrinhas
- São Luís-MA**
Colégio Adventista de São Luís
Universidade Infantil Rivana Berenice
- Fortaleza-CE**
Escola Alba Frota
Escola São Vicente de Paula
- Imperatriz-MA**
Colégio Adventista de Imperatriz
Escola Raio de Sol
Escola Aquarela
- Estreito-MA**
Igreja Alto Bonito
- Sobral-CE**
Escola de Ens. Fund. Rainha da Paz
- São João do Paraíso-MA**
Escola Municipal Natividade Marinho
Escola Municipal Sebastião Archer
- Irará-BA**
Creche Municipal Elysio Santana
Escola Municipal Allan Kardec
Escola Profª. Alzira Dantas de Oliveira
- Ipecaetá-BA**
Creche Olívia Carneiro
Escola Nova Geração
Creche Júlia Medeiros
Escola Municipal Antero Alves
Escola João Durval Carveiro
- Acarape-CE**
Núcleo Intelectivo Fênix
Creche Raio de Luz
Escola Adélia Jacó
- São Paulo-SP**
CEI Parque Casa de Pedra
Igreja Presbiteriana de Tucuruvi
Colégio Madrid
Sistema SESI/SP de Escolas
- Belo Horizonte-MG**
Escola Municipal Arthur Guimarães
Centro Educacional Casa Verde
- TanGram Sistema de Ensino*
- Rio de Janeiro-RJ**
Escola NAU
Escola CCBF
- Campinas-SP**
Escola Criativa
CEMEI Jardim Maria Rosa
EMEI Prof. Manuel Affonso Ferreira
- Guarulhos-SP**
Escola Mun. Dalva Marina R. Mingossi
- Sumaré-SP**
EMEI do Jardim São Judas Tadeu
Escola Nilza Thomazini
- Esmeraldas-MG**
Escola Pingo de Gente
- São José do Vale do Rio Preto-RJ**
Escola Municipal Santa Isabel
- Ribeirão Branco-SP**
Escola Branca de Neve
Escola Sonho Meu
Escola Hilda Matarazo
- Porto Alegre-RS**
Escola Brincar e Aprender
Escola Pequena Casa da Criança
- Curitiba-PR**
Conselho Tutelar Santa Felicidade
Escola Guaíra
Escola Nice Braga
- São José dos Pinhais-PR**
Centro de Atenção da Juventude
Escola Baba de Passarinho
Escola Meu Tesouro
Escola Mário Flores
- São Gabriel-RS**
Unidade Básica de Saúde Brandão Júnior
Escola de Ensino Médio João Pedro Nunes
- Dom Feliciano-RS**
Escola Catulino Pereira da Rosa
Projeto Social Paz e Bem
Creche Municipal Recanto Amigo
- Brasília-DF**
Jardim de Infância da 308 sul
Creche Estrela Guia
Escola Classe 206 Sul
- Goiânia-GO**
Igreja- Pastoral da Criança
Externato São José
- Valparaíso de Goiás-GO**
Igreja Católica Franciscana de Valparaíso
Escola Rainha da Paz
- Pirenópolis-GO**
Escola Tia Olívia
Creche Aldeia da Paz
E as demais não citadas...

Agradecimento aos examinadores

A realização da coleta de dados só foi possível devido à significativa participação dos psicólogos e estudantes listados a seguir, a eles nossos sinceros agradecimentos:

Alessandra Maia Vitor (Psicóloga)

Amanda Ladislau Leonardo (Psicóloga)

Ana Clara de Sousa Bittencourt Bastos (Estudante)

Ana Cristina Passos Gomes (Psicóloga)

Camila Akemi Karino (Psicóloga)

Carolina Silva Bandeira de Melo (Psicóloga)

Cinthia Figueiredo de Norões Brito (Psicóloga)

Clara Mutti Vasconcellos (Psicóloga)

Darlene Cardoso Ferreira (Psicóloga)

Elaine Arima (Psicóloga)

Enio Santos (Estudante)

Érica Pardini Marino (Psicóloga)

Iara Sousa Lima (Psicóloga)

Joice Ferreira da Silva (Psicóloga)

Leticia Seicenti Silva (Psicóloga)

Luciana de Avila Quevedo (Psicóloga)

Mariana Saissu Sousa (Estudante)

Tiago Augusto (Psicólogo)

Shana Gularte Della Vechia (Estudante)

Vilma Barbosa Pellegrino (Psicóloga)

William Martins de Oliveira (Psicologia)

Zuleide Oliveira Feitosa (Psicóloga)



Dedicatória

Dedico esse trabalho a todas as crianças com idade entre 3 e 8 anos, que tornaram possível a realização da pesquisa.



Não que sejamos capazes por nós mesmos de pensar alguma coisa como se partisse de nós, mas a nossa capacidade vem de Deus. I Cor. 3:5

Sumário

1. Introdução	1
1.1 Justificativa	3
1.2 Objetivos	4
2. A inteligência	6
2.1 O conceito de inteligência	6
2.2 Metáforas sobre a inteligência	9
2.3 Histórico de alguns modelos de inteligência	12
2.4 O modelo hierárquico de habilidades cognitivas de Cattell e Horn	14
2.5 A teoria dos três estratos	17
3. Histórico dos testes SON de 1943-2009	19
3.1 O teste não-verbal de inteligência SON-R 2½-7	22
3.1.1 Estudos de validação do SON-R 2½-7	24
3.2 O teste não-verbal de inteligência SON-R 2½-7[a]	26
3.2.1 Descrição dos subtestes do SON-R 2½-7[a]	28
4. Método	32
4.1 Amostra	32
4.1.1 Estudo piloto	32
4.1.2 Estudos de validação	32
4.1.3 Pesquisa nacional de normatização	33
4.2 Instrumentos	46
4.3 Procedimento	48
4.3.1 Procedimento utilizado na pesquisa nacional de normatização	49
4.4 Análise dos dados	53
5. Resultados do estudo piloto	54
5.1 Resultados do estudo de validação de conteúdo	54
5.2 Resultados do estudo piloto do SON-R 2½-7	58
6. Normatização do SON-R 2½-7[a]	61
6.1 O Modelo de Regressão Logística	62
6.2 Os escores Escalares	69
6.3 Resultados da Normatização	69
7. Características Psicométricas do SON-R 2½-7[a]	72
7.1 Resultados do estudo de validação	73
7.1.1 Análise da consistência interna	74
7.1.2 Análise Fatorial	78
7.1.2.1 Análise Fatorial de Informação Plena – FIFA.....	78
7.1.2.2 Análise Simultânea dos Componentes Principais – SCA.....	80
7.1.2.3 Análise Fatorial Confirmatória – CSA	82
7.1.2.4 Análise Fatorial de Segunda Ordem	84
7.1.3 Análise dos itens pela Teoria de Resposta ao Item	86
7.2 Análise por hipótese	90
8. Relação do SON-R 2 ½ -7[a] com variáveis externas	95
8.1 Nível socioeconômico (NSE).....	95
8.2 Duração da testagem	97
8.3 Influência do examinador	98
8.4 Diferenças regionais	99
8.5 Influência do gênero	100
8.6 Avaliação realizada pelos examinadores	101

9. Considerações Finais	103
10. Limitações e Agenda de Pesquisa	107
11. Referências	118

Lista de Tabelas

Tabela 4.1 Número de cidades, número de crianças de 1 a 9 anos, e percentagem de crianças de 1 a 9 anos por categoria de IDH	33
Tabela 4.2 Amostra para a pesquisa nacional	34
Tabela 4.3 Quantidade de cidades por região e quantidade de elementos a serem testados por região	35
Tabela 4.4 Número de municípios e elementos por categoria de IDHM	35
Tabela 4.5 Quantidade de elementos em cada região por categoria de IDHM	36
Tabela 4.6 Municípios selecionados em cada região	37
Tabela 4.7 Percentual de crianças que freqüentam escola nos municípios selecionados (Censo de 2000)	39
Tabela 4.8 Plano amostral segundo a quantidade de crianças escolares e não escolares	40
Tabela 4.9 Comparação dos percentuais de crianças não-escolares	41
Tabela 4.10 Plano amostral com ajuste da quantidade de crianças não-escolares, segundo dados da PNAD 2006	41
Tabela 4.11 Quantidade de crianças matriculadas na rede pública e privada (Censo Escolar de 2006)	43
Tabela 4.12 Quantitativo previsto e obtido de crianças por cidade e estado na amostra final	44
Tabela 4.13 Quantitativo de crianças por faixa de IDH das cidades	45
Tabela 4.14 Examinadores e municípios nos quais eles aplicaram o SON-R 2½-7[a]...	52
Tabela 5.1 QI de execução, raciocínio e total das crianças da creche e do CEAL	59
Tabela 5.2 Percentual de acerto, correlação item-total e fidedignidade dos subtestes do SON-R 2½-7	60
Tabela 6.1 Coeficientes da regressão linear múltipla entre o escore e a idade transformados	64
Tabela 6.2 Estatísticas descritivas dos escores digitalizados	66
Tabela 6.3. Coeficientes β da regressão logística dos escores digitalizados ajustados pela idade e o escore	66
Tabela 6.4 Coeficientes β do ajuste do e.p.m pela idade	69
Tabela 6.5 Estatísticas descritivas dos escores normatizados e do QI do SON-R 2½-7[a]	70
Tabela 6.6 Normas brasileiras para conversão dos escores brutos em escores normatizados (M=10, DP=3) do SON-R 2½-7[a] nos grupos de idade 3;3 e 7;9	71
Tabela 6.7 Normas brasileiras para conversão dos escores normatizados em QI (M=100, DP=15) do SON-R 2½-7[a]	71
Tabela 7.1 Média e desvio padrão dos escores brutos por grupo de idade	72
Tabela 7.2 Proporção de acerto dos itens dos subtestes do SON-R 2½-7[a]	73
Tabela 7.3 Fidedignidade (lambda 2 de Guttman) e desvio-padrão (DP) dos escores brutos	75
Tabela 7.4 Fidedignidade (lambda 2 de Guttman) e erro padrão de mensuração (e.p.m) dos escores normatizados do SON-R 2½ -7[a]	76
Tabela 7.5 Generalizabilidade e erro padrão de estimação para os escores normatizados da Escala de Execução (EE), Escala de Raciocínio (ER) do SON-R 2½-7[a]	77

Tabela 7.6 Cargas fatoriais dos itens do SON-R 2½-7[a] no fator único e a média por subteste	79
Tabela 7.7 Cargas fatoriais dos itens do SON-R 2½-7[a] nos dois fatores e a média por fator	80
Tabela 7.8 Autovalor e percentual de variância explicada por componente e grupo de idade dos dois componentes principais via análise SCA	82
Tabela 7.9 Coeficientes de determinação (cargas fatoriais), erro padrão e correlação múltipla ao quadrado (SMC) da AFC do SON-R 2½-7[a]	83
Tabela 7.10 Coeficientes de determinação (cargas fatoriais) do SON-R 2½-7[a] nos fatores de primeira e segunda ordem.....	85
Tabela 7.11 Solução Schmid-Leiman para o SON-R 2½-7[a] nos fatores de primeira e segunda ordem	86
Tabela 7.12 Parâmetros <i>a</i> e <i>b</i> dos 60 itens do SON-R 2½-7[a]	89
Tabela 7.13 Coeficientes de correlação entre a idade e o escore bruto nos subtestes do SON-R 2½-7[a]	91
Tabela 7.14 Coeficientes de correlação entre as escalas do SON-R 2½-7[a] e do WPPSI-III e o intervalo de confiança de 95% das correlações	92
Tabela 7.15 Coeficientes de correlação entre as escalas do SON-R 2½-7[a] e do WISC-III e o intervalo de confiança de 95% das correlações	93
Tabela 7.16 Coeficientes de correlação entre as escalas do SON-R 2½-7[a], as MPC de Raven e o CMMS (Colúmbia) e o intervalo de confiança de 95% das correlações...	94
Tabela 8.1 Grau de escolaridade dos pais ou responsáveis	95
Tabela 8.2 Cargas fatoriais no fator nível socioeconômico (NSE)	96
Tabela 8.3 Comparação dos escores normatizados nos postos percentílicos do nse.....	97
Tabela 8.4 Duração da aplicação por faixa de idade, de tempo e por subteste	98
Tabela 8.5 Estatísticas descritivas do QI das crianças por examinador	99
Tabela 8.6 Média, desvio padrão e intervalo de confiança da média para o QI das crianças por região e localidade	100
Tabela 8.7 Média, desvio padrão e estatísticas do teste <i>t</i> para a comparação de médias por sexo	101
Tabela 8.8 Percentual, média e desvio padrão por situação de testagem e grupo de idade	102

Lista de Figuras

Figura 3.1 Histórico dos testes SON de 1943 a 2009, com base em Drunen (2003)	20
Figura 3.2 Itens do subtteste Mosaicos	28
Figura 3.3 Itens do subtteste Categorias	29
Figura 3.4 Itens do subtteste Situações	30
Figura 3.5 Itens do subtteste Padrões	31
Figura 5.1 Mudanças do SON-R 2½-7 para o SON-R 2½-7[a].....	54
Figura 5.2 Mudanças no item 10 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a].....	55
Figura 5.3 Mudanças no item 11 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a].....	56
Figura 5.4 Mudanças no item 12 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a].....	56
Figura 5.5 Mudanças no item 13 (alternativas) do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a].....	57
Figura 5.6 Mudanças no item 13 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a].....	57
Figura 5.7 Mudanças no item 14 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a].....	58
Figura 8.1 Distribuição da amostra segundo o nível socioeconômico	96

Resumo

Os objetivos principais da presente tese foram estabelecer normas brasileiras para o teste não-verbal de inteligência SON-R 2½-7[a], originalmente desenvolvido na Holanda, e obter evidências sobre sua validade no Brasil. Para tanto, foi realizado inicialmente um estudo de validação de conteúdo de quatro subtestes (Mosaicos, Categorias, Situações e Padrões) do SON-R 2½-7, que avalia raciocínio abstrato, concreto e espacial. Esse estudo resultou em algumas mudanças, implementadas no subteste Categorias. A seguir foi realizado um estudo piloto com crianças de uma instituição de ensino regular e uma instituição destinada a crianças com problemas auditivos. Os resultados indicaram boa qualidade psicométrica do teste e todos os subtestes apresentaram índices de fidedignidade satisfatórios. Após o estudo piloto, os estudos de normatização e validação foram iniciados. O processo de normatização foi realizado com uma amostra nacional composta por 1.200 crianças de 3 anos e 3 meses a 7 anos e 9 meses de idade. As crianças que participaram do estudo eram provenientes de todas as regiões brasileiras, de 13 estados diferentes. Cada grupo de idade foi representado por uma amostra de 120 crianças, que foi estratificada segundo o sexo, situação educacional e variáveis demográficas. Um passo de fundamental importância na normatização foi a transformação dos escores brutos em escores distribuídos normalmente com média e desvio fixos. Tais transformações geralmente são efetuadas separadamente para cada grupo de idade. Não obstante, na pesquisa atual um método diferente foi utilizado: as distribuições dos escores para todos os grupos de idade foram ajustadas simultaneamente como uma função contínua da idade. O procedimento de ajuste foi usado a fim de minimizar as diferenças entre a distribuição observada e a distribuição populacional estimada. O método foi desenvolvido na Holanda para a normatização do SON-R 5½-17, sendo também utilizado para a normatização do SON-R 2½-7. A grande vantagem de sua utilização é que o uso da informação de todos os grupos simultaneamente torna a normatização mais precisa. Outro benefício essencial desse método é a possibilidade de calcular escores normatizados para cada idade exata na faixa pesquisada. Uma vez finalizada a normatização foi verificada a fidedignidade do teste: o índice médio de consistência interna dos subtestes foi 0,79. O subteste Categorias mostrou a fidedignidade média mais alta (0,81) e Situações a mais baixa (0,77). A fidedignidade média do SON-R QI foi 0,92. A validade de construto do teste foi verificada por meio da Análise Fatorial Confirmatória e da solução Schmid-Leiman. Os resultados obtidos embasam a distinção que é feita no SON-R 2½-7[a] entre a Escala de Execução e a Escala de Raciocínio bem como um fator geral. A validade convergente do SON-R 2½-7[a] foi verificada com os seguintes testes de inteligência: *Wechsler Preschool and Primary Scale of Intelligence-Third Edition* (WPPSI-III), a Escala Wechsler de Inteligência para Crianças-Terceira Edição (WISC-III), a Escala de Maturidade Mental Columbia (EMMC) e as Matrizes Progressivas Coloridas (MPC) de Raven. Com esses testes o SON-R 2½-7[a] mostrou os seguintes coeficientes de validação: 0,75, 0,67, 0,68 e 0,56, respectivamente. O alto índice de correlação com o WPPSI-III (0,75) sugere que o SON-R 2½-7[a] está medindo de forma mais ampla o construto de inteligência que as MPC de Raven e a EMMC. Assim, o presente estudo disponibiliza um instrumento de avaliação de habilidades cognitivas válido, fidedigno e com normas brasileiras para a avaliação de crianças pré-escolares, sendo o mesmo capaz de contribuir para diagnósticos clínicos de atrasos ou dificuldades específicas do desenvolvimento cognitivo de crianças em faixa etária prioritária para intervenção precoce.

Palavras-chave: Teste não-verbal de inteligência para crianças, SON-R 2½-7[a], normatização, validação, procedimento de ajuste simultâneo de grupos de idade separados.

Abstract

The main objectives of this doctoral thesis were to establish Brazilian norms for the non-verbal test of intelligence SON-R 2½-7[a], originally developed in the Netherlands, and to obtain evidence about its validity in Brazil. To accomplish these goals, initially, a validation study was undertaken of the content of the four subtests (Mosaics, Categories, Situations and Patterns) of the SON-R 2½-7[a] which assess abstract and concrete reasoning and spatial insight. This first study resulted in a number of changes in items of the subtest Categories. As a second step a pilot study was realized with children of an institution of regular education and an institution for children with hearing problems. The results of this study indicated good psychometric qualities of the test as a whole, and satisfactory reliability indices of the individual subtests. After the pilot study the normatization and validation studies were initiated. The normatization was performed using a national sample of children of Brazil, consisting of 1,200 subjects from 3 years and 3 months to 7 years and 9 months of age. The children that participated came from all five regions in Brazil and from 13 different states. Each age group was represented by a sample of 120 children that was stratified according to sex, educational type and demographic variables. A step of vital importance in the normatization was the transformation of raw subtest scores to normally distributed scores with a fixed mean and standard deviation. These transformations are usually carried out separately for each age group. However, in this research a different method was used: the score distributions for all age groups were fitted simultaneously as a continuous function of age. This fitting procedure was used in order to minimize the difference between the observed and the estimated population distribution. The method was developed in the Netherlands for the normatization of the SON-R 5½-17 and was also used for the normatization of the SON-R 2½-7. A major advantage of this method is that the use of information on all groups simultaneously makes the normatization much more accurate. Another essential benefit of this method is the possibility to calculate norm scores for each exact age in the examined age range. Once the normatization was concluded, the reliability of the test scores was examined: the mean internal consistency of the subtests was .79. Categories showed the highest mean reliability coefficient (.81) and Situations the lowest mean reliability index (.77). The mean reliability of the SON-R IQ score is .92. The construct validity of the SON-R was investigated by means of Confirmatory Factor Analysis and Schmid-Leiman solution. Results of this analysis supported the distinction made in the SON-R 2½-7 between the Performance Scale and the Reasoning Scale and a general factor. The convergent validity of the SON-R 2½-7[a] was examined by the combined administration of the SON-R test with the following intelligence tests: the Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III), the Wechsler Intelligence Scale for Children-Third Edition (WISC-III), the Columbia Mental Maturity Scales (CMMS) and Raven's Colored Progressive Matrices (CPM). With these tests the SON-R showed, in the above mentioned order, the following validation coefficients: .75, .67, .68 and .56. The relative high correlation with the WPPSI-III (.75), a test battery of a broad intelligence construct, suggests that the SON-R 2½-7 [a] is measuring a broader intelligence construct than Raven's CPM and the Columbia. With the present study a valid, reliable instrument that assesses cognitive abilities of young children with Brazilian norms becomes available. This instrument will hopefully contribute to improve clinical diagnoses of specific difficulties in the cognitive development of children in the priority age group for early intervention.

Key words: Non-verbal test of intelligence for young children, SON-R 2½-7[a], normatization, validation, simultaneously fitting procedure of separate age groups.

1. Introdução

De acordo com Snijders, Tellegen e Laros (1989), em termos gerais a inteligência pode ser descrita como o quanto alguém é capaz de desempenhar tarefas que exigem habilidades cognitivas. Ser capaz de usar essas habilidades é visto como uma característica individual, a qual se quer avaliar.

Acerca desse tema, Brody (2000) argumenta que todos os pioneiros da área da inteligência acreditavam que esta se tratava de um traço hereditário. Entretanto, todos eles também tinham consciência de que o desenvolvimento da inteligência poderia ser atrapalhado pela inadequabilidade do meio. Spearman (1904), por exemplo, verificou que a habilidade de discriminação sensorial, que ele equivocadamente pensava ser o coração da inteligência, era influenciada pelo meio (associado ao nível socioeconômico). No teste de Binet também eram notadas diferenças no desempenho de crianças com diferentes níveis socioeconômicos.

Ao longo dos tempos surgiram muitos modelos que procuraram dar conta do que seria a inteligência e qual é sua fonte ou o que pode influenciá-la. Os modelos hierárquicos, por exemplo, têm como pressuposto básico que a estrutura da inteligência pode ser descoberta por meio da análise da interrelação de escores em testes de habilidade mental. Assim, um conjunto de habilidades interrelacionadas é visto como um fator (Davidson & Downing, 2000). Os modelos hierárquicos atuais podem ser melhor compreendidos em termos de duas teorias psicométricas anteriores, a de Spearman e a de Thurstone. E dois modelos hierárquicos contemporâneos ajudam a conciliar a visão dos dois, é a Teoria de Inteligência Fluida e Cristalizada (Teoria *Gf-Gc*).

A Teoria *Gf-Gc* explica e prediz o desenvolvimento da inteligência ao longo da vida (Horn, 1994; Horn, Donaldson & Engstrom, 1981). A inteligência fluida é vista como o potencial para aprender, é menos suscetível à escolarização e oportunidades para aprender; a

inteligência cristalizada, por sua vez, tem a ver com conhecimentos e habilidades adquiridos por meio da escolarização e oportunidades de aprendizagem.

Os testes tradicionais de inteligência geral, como o Stanford-Binet (ver Thorndike, Hagen & Sattler, 1986) e os testes de inteligência Wechsler (Weiss, 1982) se centram mais na inteligência cristalizada, ou seja, mais no resultado final da aprendizagem do que no potencial para a aprendizagem. Segundo Tellegen e Laros (2004; 2005), testes de inteligência que aferem meramente o resultado final da aprendizagem subestimam a habilidade para aprender de pessoas que tiveram poucas oportunidades para adquirir conhecimento e habilidades para obterem um bom desempenho em uma situação de teste. Em particular, membros de minorias étnicas, pessoas com baixo nível socioeconômico e pessoas com problemas de aprendizagem estariam em desvantagem quando testadas com um teste tradicional de inteligência geral.

Além disso, de acordo com Tellegen e Laros (2005), os testes tradicionais de inteligência vêm sendo criticados em seu conteúdo por defensores de testes de inteligência culturalmente justos. Pelo fato dos testes tradicionais de inteligência geralmente fazerem apelo para habilidades de linguagem específicas, tanto nos conteúdos quanto nas instruções do teste, esses testes colocariam membros de minorias culturais em desvantagem (ver Jesus & Laros, 2006). Este argumento também se aplica a pessoas com problemas de linguagem e auditivos. Para todos esses grupos, um baixo desempenho no teste poderia refletir, primariamente, um conhecimento verbal pobre, em vez de raciocínio ou habilidade para aprendizagem pobre. Tais críticas levaram ao desenvolvimento de testes não-verbais de inteligência, com o objetivo de minimizar a dependência do conhecimento adquirido e da habilidade verbal, tais como as “Matrizes Progressivas de Raven” (Raven, 1938), os testes SON (Snijders-Oomen, 1943) e o “Cattell's Culture Fair Intelligence Test” (Cattell, 1950).

Uma das grandes vantagens dos testes não verbais diz respeito à maior facilidade para adequação a diferentes culturas, uma vez que o processo de adaptação de testes não-verbais é

menos complicado do que o exigido para testes que utilizam linguagem escrita ou falada como parte do seu conteúdo. Não obstante, o fato de os testes não-verbais não exigirem tradução não significa que esses instrumentos possam ser utilizados sem um estudo empírico que verifique sua adequação à cultura na qual será utilizado.

A esse respeito, Van de Vijver e Poortinga (1997) assinalam que não se pode assumir que instrumentos psicológicos desenvolvidos em uma cultura, quando utilizados em outra, tenham propriedades psicométricas como validade e fidedignidade simplesmente assumidas, mas precisam ser empiricamente demonstradas. O viés, por exemplo, pode mudar as propriedades psicométricas de um instrumento quando ele é usado em uma cultura diferente.

Os testes SON são indicados como exemplo de testes com conteúdo cultural reduzido, ou seja, esses testes aferem uma quantidade limitada de conteúdo cultural, conteúdo este que eles não objetivam mensurar. Ademais, os testes SON têm como foco a inteligência fluida, isto é, o potencial para aprender. Dessa forma, o Teste Não-verbal de inteligência SON-R 2½-7 foi escolhido como objeto de estudo na presente tese.

1.1 Justificativa

No contexto brasileiro, praticamente não há teste não-verbal de inteligência normatizado para toda a faixa etária contemplada pelo SON-R 2½-7[a]. De acordo com dados do Conselho Federal de Psicologia (disponível no site desse órgão), que vem avaliando todos os testes psicológicos disponibilizados no mercado brasileiro, entre os testes de inteligência para crianças encontra-se com avaliação favorável, até agosto de 2007: o Teste Não-Verbal de Raciocínio para Crianças – TNVRI, destinado à faixa etária entre 5 anos e 9 meses e 13 anos e 3 meses; a Escala de Inteligência Wechsler para Crianças (WISC-III), adaptada para crianças com idades entre 6 e 16 anos; o teste R-2 (Teste Não-Verbal de Inteligência para Crianças), adaptado com uma população de 5 a 11,5 anos (Rosa & Alves, 2000); o Colúmbia - CMMS

(Escala de Maturidade Mental Colúmbia) indicado para crianças entre 3 anos e 6 meses e 9 anos e 11 meses.

A presente pesquisa visa contribuir para amenizar a lacuna existente nesse campo no Brasil, pois se propõe a validar e estabelecer normas brasileiras para o Teste Não-verbal de Inteligência SON-R 2½-7[a], com uma amostra proveniente de todas as regiões brasileiras e representante de diferentes extratos socioeconômicos do país, uma vez que os testes disponibilizados até o momento foram validados com populações específicas de determinadas regiões e áreas metropolitanas e não com representantes das cinco regiões brasileiras (ver Alchieri, Noronha & Primi, 2003).

Disponibilizar um instrumento com normas nacionais para a avaliação de crianças pequenas é de fundamental importância para a área de avaliação psicológica e neuropsicológica infantil. Por exemplo, Campbell e Ramey (1995) apontam que a intervenção precoce é crucial para o desenvolvimento cognitivo. Esses autores realizaram um importante estudo longitudinal com crianças estadunidenses negras no qual verificaram que até aos 15 anos de idade ainda se podia verificar o efeito, sobre o desenvolvimento cognitivo, da intervenção realizada no período da infância até os cinco anos de idade.

1.2 Objetivos

Objetivo Geral

Nesse contexto, o objetivo geral da presente tese foi estabelecer normas e obter dados sobre a validade do Teste Não-verbal de Inteligência SON-R 2½-7 [a] para o Brasil.

Objetivos específicos

1. providenciar normas brasileiras do teste não-verbal de inteligência SON-R 2½-7[a] para crianças na faixa entre 2 e 8 anos;

2. verificar a estrutura fatorial do teste não-verbal de inteligência SON-R 2½-7[a] no contexto brasileiro, isto é, verificar a validade de construto do instrumento;
3. verificar a validade convergente do SON-R 2½-7[a].

A presente tese está dividida em nove seções, além da introdução, na próxima seção é apresentada uma explanação sobre o conceito e os modelos de inteligência. A seguir é apresentado o histórico dos testes SON e o método da presente pesquisa. A primeira seção de resultados é apresentada com os resultados do estudo piloto com o teste, a seguir é apresentado o modelo de normatização utilizado, a seção seguinte trata da validade de construto do teste e a próxima seção sobre a relação do teste com variáveis externas e, por fim, são traçadas considerações finais sobre o estudo.

2. A inteligência

Os indivíduos diferenciam-se uns dos outros em vários aspectos, tanto em suas habilidades de entender ideias complexas quanto na adaptação ao ambiente, a aprendizagem por meio da experiência e as formas variadas de utilização do raciocínio. Dessa forma, é razoável entender por que o desempenho intelectual de uma pessoa pode variar em ocasiões e domínios diferentes, e segundo os diferentes critérios pelos quais for julgada (Neisser & cols, 1996).

Segundo Neisser e cols. (1996) muitos teóricos argumentam que há diferentes “inteligências” (sistemas de habilidades) e apenas poucas delas podem ser aferidas com testes psicométricos tradicionais. Por sua vez, outros teóricos enfatizam o papel da cultura, determinando concepções diferentes de inteligência e influenciando a aquisição de habilidades intelectuais (Hu & Oakland, 1991). Seguindo outra direção, estudiosos do desenvolvimento humano frequentemente enfocam mais os processos pelos quais todas as crianças vêm a pensar inteligentemente do que na medição das diferenças individuais entre elas.

Nesse contexto, a seguir será tratado o conceito de inteligência, bem como os modelos propostos para o seu estudo.

2.1 O conceito de inteligência

Boring (1923) propôs que a inteligência é o que os testes de inteligência aferem. Apesar da operacionalidade dessa definição, o autor não esperava que a mesma colocasse um ponto final na discussão sobre o tema. Sternberg (2000) ressalta que nessa época não era claro o que os testes de inteligência mensuravam, portanto, a definição de Boring não deixava claro o que era a inteligência.

Um dos estudos mais famosos sobre a definição de inteligência foi feito pelos editores do *Journal of Educational Psychology* (“Intelligence and its measurement,” 1921) em um simpósio com especialistas, o qual obteve definições como as seguintes (Sternberg, 2000, p. 8):

O poder de dar boas respostas sob o ponto de vista da verdade ou dos fatos (E. L. Thorndike);

Habilidade para raciocinar abstratamente (L. M. Terman);

Capacidade sensorial, capacidade para reconhecimento perceptual, extensão ou flexibilidade de associação, facilidade e imaginação, atenção, rapidez ou prontidão para responder (F. N. Freeman);

Habilidade para aprender ou aprender a se ajustar ao meio (S. S. Colvin);

Habilidade para se adaptar adequadamente a situações da vida relativamente novas (R. Pintner);

A capacidade para saber e ser dotado de conhecimento (B. A. C. Henmon);

Um mecanismo biológico por meio do qual os efeitos da complexidade dos estímulos são agregados para produzir um efeito unificado no comportamento (J. Peterson);

A capacidade para inibir um ato instintivo, a capacidade para redefinir essa inibição à luz de experiências imaginadas, de tentativas e erros, e a capacidade de perceber a modificação produzida no comportamento social como vantagem para o indivíduo como um animal social (L. L. Thurstone);

A capacidade para adquirir capacidade (H. Woodrow);

A capacidade para aprender ou para se beneficiar das experiências (W. F. Dearborn);
e

Sensação, percepção, associação, memória, imaginação, discriminação, julgamento e raciocínio (N. E. Haggerty).

De acordo com Sternberg (2000), existem algumas características em comum nas definições apresentadas nesse simpósio. Alguns pontos em comum são a habilidade para se adaptar ao meio e a habilidade para aprender. Em outro simpósio realizado em 1986, 65 anos mais tarde, os especialistas convidados buscaram atualizar as definições do primeiro. Entre os participantes estavam: Anne Anastasi, Paul Bates, Jonathan Baron, John Berry, Ann Brown,

Earl Butterfield, John B. Carroll, Howard Gardner, Robert Glaser, Jacqueline Goodnow, John Horn, Lloyd Humphreys, Arthur Jensen e Robert Sternberg. Alguns temas proeminentes em ambos os simpósios foram: atributos como adaptação ao meio, processos mentais básicos e pensamento de ordem superior (por exemplo: raciocínio, resolução de problemas e tomada de decisão) e a quantidade de inteligências, se uma ou muitas, embora não se tenha chegado a qualquer consenso. Assim como no primeiro simpósio, no segundo alguns especialistas definiram a inteligência estritamente em termos de elementos biológicos ou cognitivos, enquanto outros incluíram muitos outros elementos, como motivação e personalidade.

Da mesma forma que nos dias atuais, também nesses dois simpósios não houve um consenso geral entre os especialistas acerca da natureza da inteligência. E, embora seja um tema corrente no campo da psicologia e muito estudado por especialistas que adotam diferentes vertentes de estudo, ainda há muita divergência sobre o que é e como pode ser mensurada a inteligência humana.

Neisser (1979) propôs que a inteligência pode ser prototipicamente organizada, ou seja, para esse autor não existem atributos claros que possam defini-la, mas apenas atributos característicos que tendem a ser típicos em pessoas inteligentes. Segundo essa visão, uma pessoa inteligente é alguém que possui certos atributos, mas não há atributos particulares identificados como necessários e suficientes para descrever uma pessoa inteligente.

Mesmo Binet não conseguiu definir claramente a inteligência. No final de sua vida, ele notou que a inteligência está contida no significado de quatro termos: compreensão, inventividade, direção e crítica. Ele argumentou que a inteligência é um processo direcionado à compreensão de um mundo externo em direção à reconstrução. E a reconstrução deve estar sujeita a uma avaliação crítica por parte do sujeito (Brody, 2000).

Segundo Sternberg (1990), um bom método para entender a inteligência é por meio de modelos teóricos. E esses modelos frequentemente utilizam metáforas para explicar as

pesquisas existentes em inteligência bem como para clarear questões futuras. Na próxima seção serão tratadas algumas dessas metáforas.

2.2 Metáforas sobre a inteligência

No livro editado por R. J. Sternberg no ano de 2000, *Handbook of intelligence*, o capítulo escrito por ele: “The concept of intelligence”, apresenta várias metáforas sobre a inteligência, relatadas a seguir:

I. Metáfora geográfica

A inteligência é vista como um mapa da mente. Alguns teóricos dessa visão são: Spearman (1927), Thurstone (1938), Guilford (1985), Cattell (1967), Vernon (1971) e Carroll (1993). A unidade básica de análise nessa metáfora é o fator, tipicamente eleito como a fonte das diferenças individuais. Algumas vantagens dessa visão: (a) especificação clara das estruturas mentais propostas; (b) operacionalização direta por meio de testes mentais; e (c) disponibilidade de maquinário sofisticado para a implementação.

II. Metáfora computacional

A unidade básica de análise é o processo elementar de informação. Alguns dos teóricos adeptos dessa visão são: Simon (1976), Hunt (1978) e Sternberg (1977). Os expoentes dessa visão geralmente utilizam em seus estudos análise do tempo de reação, análise de protocolos e simulação no computador. Entre as vantagens dessa visão estão: (a) sua especificação detalhada dos processos e estratégias mentais; (b) análise do tempo real em tarefas de execução; e (c) disponibilidade de maquinário sofisticado para a implementação.

III. Metáfora biológica

Entre os principais teóricos dessa perspectiva estão: Luria (1973), Hebb (1942), Halstead (1951) e Vernon (1971). A unidade principal de análise varia entre as teorias. Para Hebb, era a conjunção celular; para Vernon, é a velocidade da condução neuronal. Outros, como Luria e Halstead, propuseram teorias estruturais ligando partes do cérebro a várias funções intelectuais. Os métodos de análise utilizados incluem mensuração de potenciais evocados, de velocidade da condução neuronal, avaliação da especialização hemisférica e escaneamento de partes do cérebro envolvidas em diferentes tipos de tarefas mentais. Algumas vantagens dessa metáfora: (a) relaciona a inteligência com sua fonte no cérebro; e (b) geralmente utiliza técnicas de mensuração e experimentos precisos.

IV. Metáfora dos sistemas

Baseia-se na noção de que a inteligência é um sistema complexo que integra muitos níveis de análise, incluindo o geográfico, o computacional, o biológico, o antropológico e o sociológico. O sistema e seus elementos em interação são a unidade de análise. Entre os principais teóricos estão Gardner (1993) e Sternberg (1997). Algumas vantagens dessa metáfora são: (a) o reconhecimento da complexidade da inteligência; (b) a integração de múltiplos níveis de análise; e (c) a extensão de habilidades incluídas nas teorias.

V. Metáfora sociológica

Essa metáfora enfatiza a importância da socialização para a inteligência. Entre os teóricos mais conhecidos são incluídos Vygotsky bem como Feuerstein (1980). Um construto bastante relevante para Vygotsky é a internalização, por meio da qual a criança orienta o seu comportamento nas interações sociais e então internaliza aspectos importantes da situação e os torna seus. Para Feuerstein um construto chave é a aprendizagem mediada, ou seja, a

aquisição de conhecimento que ocorre quando um mediador, em geral um parente ou professor, explica o ambiente para a criança. Entre as vantagens dessa metáfora estão: (a) o reconhecimento da importância da internalização de experiências inicialmente obtidas com outros; (b) o reconhecimento do papel do mediador na internalização; e (c) o reconhecimento da diferença entre capacidade latente e habilidade desenvolvida.

VI. Metáfora genética-epistemológica

O principal teórico é Piaget (1972) e a unidade fundamental de análise é o esquema. Para Piaget o esquema está presente durante todo o desenvolvimento. Eles são estruturas que estão presentes desde o nascimento e vão se expandindo tanto quantitativamente, o que chamou de assimilação, quanto qualitativamente – acomodação. Esses esquemas são construtos hipotéticos, estruturas intelectuais que organizam os eventos como eles são percebidos pelo organismo e classificados em grupos, de acordo com características comuns. Assim sendo, os esquemas do adulto emergem dos esquemas da criança, os quais no momento do nascimento são de natureza reflexa (Wadsworth, 1997).

O método típico de pesquisa sob essa visão é a observação via estudos de caso e experimentação. Algumas vantagens dessa metáfora são: (a) a compreensão que se tem dela como uma teoria de inteligência e desenvolvimento intelectual; (b) a quantidade expressiva de pesquisas realizadas sob o enfoque dessa metáfora com crianças de todas as idades ao redor do mundo; e (c) os detalhes com os quais muitas estruturas e processos são descritos. Entre as desvantagens estão: (a) a concentração da teoria em aspectos lógicos e científicos da inteligência; (b) o estabelecimento do começo do último estágio de desenvolvimento intelectual aproximadamente aos 11 ou 12 anos de idade, sendo quase certo que os indivíduos alcancem maturidade intelectual antes dessa idade; e (c) a superestimação das idades nas quais as crianças são capazes de apresentar vários avanços intelectuais. Tendo como base a

metáfora geográfica, adotada no presente estudo, na seção a seguir será delineado um histórico de alguns dos modelos mais importantes na área de inteligência.

2.3 Histórico de alguns modelos de inteligência

Francis Galton, seis anos depois da publicação da célebre obra de Darwin ‘A origem das espécies’, publicou dois artigos, ambos intitulados ‘Características e talentos hereditários’. Segundo esse autor a hereditariedade explicava as características cognitivas, ele também acreditava em hierarquias raciais. Em 1882 Francis Galton montou um laboratório no Sul de Kensington, no museu de Londres com o objetivo de mensurar diferenças individuais. Ele aferia, por exemplo, habilidades de discriminação sensorial auditiva e visual, bem como tempo de reação a estímulos (Brody, 2000).

Galton (1869) verificou a relação entre a inteligência de parentes e constatou uma relação significativa apenas entre parentes de primeiro grau (pais e filhos). Esse pesquisador introduziu o estudo com gêmeos e sugere que gêmeos que são similares desde o nascimento continuam similares e gêmeos que nascem dissimilares continuam dissimilares, mesmo quando expostos a ambientes similares. Ele argumentava que esses dados evidenciavam a dominância da natureza sobre o meio.

Binet se opôs às idéias de Galton, ele não concordava com a mensuração da inteligência focada em processos cognitivos elementares. Ele estava convencido de que a inteligência deveria ser estudada focando-se em processos mentais complexos. Binet e Henri (1896 citado em Brody, 2000) expressaram suas dúvidas sobre a pesquisa em inteligência direcionada ao estudo de diferenças individuais em funções simples. Eles ficaram espantados com o espaço dado às sensações e processos simples e com a pouca atenção dada aos processos superiores.

De acordo com Brody (2000), o debate sobre a relevância de tarefas simples ou complexas para a mensuração da inteligência é apresentada em dois dos mais importantes

artigos na história da pesquisa em inteligência: um de Spearman publicado em 1904 e um de Binet e Simon publicado um ano depois. Spearman desenvolveu uma teoria, Binet e Simon desenvolveram um teste.

Spearman (1904) fez estudos de correlação entre várias medidas de discriminação sensorial e encontrou correlações positivas; também correlacionou estas com desempenho acadêmico e também encontrou índices positivos. Ele chegou à conclusão de que “o elemento comum e essencial nas inteligências coincide com o elemento comum e essencial nas funções sensoriais” (Spearman, 1904, p. 269). Assim, Spearman assumiu que deve existir uma habilidade intelectual comum, responsável pelas principais correlações positivas encontradas. Ele chamou de inteligência geral, *g*, essa habilidade. Ele assumiu que a variância em uma medida pode ser explicada por uma teoria de dois fatores: um componente atribuído a *g* e uma segunda fonte específica de variância, *s*.

Segundo Brody (2000, p. 19), “Binet e Spearman discordaram substancialmente” no âmbito do estudo da inteligência. Binet escreveu que além de ficar espantado com a visão de Spearman de que a inteligência estaria relacionada com a habilidade de discriminação sensorial, também assinalou que a noção de *g* como uma entidade singular era enganosa, pois dois indivíduos poderiam obter o mesmo escore usando conhecimentos e habilidades muito diferentes.

A visão de que o fator *g* não era suficiente para explicar a inteligência e a ideia de que poderia haver fatores específicos, além do *g*, levou Thurstone a elaborar sua teoria (Thurstone, 1931, 1938 citado em Brody, 2000; Thurstone e Thurstone, 1941). Ele desenvolveu um método de análise fatorial múltipla para verificar fatores independentes. A análise de Thurstone (1938) revelou sete fatores independentes, ou habilidades mentais primárias, que poderiam obter interpretação psicológica, a saber: (1) compreensão verbal - saber as definições de palavras, (2) fluência verbal - produzir rapidamente um grande número

de palavras, (3) facilidade numérica - resolução de problemas matemáticos, (4) raciocínio espacial - rotacionar mentalmente figuras ou objetos, (5) velocidade perceptual - reconhecer rapidamente características de estímulos, (6) indução - resolver analogias e encontrar regras ou princípios para completar problemas e (7) memória - lembrar listas de palavras, figuras ou números (Davidson & Downing, 2000).

Não obstante, Thurstone notou que não era possível obter uma estrutura simples de fatores independentes; ele chegou à conclusão de que os fatores tinham que ser oblíquos, então, era possível realizar análises com fator de segunda ordem, o que poderia levar ao *g*.

Guilford desenvolveu um modelo de inteligência que não era hierárquico (ver Guilford, 1967, 1981, 1985). Ele desenvolveu uma taxonomia tri-dimensional que lhe permitia classificar qualquer teste quanto à sua posição nas dimensões de operação, produção e conteúdo. Ele distinguiu cinco operações: cognição (conhecimento), memória, produção divergente (geração de alternativas), produção convergente e avaliação.

Dois modelos hierárquicos contemporâneos ajudam a conciliar a visão de Spearman e Thurstone. A Teoria de Inteligência Fluida e Cristalizada (Teoria *Gf-Gc*) clarifica a noção de Spearman de um fator geral e é útil para explicar o desenvolvimento da infância até a idade adulta (Horn, 1994). Por outro lado, a Teoria dos Três Estratos de Carroll (1993) é baseada na meta-análise de uma grande quantidade de resultados psicométricos, incluindo os dados que formaram a Teoria *Gf-Gc* (Davidson & Downing, 2000). As seções a seguir tratarão dessas duas teorias de forma mais detalhada.

2.4 O modelo hierárquico de habilidades cognitivas de Cattell e Horn

Raymond B. Cattell foi estudante e pesquisador associado de Spearman nos anos de 1930. Ele se mudou para os Estados Unidos em 1937 e por muitos anos dedicou-se quase que exclusivamente à pesquisa e ao estudo de metodologias de análise de fatores, especialmente no campo da personalidade. Não obstante, ele fez muitas incursões na área de inteligência.

Em um artigo publicado em 1943 sobre inteligência de adultos (Cattell, 1943), ele propôs a possível existência de dois tipos de inteligência: uma “fluida”, que refletiria habilidades básicas de raciocínio e estaria relacionada a processos mentais mais superiores; e uma inteligência “cristalizada”, refletindo o quanto o indivíduo é capaz, parcialmente com base no nível de “inteligência fluida”, de aprender e aproveitar de suas experiências culturais, educacionais etc. A Inteligência Fluida foi assim nomeada porque foi concebida como sendo capaz de permear muitos tipos de atividades mentais; a cristalizada foi assim chamada porque foi pensada como um tipo de produto final de experiências até um determinado ponto na vida de um indivíduo (Carroll, 1993).

Cattell notou o paralelismo dessa distinção com a distinção de Hebb (1942) entre “Inteligência A” e “Inteligência B”, sendo a primeira a capacidade determinada biologicamente e a segunda, a inteligência gerada por meio da experiência e da educação. Assim, a teoria das inteligências fluida e cristalizada, agora geralmente chamada Teoria Gf-Gc, passou por muitas checagens e testes experimentais (Cattell, 1963, 1967). Um estudante de Cattell, John Horn, forneceu o primeiro teste claro da teoria em sua tese de doutorado, usando as técnicas fatoriais de ordem superior thurstonianas (Carroll, 1993).

Estudos de Horn, Cattell e outros têm refinado a Teoria Gf-Gc; por exemplo, um grande estudo de vinte fatores primários de Hakstian e Cattell (1978) revelou seis fatores de segunda ordem, incluindo não somente Gf e Gc, mas também Gv (Capacidade de Visualização), Gps (Velocidade Geral Perceptual), Gm (Capacidade Geral de Memória) e Gr (Capacidade Geral de Recuperar). Ademais, foram feitas correções entre esses fatores de segunda ordem, sugerindo a existência de três fatores de terceira ordem, “inteligência fluida original”, “capacidade de concentração” e “cultura escolar”. Todavia, esses achados devem ser considerados com cautela, sendo necessários estudos futuros (Carroll, 1993).

Uma característica interessante da Teoria Gf-Gc, desenvolvida por Horn (1985; Horn, Donaldson & Engstrom, 1981) é que as habilidades da inteligência fluida tendem a declinar com o avanço da idade, pelo menos em alguns indivíduos, enquanto as habilidades cristalizadas, como vocabulário, não (Carroll, 1993). A esse respeito, Bugg, Zook, DeLosh, Davalos e Davis (2006) encontraram, em um estudo realizado com pessoas entre 20 e 89 anos de idade, um declínio da inteligência fluida com o avanço da idade. Esses autores concluíram que esse fator era devido ao processo de envelhecimento geral e ao declínio das funções frontais do cérebro, bem como a outros fatores não identificados.

Em seu maior livro sobre habilidades, Cattell (1971 citado em Carroll, 1993) incorporou a Teoria Gf-Gc em uma teoria nova, mais geral – a “teoria triádica”, que propõe que as habilidades cognitivas encaixam-se em três tipos, a saber: “capacidades”, que são habilidades que refletem “limites para a ação do cérebro como um todo”; “capacidades internas”, que são tipos de ‘organização local’ para diferentes modalidades motoras e sensoriais; e “ações”, que são habilidades para realizar coisas em diferentes áreas de conteúdo cultural, adquiridas por meio de investimento da inteligência fluida no aprendizado.

Ademais, Cattell postulou um número de dimensões por meio das quais os fatores de habilidade podem ser caracterizados. Duas dessas dimensões são relacionadas com “fases de ação” – entrada, processo e saída, e duas delas referem-se a conteúdo – envolvimento de dimensões culturais e de experiência, e envolvimento de dimensões organizacional-neural. O restante são “parâmetros de processo”, tais como nível de complexidade, grau de exigência da memória e velocidade demandada (Carroll, 1993).

Segundo Carroll (1993), o modelo de Horn-Cattell é um verdadeiro modelo hierárquico que cobre todos os maiores domínios do funcionamento intelectual. Numerosos detalhes necessitam ser melhor estudados, mas entre os modelos disponíveis esse parece

oferecer a abordagem mais razoável e bem estabelecida para uma teoria aceitável acerca da estrutura das habilidades cognitivas.

2.5 A teoria dos três estratos

A teoria dos três estratos (Carroll, 1993) está baseada na análise de mais de 460 estudos. Esse modelo é similar a outros modelos hierárquicos, mas difere da Teoria *Gf-Gc*, por exemplo, na inclusão do fator geral de terceira ordem no ápice da hierarquia (Davidson & Downing, 2000). Carroll (1993) descreve a estrutura da inteligência como uma pirâmide. No topo da pirâmide está o Estrato III, equivalente conceitualmente ao fator *g* de Spearman. Embora Carroll não apoie a interpretação de Spearman do *g* como representando energia mental, ele concorda que o *g* está por trás de todas as atividades intelectuais e tem um alto grau de hereditariedade.

No meio da pirâmide, Estrato II, estão oito fatores que são influenciados de diferentes formas pelo *g*. Essas habilidades gerais, que são similares aos fatores de segunda ordem na Teoria *Gf-Gc*, representam a permanência de características individuais que podem influenciar a performance em determinado domínio. Essas oito habilidades são: inteligência fluida, inteligência cristalizada, memória e aprendizagem geral, percepção visual ampla, vasta percepção auditiva, ampla habilidade de recuperação, ampla velocidade cognitiva e velocidade de processamento (Davidson & Downing, 2000).

A base da pirâmide, Estrato I, consiste de numerosas habilidades específicas, tais como raciocínio quantitativo e conhecimento léxico. De acordo com Davidson e Downing (2000), assim como no modelo *gf-gc*, alguns desses fatores representam as habilidades mentais primárias de Thurstone. Cada habilidade no Estrato I está relacionada com uma ou mais das oito habilidades que compõem o Estrato II.

Não obstante a estrutura proposta por Carroll (1993), ele enfatiza que seus três estratos não são definidos de forma rígida. As habilidades em cada estrato refletem apenas seu

grau de generalidade no comando de grande parte do comportamento inteligente. Davidson e Downing (2000) assinalam que o Estrato I representa habilidades muito especializadas, que refletem a aquisição de estratégias particulares ou tipos específicos de conhecimento; o Estrato II, por sua vez, contém habilidades não muito especializadas, presentes em muitos domínios do comportamento inteligente; enquanto o Estrato III tem uma habilidade, *g*, que está por trás de todos os aspectos da atividade intelectual. Devido ao fato de a generalidade ser uma questão de grau, algumas habilidades podem ser difíceis de alocar em um estrato. Portanto, Carroll admite que podem existir estratos intermediários entre os três estratos que ele identificou.

Muitos estudos ainda precisam ser feitos nessa área para que se possa compreender melhor os pontos menos estabelecidos desse modelo. No Brasil, alguns estudiosos do campo da inteligência têm construído instrumentos e realizado pesquisas com o intuito de compreender melhor e aferir a inteligência fluida (Pasquali, 2005; Primi, 2001). O presente trabalho também se configura como uma contribuição para esse campo de estudo, especialmente no âmbito da psicologia infantil.

3. Histórico dos testes SON de 1943-2009

No início dos anos 40 Snijders-Oomen (1943) construiu uma escala de inteligência não-verbal (SON) objetivando avaliar crianças surdas. A inteligência foi operacionalizada por essa autora em termos da habilidade para aprender; do quanto as crianças podem aproveitar da instrução adquirida na escola. O teste SON desenvolvido por Snijders-Oomen foi o primeiro teste que contemplou uma grande área da inteligência sem ser dependente do uso da linguagem.

O diagrama a seguir sumariza o histórico dos testes SON, desde sua primeira versão até o estudo atual de normatização no Brasil.

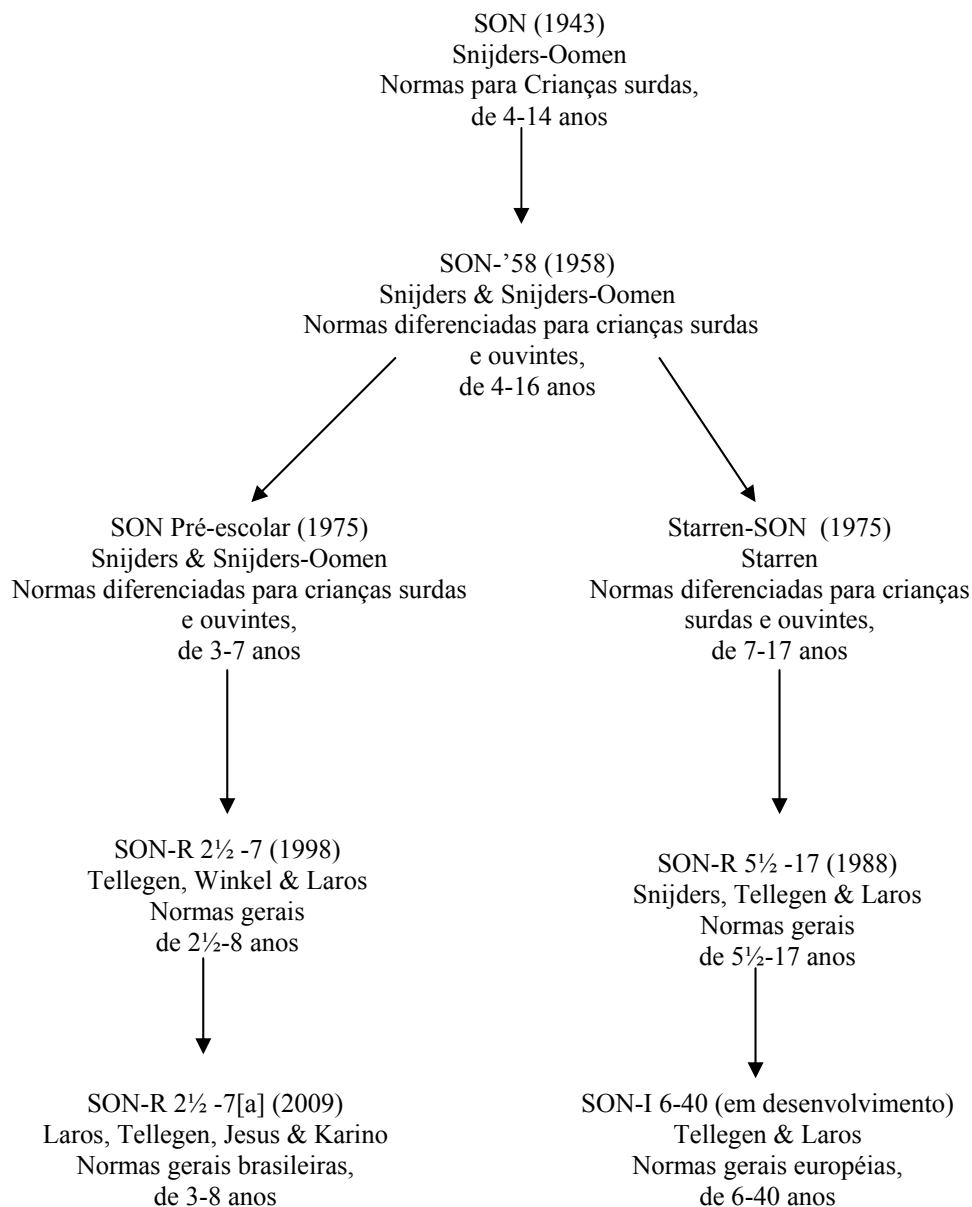


Figura 3.1. Histórico dos testes SON de 1943 a 2009, com base em Drunen (2003)

Conforme observado nesse diagrama, a primeira versão do teste SON foi desenvolvida para a avaliação do funcionamento cognitivo de crianças surdas com idade entre 4 e 14 anos (Snijders-Oomen, 1943). Fazendo uso de tarefas novas e tarefas recém desenvolvidas na época, Snijders-Oomen desenvolveu uma bateria de testes que incluiu diversas tarefas não-verbais relacionadas à habilidade espacial e aos raciocínios abstrato e concreto. O objetivo do

teste era fornecer uma indicação clara da habilidade de aprendizagem e das chances de sucesso na escola. Um requisito para a bateria de testes foi que as vivências familiares e a educação formal deveriam influenciar minimamente os resultados do teste.

Para essa mesma autora “forma”, “coerência concreta”, “abstração” e “memória de curto-prazo” foram vistos como sendo representações aceitáveis de funcionamento intelectual típico para sujeitos que estavam sofrendo de surdez precoce. O objetivo dessa bateria de testes era quebrar o viés dos testes de execução não-verbal empregados naquela época e permitir que funções como abstração, simbolismo, entendimento de situações comportamentais e memória ficassem mais acessíveis para a avaliação não-verbal (Tellegen, Winkel, Wijnberg-Williams & Laros, 1998).

A primeira revisão do teste foi publicada em 1958, o SON-'58 (Snijders & Snijders-Oomen, 1958). Nesta a primeira versão foi expandida e normatizada tanto para crianças com audição como para crianças surdas com idade entre 4 e 16 anos.

Na segunda revisão foram desenvolvidas duas baterias distintas de testes, pois em todos os subtestes do SON original um tipo de item diferente parecia mais apropriado para crianças acima de seis anos. Assim, o SSON (Starren, 1975) foi desenvolvido para crianças de 7 a 17 anos e o SON 2½-7, geralmente chamado de SON Pré-escolar, ou P-SON (Snijders & Snijders-Oomen, 1976), para crianças de 2½ a 7 anos.

Em 1988 foi publicado o SON-R 5½-17 como resultado da terceira revisão do teste (Laros & Tellegen, 1991; Snijders, Tellegen & Laros, 1989; Tellegen & Laros, 1993). Ele foi desenvolvido para uso tanto com crianças ouvintes quanto com crianças surdas, com idade entre 5½ e 17 anos. Foi feito um esforço na construção desse teste para combinar as vantagens do SSON e do SON-'58.

A COTAN (*Commissie Testaangelegenheden van het nederlandse Instituut van Psychologen*), Comissão do Instituto Nacional dos Psicólogos da Holanda, responsável pela

avaliação de testes, analisou o SON-R 5½-17. Todos os aspectos do teste avaliados (base teórica para a construção, execução do manual e dos materiais, normas, fidedignidade e validade) foram julgados como ‘bom’, que é o maior conceito em uma escala de quatro níveis (Evers, Van Vliet-Mulder & Ter Laak, 1992). Isto quer dizer que o SON-R 5½-17 é considerado um dos testes de melhor credibilidade na Holanda (Sijtsma, 1993).

Finalizada a revisão do SON-R 5½-17, foi iniciada a revisão do SON Pré-escolar, o que resultou na publicação do SON-R 2½-7 (1998). Este é o teste utilizado na presente pesquisa. Dessa forma, a próxima seção tratará de forma mais pormenorizada do mesmo.

3.1 O teste não-verbal de inteligência SON-R 2½-7

As principais razões para a revisão do SON Pré-escolar (1975) foram: (1) a necessidade de atualizar as normas, modernizar os materiais do teste e (2) proporcionar uma boa comparação com os itens destinados às crianças mais novas do SON-R 5½ -17 (Tellegen e cols., 1998).

Segundo Tellegen e cols. (1998), como o SON Pré-escolar foi publicado em 1975 era necessário uma revisão, pois Cronbach (1996) aconselha revisar testes de inteligência depois de um período de mais de 20 anos. As normas têm a tendência de ficarem obsoletas ao longo do tempo. Ademais, alguns pesquisadores fizeram estudos nos quais observaram que o desempenho em testes de inteligência em 10 anos aumenta cerca de 2,5 pontos na escala de QI (100, 15) (Lynn & Hampson, 1986; Flynn, 1987). Em geral, por questões práticas, como financiamento e tempo, não se revisa as normas dos testes a cada 10 anos. E quanto aos testes SON, o programa de correção do teste já calcula esse efeito do tempo sobre o QI. É oferecida uma estimativa do QI chamada QI Flynn.

De acordo com Tellegen e cols. (1998) outra razão para a revisão do SON Pré-escolar foi devido aos materiais, que já estavam antiquados. O terceiro motivo para a revisão foi buscar melhorar a fidedignidade e a generalizabilidade do teste, em especial para os limites

superiores e inferiores da faixa etária. No manual do SON Pré-escolar são apresentados dados que mostram que os subtestes diferenciam muito pouco os grupos limítrofes de idade. Além disso, no SON Pré-escolar não havia normas para os subtestes individuais, apenas para o escore total. Assim, não era possível calcular o QI se um subteste deixasse de ser administrado. Por fim, as normas eram apresentadas por grupo de idade de seis meses, o que poderia levar a um desvio de seis pontos de QI se a idade não correspondesse ao meio do intervalo.

A última razão para a revisão do SON Pré-escolar foi possibilitar a comparação dos resultados do SON-R 2½ -7 com os do SON-R 5½-17. Assim, o teste para crianças mais novas deveria ser muito similar ao teste para crianças mais velhas. Uma sobreposição nas faixas de idade dos testes também era considerada desejável (Tellegen & cols., 1998). A revisão não se restringiu à construção de normas novas, mas os itens, os subtestes e as instruções foram também submetidos a uma ampla revisão, em todos os aspectos.

Tellegen e cols. (1998) referem que o SON-R 2½-7 foi construído nos anos de 1991 e 1992. Nesse período foram elaboradas três versões experimentais do teste, administradas em mais de 1.850 crianças com idades entre 2½ e 7 anos. A versão final foi construída com base nos dados dessa pesquisa, nas experiências e observações dos examinadores, comentários e sugestões de psicólogos e educadores.

A pesquisa de normatização do SON-R 2½-7 foi realizada entre os anos de 1993 e 1994 com mais de 1.000 crianças, com idade na faixa etária do teste. Para a validação do teste foram aplicados outros testes de inteligência e linguagem em grande parte das crianças que participaram da pesquisa de normatização (Tellegen & cols., 1998). A seguir serão relatados alguns estudos sobre a validação do SON-R 2½-7.

3.1.1 Estudos de validação do SON-R 2½-7

A validade do SON-R 2½-7 foi verificada em vários estudos de validação na Holanda e outros países como Inglaterra, Estados Unidos e Austrália. Nesses estudos foram comparados os resultados no SON-R 2½-7 com os resultados em outros testes de inteligência e cognitivos.

Os resultados do SON-R 2½-7 foram comparados na Holanda com os dos seguintes testes: WISC-R, WPPSI-R, TONI-2, Stutsman, Kaufman-ABC, BOS 2-30, LDT, RAKIT, TOMAL, DTVP-2 e os testes de desenvolvimento da linguagem Reynell e Schlichting. O tamanho da amostra dos vários estudos de validação variou de 26 a 558 sujeitos. As 21 correlações do SON-R 2½-7 com outros testes não-verbais de inteligência variaram de 0,45 a 0,83, com média de 0,65. As 12 correlações com medidas gerais de inteligência variaram de 0,54 a 0,87, com média de 0,65. As 19 correlações do SON-R 2½-7 com medidas de habilidade verbal e inteligência verbal variaram de 0,20 a 0,71, com média de 0,48 (Tellegen & cols., 1998).

O estudo de validação do SON-R 2½-7 na Austrália foi realizado em 1996. A amostra do estudo tinha 155 sujeitos, 72 meninos e 83 meninas, com idade média de 4 anos e 5 meses (DP=10 meses). Nessa amostra havia três grupos diferenciados de crianças: (1) sem nenhum tipo de deficiência (N = 59), (2) com problemas auditivos (N = 59) e (3) com retardo no desenvolvimento (N = 37). A correlação do SON-R 2½-7 com o WPPSI-R na amostra total da Austrália foi 0,78. Considerando os três grupos de crianças citados, a correlação entre o QI da escala de execução do WPPSI-R e o SON-R 2½-7 foi 0,74, 0,74 e 0,75, respectivamente. No grupo de crianças sem qualquer tipo de deficiência a correlação do SON-R 2½-7 com a escala verbal ($r=0,54$) foi menor do que a correlação com a escala de execução ($r=0,74$) do WPPSI-R (Tellegen e cols., 1998).

O estudo de validação realizado nos Estados Unidos (Oeste da Virgínia) foi efetuado por meio da aplicação do SON-R 2½-7 e cinco outros testes cognitivos: WPPSI-R, Kaufman-ABC, Escalas de Habilidades Infantis de McCarthy (MSCA), Teste de Figura e Vocabulário Peabody – Revisado (PPVT-R) e a Escala Pré-escolar de Linguagem – 3 (PLS-3). A correlação entre o SON-R 2½-7 e o escore total do WPPSI-R foi 0,59; as correlações com as escalas de execução e verbal foram 0,60 e 0,43, respectivamente. A idade média das 75 crianças que fizeram ambos os testes foi 5,1 anos (Tellegen & cols., 1998).

A correlação do SON-R 2½-7 com o Kaufman-ABC foi 0,66; com a escala não-verbal do Kaufman-ABC foi 0,61 e a idade média das 31 crianças que fizeram tanto o Kaufman-ABC quanto o SON-R 2½-7 foi 4,6 anos. Com o índice cognitivo geral do MSCA o SON-R 2½-7 mostrou uma correlação de 0,61; com a escala verbal foi 0,48 e com a escala de execução perceptual foi 0,61 (Tellegen & cols., 1998).

Por fim, a correlação do SON-R 2½-7 com o PPVT-R foi 0,47, a idade média das 29 crianças que fizeram ambos os testes foi 5,5 anos. Com o escore total de linguagem do PLS-3 a correlação foi 0,61, com a Escala de Compreensão Auditiva foi 0,59 e com a Escala de Comunicação Expressiva foi 0,56. A idade média das 47 crianças que fizeram ambos os testes foi 4,6 anos (Tellegen & cols., 1998).

De acordo com Tellegen e cols. (1998) a pesquisa de validação na Inglaterra foi realizada em 1996. Além do SON-R 2½-7 foram aplicadas as Escalas Britânicas de Habilidades (BAS). Ambos os testes foram aplicados em 58 crianças, 34 meninos e 24 meninas. A idade média das crianças foi 6,3 anos (DP=3 meses). O intervalo entre a aplicação dos testes variou de alguns dias a algumas semanas. Nessa amostra há três grupos diferenciados de crianças: um grupo sem deficiências específicas (N = 20), um grupo que tem o inglês como segunda língua (N = 22) e um grupo com problemas de aprendizagem (N = 16).

A correlação do SON-R 2½-7 com a versão reduzida da BAS foi 0,80. Quando os dois subtestes não-verbais do BAS foram incluídos a correlação aumentou para 0,87. A correlação com a parte verbal (três subtestes verbais) da versão reduzida do BAS foi 0,71, enquanto a correlação com a parte de execução (três subtestes não-verbais) foi 0,78. As correlações no grupo de crianças sem deficiências foram consideravelmente menores do que nos outros dois grupos (0,56 vs. 0,76 e 0,78).

No Brasil o SON-R 2½-7 foi administrado por Quesada (2007) em um estudo com 30 crianças epiléticas, com idade entre 3 anos e 7 meses e 6 anos e 11 meses. Além do SON-R 2½-7 foi aplicado o WPPSI-III e a correlação entre os dois testes mostrou altos índices entre todos os tipos de QI avaliados ($r > 0,50$). O QI total, de raciocínio e de execução do SON-R 2½-7 obteve correlações variando de 0,55 a 0,82, com o QI verbal, de execução, total, velocidade de processamento, organização perceptual e resistência à distração do WPPSI-III.

3.2 O teste não-verbal de inteligência SON-R 2½-7[a]

O SON-R 2½-7[a] é uma versão abreviada do SON-R 2½-7, pois dos seis subtestes que compõem este último apenas quatro fazem parte do SON-R 2½-7[a], a saber: Mosaicos, Categorias, Situações e Padrões, administrados nessa ordem. O SON-R 2½-7[a], objeto da presente pesquisa, é um teste de inteligência geral para crianças com idade de até 7 anos e nove meses. Ele avalia um espectro largo de habilidades cognitivas e as instruções podem ser dadas tanto de maneira verbal como não-verbal, dependendo das possibilidades de comunicação da criança. Os subtestes do SON-R 2½-7[a] podem ser agrupados em dois tipos de teste: de raciocínio (Categorias e Situações) e de execução com enfoque espacial (Mosaicos e Padrões).

As habilidades de raciocínio têm sido consideradas tradicionalmente como a base da função de inteligência (Carroll, 1993). Testes de raciocínio são centrais na maioria dos testes de inteligência e podem ser divididos em testes de raciocínio abstrato e concreto. Testes de

raciocínio abstrato, como Categorias, são embasados no relacionamento entre conceitos que são abstratos, isto é, os que não são determinados pelo tempo ou espaço. Nos testes de raciocínio abstrato, um princípio de ordem precisa ser derivado do material que foi apresentado como estímulo, que precisa ser aplicado ao material novo. Nos testes de raciocínio concreto, como no de Situações, o objetivo é criar uma conexão real de tempo-espaço entre pessoas e objetos (Laros & Tellegen, 1991).

Os testes espaciais são similares aos testes de raciocínio concreto porque, em ambos os casos, um relacionamento dentro de um todo espacial precisa ser construído. A diferença está no fato de que os testes de raciocínio concreto dizem respeito a um relacionamento entre partes de uma figura que faz sentido e os testes espaciais se referem a um relacionamento de “forma” entre partes de uma figura (ver Carroll, 1993). Testes espaciais têm sido utilizados por muito tempo como componentes integrais dos testes de inteligência. Os subtestes espaciais incluídos no SON-R 2½-7[a] são Mosaicos e Padrões.

Uma característica importante que Mosaicos e Padrões têm em comum é que o item é resolvido enquanto os materiais do teste estão sendo manipulados. Por essa razão, esses dois subtestes são chamados testes de execução (ver Tellegen & cols., 1998).

A administração do SON-R 2½-7 é individual e o tempo médio de resposta é de 50 minutos. Um elemento importante das instruções do SON-R 2½-7 é o “mostrar como fazer” por parte do aplicador em parte dos itens. Outro aspecto muito importante da administração é o *feedback* que o examinador oferece depois de cada item. Esse *feedback* não é apenas informar se a resposta da criança foi certa ou errada, mas o examinador ajuda a criança a encontrar a solução correta. Devido a esse aspecto o SON-R 2½-7 possui mais similaridade com testes que aferem potencial para aprender do que com testes tradicionais de inteligência (Tellegen & Laros, 1993).

3.2.1 Descrição dos subtestes do SON-R 2½-7[a]

Mosaicos

O subteste Mosaicos é composto por 15 itens e um exemplo. Na parte I de Mosaicos a criança precisa copiar padrões simples de mosaicos numa moldura utilizando 3 a 5 quadrados vermelhos. O número de quadrados a serem utilizados e a possibilidade de o examinador demonstrar o item primeiro ou não determina a dificuldade dos itens.

Por sua vez, a parte II de Mosaicos requer que padrões diversificados de mosaicos sejam copiados com quadrados vermelhos, amarelos e vermelhos/amarelos em uma moldura. Nos itens mais fáceis da parte II, somente quadrados vermelhos e amarelos são utilizados e o desenho é impresso no tamanho real. Nos itens mais difíceis, todos os tipos de quadrados são utilizados e o desenho é apresentado numa escala menor. A Figura 3.2 apresenta um exemplo desse subteste.

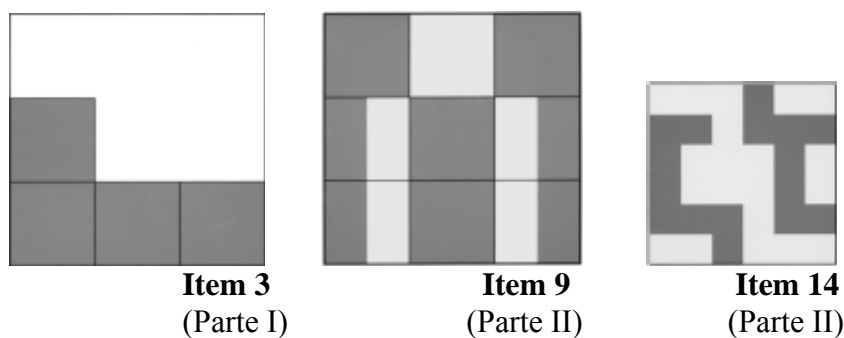


Figura 3.2 Itens do subteste Mosaicos

Categorias

O subteste Categorias possui 15 itens. Na primeira parte 4 ou 6 cartões precisam ser postos na categoria correta. Nos primeiros itens, os desenhos nos cartões que pertencem à mesma categoria possuem uma semelhança clara. Por exemplo, uma fatia de bolo é apresentada em diferentes posições. Nos últimos itens da parte I o conceito adjacente à

categoria precisa ser compreendido para que o item possa ser finalizado com êxito. Por exemplo, veículos motorizados são contrastados com os que não são.

Os itens da parte II do subtteste Categorias são de múltipla escolha. São apresentados três desenhos para a criança que têm algo em comum e dois espaços vazios devem ser preenchidos por desenhos que têm esse mesmo algo em comum, mas estão em outra página que tem cinco desenhos. A criança deve escolher, entre os cinco, os dois objetos adequados à categoria em análise. O nível de dificuldade é determinado pela quantidade de abstração necessária para descobrir a categoria adjacente a cada item. Na figura a seguir é apresentado um exemplo desse subtteste.

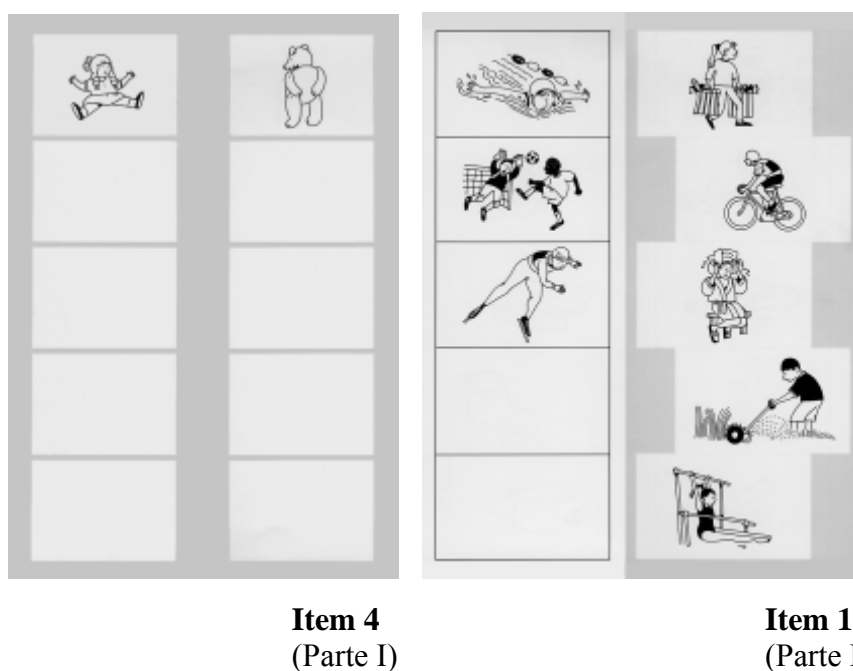
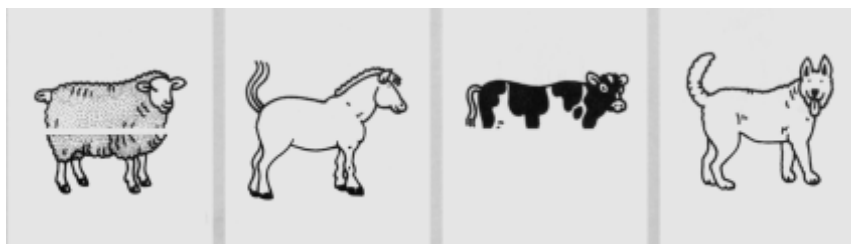


Figura 3.3 Itens do subtteste Categorias

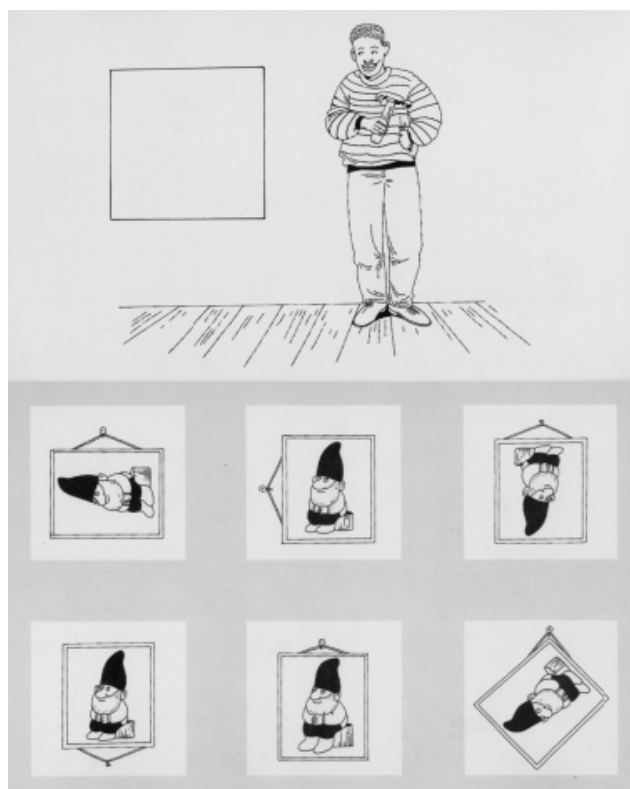
Situações

O subtteste Situações possui 14 itens. Na parte I desse subtteste são apresentadas à criança, no livreto, 4 figuras faltando a metade cada uma. A criança recebe cartões com diferentes metades. Ela precisa colocar a metade certa no desenho correto. O primeiro item é

impresso em cores para deixar o princípio claro. A quantidade de semelhança entre as metades diferentes determina o nível de dificuldade dos itens. A segunda parte desse subtteste é de múltipla escolha. Cada item consiste em uma situação desenhada em que uma ou duas peças estão faltando (ver Figura 3.4). A peça (ou peças) correta precisa ser escolhida entre várias alternativas a fim de deixar a situação consistente. O número de peças faltando determina o nível de dificuldade.



Item 5
(Parte I)



Item 10
(Parte II)

Figura 3.4 Itens do subtteste Situações

Padrões

O subtteste Padrões é formado por 16 itens. Este é o único subtteste em que a segunda parte não é precedida por um exemplo. Solicita-se à criança que copie uma forma que lhe é apresentada utilizando um lápis. Os primeiros itens são desenhados livremente, depois pontos precisam ser conectados a fim de que a forma fique parecida com o exemplo. Os itens em Padrões I são demonstrados pelo examinador primeiro e consistem em não mais do que 5 pontos. A seguir (Figura 3.5) é apresentado um exemplo desse subtteste.

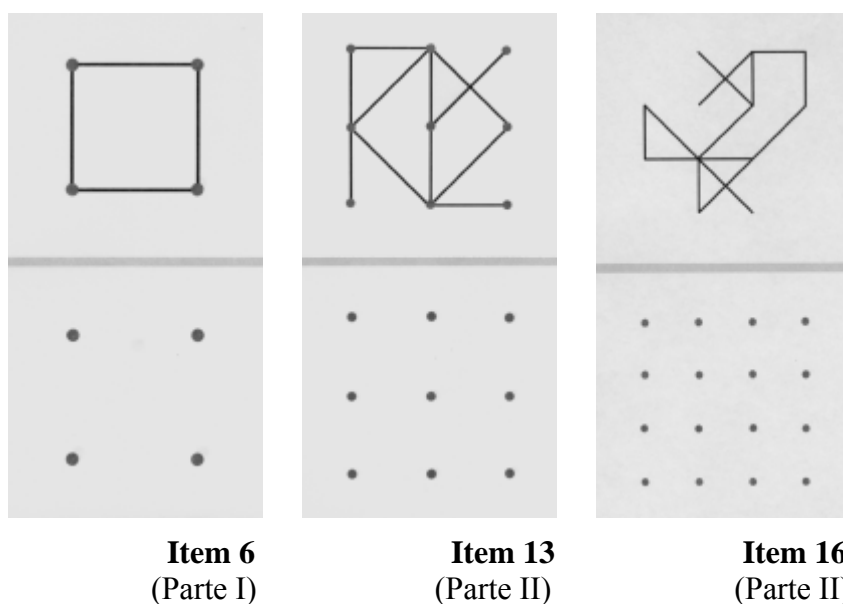


Figura 3.5 Itens do subtteste Padrões

Os itens em Padrões II consistem de 5, 9 ou 16 pontos e precisam ser copiados pela criança sem ajuda. O nível de dificuldade é determinado pelo número de pontos e se os pontos são representados no exemplo ou não.

4. Método

4.1 Amostra

4.1.1 Estudo piloto

Foi realizado um estudo piloto em Brasília com 92 alunos de uma creche e 38 alunos do Centro Educacional da Audição e Linguagem - Ludovico Pavoni (CEAL), totalizando 130 participantes, a maioria do sexo masculino (52,3%). A idade média das crianças da creche foi 4,9 anos (DP = 1,3) e das crianças do CEAL foi 5,3 anos (DP = 1,0). A média geral de idade foi 5 anos (DP = 1,2). Todas as crianças do CEAL apresentavam surdez profunda.

4.1.2 Estudos de validação

Foram realizados estudos de validação do SON-R 2½-7[a] com outros testes de inteligência. A verificação da validade convergente do SON-R 2½-7[a] com o WPPSI-III foi realizada com uma amostra proveniente de três ambientes: uma creche, uma escola e uma pré-escola de Brasília-DF. Ao todo, participaram 49 crianças, equitativamente distribuídas quanto ao sexo: 49% do sexo masculino e 51% feminino, com idade média de 4,7 anos (DP = 1,2).

O estudo de validação do SON-R 2½-7[a] com o WISC-III foi realizado com uma amostra de 50 crianças de Brasília-DF, São José dos Pinhais-PR, Curitiba-PR e Porto Alegre-RS, a maioria do sexo feminino (56%), todos com idade entre 6 e 7 anos. Quanto à escolaridade, a maioria dos participantes estava cursando o 2º ano do ensino fundamental (68%).

A amostra de validação do SON-R 2½-7[a] com a Escala de Maturidade Mental Colúmbia (EMMC) e as Matrizes Progressivas Coloridas de Raven (MPC) foi composta por 120 crianças, com idade entre 5 e 7 anos, a maioria do sexo masculino (56,7%). Os testes foram aplicados em escolas públicas e privadas do Distrito Federal, selecionadas tendo como critério a existência de crianças com idade entre seis e sete anos.

4.1.3 Pesquisa nacional de normatização

Para a pesquisa nacional deu-se início ao processo de amostragem, realizado tendo como base os dados do censo populacional brasileiro mais recente, de 2000 (IBGE, 2000). A população base para o processo de amostragem foi as crianças recenseadas de 1 a 9 anos de idade. Durante a elaboração do plano amostral foram consideradas as seguintes informações: (1) densidade demográfica e (2) grau de desenvolvimento socioeconômico de todas as cidades brasileiras, conforme os dados do Índice de Desenvolvimento Humano - IDH. Nesse estudo o IDH adotado é composto por três fatores: renda, longevidade e educação, chamado IDHM. Foi utilizada a classificação de IDH adotada pelo Programa das Nações Unidas para o Desenvolvimento (PNUD) de 2000, que classifica esse índice em cinco categorias. Na Tabela 4.1 pode-se verificar a quantidade de cidades e o número de crianças em cada umas das cinco categorias de IDHM.

Tabela 4.1 Número de cidades, número de crianças de 1 a 9 anos, e percentagem de crianças de 1 a 9 anos por categoria de IDHM

Categorias de IDHM	Número de cidades	Número de crianças de 1 a 9 anos	Percentagem de crianças de 1 a 9 anos
IDHM \geq 0,801	558	9.407.143	31,67
IDHM entre 0,751 e 0,800	1.305	7.832.807	26,37
IDHM entre 0,701 e 0,750	1.117	3.960.039	13,33
IDHM entre 0,651 e 0,700	787	2.982.895	10,04
IDHM \leq 0,650	1.740	5.521.861	18,59
Total	5.507	29.704.745	100,00

Nota: IDHM é a média de três indicadores: expectativa de vida, saúde e educação.

A Tabela 4.1 mostra que, segundo os dados do censo de 2000, cerca de 30 milhões de crianças no Brasil estava na faixa etária de 1 a 9 anos de idade, o que representava cerca de 18% da população total. O mapa a seguir apresenta a situação dos 26 estados brasileiros e do Distrito Federal em relação às cinco categorias de IDHM.

A determinação do número de elementos por região (o primeiro estrato da seleção) foi baseada no seguinte cálculo: [número de crianças com idade de 1 a 9 anos em uma determinada região] dividido pelo [número total de crianças com idade de 1 a 9 anos] multiplicado por 60. A multiplicação com o fator 60 foi feita porque a amostra consiste de 60 elementos. Os resultados desses cálculos são apresentados na próxima tabela.

Tabela 4.3 Quantidade de cidades por região e quantidade de elementos a serem testados por região

Região	Número de cidades	Número de elementos	Número de elementos (truncado*)	Número de crianças a testar
Norte	449	5,90	6	120
Nordeste	1.787	18,58	19	380
Sudeste	1.666	23,04	23	460
Sul	1.159	8,28	8	160
Centro Oeste	446	4,20	4	80
Total	5.507	60,00	60	1.200

Nota: *O número de elementos foi truncado devido à impossibilidade de trabalhar com números decimais para representar quantidade de crianças.

A Tabela 4.3 mostra que o maior percentual de crianças a testar é na região Sudeste, seguida pela região Nordeste. A seguir, a Tabela 4.4 apresenta a quantidade de elementos a testar em cada categoria de IDH.

Tabela 4.4 Número de municípios e elementos por categoria de IDHM

Categorias de IDH	Nº de municípios	Nº de elementos
IDHM $\geq 0,801$	558	19,00
IDHM entre 0,751 e 0,800	1.305	15,82
IDHM entre 0,701 e 0,750	1.117	8,00
IDHM entre 0,651 e 0,700	787	6,03
IDHM $\leq 0,650$	1.740	11,15
Total	5.507	60,00

A fim de estabelecer o número de elementos em cada região por categoria de IDH foi feita uma combinação do critério de estratificação de elementos por região e o número de elementos por categoria de IDHM. A Tabela 4.5 apresenta os resultados dessa combinação.

Tabela 4.5 Quantidade de elementos em cada região por categoria de IDHM

Região	Categorias de IDHM	Nº de elementos	Nº de elementos*	Nº de crianças
Norte	IDHM \geq 0,801	0,44	1	20
	IDHM entre 0,751 e 0,800	1,53	1	20
	IDHM entre 0,701 e 0,750	1,19	1	20
	IDHM entre 0,651 e 0,700	1,52	2	40
	IDHM \leq 0,650	1,22	1	20
	Total		5,90	6
Nordeste	IDHM \geq 0,801	0,75	1	20
	IDHM entre 0,751 e 0,800	2,84	3	60
	IDHM entre 0,701 e 0,750	2,28	2	40
	IDHM entre 0,651 e 0,700	3,28	3	60
	IDHM \leq 0,650	9,43	10	200
	Total		18,58	19
Sudeste	IDHM \geq 0,801	11,79	12	240
	IDHM entre 0,751 e 0,800	7,45	7	140
	IDHM entre 0,701 e 0,750	2,47	2	40
	IDHM entre 0,651 e 0,700	0,87	1	20
	IDHM \leq 0,650	0,45	1	20
	Total		23,04	23
Sul	IDHM \geq 0,801	4,38	4	80
	IDHM entre 0,751 e 0,800	2,66	3	60
	IDHM entre 0,701 e 0,750	0,99	1	20
	IDHM entre 0,651 e 0,700	0,22	-	-
	IDHM \leq 0,650	0,03	-	-
	Total		8,28	8
Centro-Oeste	IDHM \geq 0,801	1,64	2	40
	IDHM entre 0,751 e 0,800	1,34	1	20
	IDHM entre 0,701 e 0,750	1,06	1	20
	IDHM entre 0,651 e 0,700	0,14	-	-
	IDHM \leq 0,650	0,03	-	-
	Total		4,20	4
Total	IDHM \geq 0,801	19,00	19	380
	IDHM entre 0,751 e 0,800	15,82	16	320
	IDHM entre 0,701 e 0,750	8,00	8	160
	IDHM entre 0,651 e 0,700	6,03	6	120
	IDHM \leq 0,650	11,15	11	220
	Total		60,00	60

Nota: *Considerou-se apenas números inteiros para a quantidade de elementos.

Para a seleção dos municípios em cada região, além do IDHM do município, foram utilizados os seguintes critérios: (a) dentro de cada região deveria haver sempre capital e interior; (b) as cidades selecionadas deveriam contemplar os maiores estados da região; (c) estados com maiores e menores IDH do país deveriam ser contemplados na amostra; (d) disponibilidade de aplicadores no município ou próximo a ele e (e) a facilidade de acesso. A Tabela 4.6 mostra o plano que foi elaborado com base nesses critérios para a execução da pesquisa de normatização do SON-R 2½-7[a].

Tabela 4.6 Municípios selecionados em cada região

Região	Categorias de IDHM	Nº de crianças	Municípios
Norte	IDHM >= 0,801	20	Belém (PA)
	IDHM entre 0,751 e 0,800	20	Manaus (AM)
	IDHM entre 0,701 e 0,750	20	Araguaína (TO)
	IDHM entre 0,651 e 0,700	40	Tocantinópolis (TO)
	IDHM <= 0,650	20	Augustinópolis (TO)
	Total	120	
Nordeste	IDHM >= 0,801	20	Salvador (BA)
	IDHM entre 0,751 e 0,800	60	São Luis (MA), Fortaleza (CE)
	IDHM entre 0,701 e 0,750	40	Imperatriz (MA)
	IDHM entre 0,651 e 0,700	60	Estreito (MA), Sobral (CE)
	IDHM <= 0,650	200	São J. do Paraíso (MA), Irará (BA), Ipecaetá (BA), Acarape (CE), Chorozinho (CE)
	Total	380	
Sudeste	IDHM >= 0,801	240	São Caetano do Sul (SP), São Paulo (SP), Belo Horizonte (MG), Rio de Janeiro (RJ), Campinas (SP)
	IDHM entre 0,751 e 0,800	140	Guarulhos (SP), Sumaré (SP), Esmeraldas (MG)
	IDHM entre 0,701 e 0,750	40	São José do Vale do Rio Preto (RJ)
	IDHM entre 0,651 e 0,700	20	Paulistas (MG)
	IDHM <= 0,650	20	Ribeirão Branco (SP)
	Total	460	
Sul	IDHM >= 0,801	80	Porto Alegre (RS), Curitiba (PR)
	IDHM entre 0,751 e 0,800	60	São José dos Pinhais (PR), São Gabriel (RS)
	IDHM entre 0,701 e 0,750	20	Dom Feliciano (RS)
	IDHM entre 0,651 e 0,700	-	
	IDHM <= 0,650	-	
	Total	160	

Região	Categorias de IDHM	Nº de crianças	Municípios
Centro-	IDHM \geq 0,801	40	Brasília (DF), Goiânia (GO)
Oeste	IDHM entre 0,751 e 0,800	20	Valparaíso de Goiás (GO)
	IDHM entre 0,701 e 0,750	20	Pirenópolis (GO)
	IDHM entre 0,651 e 0,700	-	
	IDHM \leq 0,650	-	
	Total	80	

A Tabela 4.6 apresenta os 36 municípios, localizados em 13 estados brasileiros, para a realização da pesquisa com o SON-R 2½-7[a]. O número de municípios por região é o seguinte: Norte (5), Nordeste (11), Sudeste (11), Sul (5) e Centro-Oeste (4).

Após a seleção dos municípios foram determinados os locais de aplicação. Considerando que um grande percentual das crianças menores ainda não freqüenta escola ou creche, muitas crianças com menos de 6 anos de idade precisaram ser testadas em outros lugares, que não em escolas, tais como: igrejas, pastoral da criança, ONGs ou em casa. A decisão sobre o local de testagem (em escola versus fora da escola) foi tomada considerando o percentual de crianças matriculadas em creches e pré-escolas, segundo o Censo de 2000. A Tabela 4.7 apresenta os percentuais por município.

Tabela 4.7 Percentual de crianças que freqüentam escola nos municípios selecionados (Censo de 2000)

Município	Percentual de crianças entre 4 e 5 anos freqüentando escola*	Percentual de crianças entre 5 e 6 anos freqüentando escola***
Manaus (AM)	43,51	61,38
Belém (PA)	69,23	81,86
Araguaína (TO)	44,89	67,62
Augustinópolis (TO)	29,47	51,12
Tocantinópolis (TO)	62,50	85,33
Estreito (MA)	30,97	64,12
Imperatriz (MA)	85,04	93,21
São João do Paraíso (MA)	40,35	72,02
São Luís (MA)	80,10	89,95
Acarape (CE)	87,86	89,32
Chorozinho (CE)	68,18	86,38
Fortaleza (CE)	76,07	85,65
Sobral (CE)	69,42	80,30
Irará (BA)	66,96	86,21
Ipecaetá (BA)	55,61	72,81
Salvador (BA)	74,16	83,44
Belo Horizonte (MG)	63,61	80,33
Esmeraldas (MG)	25,04	54,95
Paulistas (MG)	29,24	69,49
Rio de Janeiro (RJ)	70,00	84,28
São José do Vale do Rio Preto (RJ)	53,64	80,77
Campinas (SP)	52,48	71,63
Guarulhos (SP)	37,87	66,24
Ribeirão Branco (SP)	09,04	31,41
São Caetano do Sul (SP)	91,24	94,74
São Paulo (SP)	54,05	75,21
Sumaré (SP)	37,05	63,50
Curitiba (PR)	52,00	72,01
São José dos Pinhais (PR)	28,92	56,23
Dom Feliciano (RS)	07,23	28,32
Porto Alegre (RS)	41,21	60,84
São Gabriel (RS)	23,56	52,76
Pirenópolis (GO)	32,91	70,01
Goiânia (GO)	55,20	79,14
Valparaíso de Goiás (GO)	55,04	73,51
Brasília (DF)	54,59	74,31

Nota: *Esses percentuais foram usados para calcular o número de crianças entre 3;3 e 4;9.

** Esses percentuais foram usados para calcular o número de crianças entre 5;3 e 7;9.

Considerando as informações apresentadas nas Tabelas 4.6 e 4.7, foi calculada a quantidade de crianças que precisavam ser testadas em cada condição: frequentando e não frequentando escola. A Tabela 4.8 sumariza os resultados.

Tabela 4.8 Plano amostral segundo a quantidade de crianças escolares e não escolares*

Município	Número total de crianças	Número de crianças escolares	Número de crianças não-escolares	Percentual de crianças não-escolares
Belém (PA)	20	16	4	
Manaus (AM)	40	22	18	
Araguaína (TO)	20	12	8	
Tocantinópolis (TO)	20	15	5	
Augustinópolis (TO)	20	8	12	
Norte	120	73	47	39,17
Porto Alegre (RS)	40	22	18	
São Gabriel (RS)	20	8	12	
Dom Feliciano (RS)	20	4	16	
Curitiba (PR)	40	25	15	
São José dos Pinhais (PR)	40	18	22	
Sul	160	77	83	51,88
Salvador (BA)	20	16	4	
Irará (BA)	40	32	8	
Ipecaetá (BA)	40	24	16	
São Luiz (MA)	20	17	3	
Estreito (MA)	20	10	10	
São João do Paraíso (MA)	40	23	17	
Imperatriz (MA)	40	36	4	
Fortaleza (CE)	40	33	7	
Chorozinho (CE)	40	32	8	
Acarape (CE)	40	31	9	
Sobral (CE)	40	30	10	
Nordeste	380	284	96	25,26
Belo Horizonte (MG)	60	44	16	
Esmeraldas (MG)	60	26	34	
Paulistas (MG)	20	10	10	
Rio de Janeiro (RJ)	40	31	9	
S. J. do V. R. Preto (RJ)	40	28	12	
Ribeirão Branco (SP)	20	5	15	
Sumaré (SP)	40	21	19	
Guarulhos (SP)	40	22	18	
São Paulo (SP)	60	40	20	
São Caetano do Sul (SP)	40	38	2	
Campinas (SP)	40	25	15	
Sudeste	460	290	170	36,96
Brasília (DF)	20	13	7	
Goiânia (GO)	20	13	7	
Valparaíso (GO)	20	13	7	
Pirenópolis (GO)	20	10	10	
Centro-Oeste	80	49	31	38,75
Total	1200	773	427	35,58

Nota: *Baseado no Censo de 2000 (IBGE, 2000).

Ajustes no plano amostral

No Segundo semestre de 2007 os dados da Pesquisa Nacional por Amostragem de Domicílios realizada em 2006 foram publicados (IBGE, 2007). Conforme esperado, o percentual de crianças escolares aumentou, quando comparado com o percentual apresentado no Censo de 2000. A próxima tabela apresenta uma comparação por região entre os dados do Censo de 2000 e da PNAD 2006.

Tabela 4.9 Comparação dos percentuais de crianças não-escolares

Região	% Censo de 2000	% PNAD 2006*
Norte	39,2	35,8
Sul	51,9	33,6
Nordeste	25,3	19,6
Sudeste	37,0	19,1
Centro-Oeste	38,8	33,0
Total	35,6	24,0

*Nota: Baseado no número de crianças de 4 a 6 anos de idade.

O plano amostral foi ajustado, durante a pesquisa, para adequar aos percentuais da PNAD 2006 no que se refere à quantidade de crianças não-escolares. A Tabela 4.10 apresenta o plano ajustado.

Tabela 4.10 Plano amostral com ajuste da quantidade de crianças não-escolares, segundo dados da PNAD 2006

Município	Nº de crianças	Escolares	Não-Escolares	Rede Pública	Rede Privada
Belém (PA)	20	16	4	X	
Manaus (AM)	40	26	14	X	
Araguaína (TO)	20	12	8		X
Tocantinópolis (TO)	20	15	5		X
Augustinópolis (TO)	20	8	12	X	
Norte	120	77	43		
Porto Alegre (RS)	40	28	12		X
São Gabriel (RS)	20	14	6	X	
Dom Feliciano (RS)	20	10	10	X	
Curitiba (PR)	40	30	10	X	
São José dos Pinhais (PR)	40	24	16	X	
Sul	160	106	54		
Salvador (BA)	20	20	-	X	

Município	Nº de crianças	Escolares	Não-Escolares	Rede Pública	Rede Privada
Irará (BA)	40	32	8	X	
Ipecaetá (BA)	40	26	14	X	
São Luiz (MA)	20	20	-		X
Estreito (MA)	20	10	10	X	
São João do Paraíso (MA)	40	26	14	X	
Imperatriz (MA)	40	36	4		X
Fortaleza (CE)	40	36	4	X	
Chorozinho (CE)	40	32	8	X	
Acarape (CE)	40	34	6		X
Sobral (CE)	40	34	6		X
Nordeste	380	306	74		
Belo Horizonte (MG)	60	52	8	X	X
Esmeraldas (MG)	60	36	24		X
Paulistas (MG)	20	16	4	X	
Rio de Janeiro (RJ)	40	38	2		X
S. J. do V. R. Preto (RJ)	40	34	6	X	
Ribeirão Branco (SP)	20	10	10	X	
Sumaré (SP)	40	32	8	X	
Guarulhos (SP)	40	32	8	X	
São Paulo (SP)	60	50	10	X	X
São Caetano do Sul (SP)	40	40	-	X	X
Campinas (SP)	40	32	8	X	X
Sudeste	460	372	88		
Brasília (DF)	20	14	6	X	
Goiânia (GO)	20	14	6		X
Valparaíso (GO)	20	14	6		X
Pirenópolis (GO)	20	12	8	X	
Centro-Oeste	80	54	26		
Total	1.200	915	285		

A decisão sobre o percentual de escolas da rede pública e privada foi tomada tendo como base as informações do Censo escolar de 2006. A Tabela 4.11 sumariza essas informações para o Brasil.

Tabela 4.11 Quantidade de crianças matriculadas na rede pública e privada (Censo Escolar de 2006)

Idade	Estadual	Federal	Municipal	Particular	Total	% Estad	% Feder	% Munic	% Part
3 anos	8.540	332	400.795	315.639	725.306	1,18	0,05	55,26	43,52
4 anos	28.790	406	915.267	449.143	1.393.606	2,07	0,03	65,68	32,23
5 anos	68.189	513	1.535.504	490.554	2.094.760	3,26	0,02	73,30	23,42
6 anos	10.1916	588	1.479.771	403.501	1.985.776	5,13	0,03	74,52	20,32
7 anos	27.827	176	194.924	87.530	310.457	8,96	0,06	62,79	28,19

Considerando os percentuais para o Brasil e para cada região, procurou-se representar na amostra, de certa forma, o percentual de crianças que estuda na rede pública (cerca de 70%) e particular (cerca de 30%). Optou-se por selecionar escolas particulares tanto na capital quanto no interior, algumas cidades não possuíam escola particular (de acordo com os dados do Censo Escolar de 2006), fazendo com que não houvesse escolha quanto ao tipo de rede. Também optou-se por não dividir os elementos ao definir o tipo de rede, ou seja, ao ser definida a rede pública para uma cidade, na maioria dos casos, todos os escolares daquela cidade foram coletados nesse tipo de escola. Outro critério utilizado ao definir o tipo de rede escolar em cada cidade foi o percentual de crianças em cada idade, ou seja, se em determinada cidade havia um percentual muito pequeno de crianças menores em escola pública, optou-se por definir a coleta naquela cidade em rede particular. Assim, todas as idades daquela cidade foram coletadas em escola particular também.

Amostra final

A amostra final da pesquisa nacional de normatização ficou composta por 1.200 crianças, divididas equitativamente quanto ao sexo e conforme o plano amostral no que tange à divisão por sexo e idade. O desvio máximo permitido para a idade foi de 2 meses, para mais ou para menos –, embora essa situação não tenha sido recorrente no processo de coleta. No que se refere ao quantitativo previsto por cidade e estado, a amostra final sofreu pequenas alterações, como pode ser visto na tabela a seguir.

Tabela 4.12 Quantitativo previsto e obtido de crianças por cidade e estado na amostra final

Município	Nº de crianças previstas	Nº de crianças obtidas	Nº de crianças substituídas	Cidade original
Belém (PA)	20	20	-	
Manaus (AM)	40	40	-	
Araguaína (TO)	20	20	-	
Tocantinópolis (TO)	20	20	-	
Augustinópolis (TO)	20	20	-	
Norte	120	120	-	
Porto Alegre (RS)	40	40	-	
São Gabriel (RS)	20	20	-	
Dom Feliciano (RS)	20	20	-	
Curitiba (PR)	40	40	-	
São José dos Pinhais (PR)	40	39	1	Porto Alegre (RS)
Sul	160	159	1	
Salvador (BA)	20	20	-	
Irará (BA)	40	40	-	
Ipecaetá (BA)	40	39	1	Irará (BA)
São Luiz (MA)	20	20	-	
Estreito (MA)	20	20	-	
São João do Paraíso (MA)	40	40	-	
Imperatriz (MA)	40	40	-	
Fortaleza (CE)	40	40	-	
Chorozinho (CE)	40	40	-	
Acarape (CE)	40	40	-	
Sobral (CE)	40	40	-	
Nordeste	380	379	1	
Belo Horizonte (MG)	60	60	-	
Esmeraldas (MG)	60	34	26	Raposos (MG)
Paulistas (MG)	20	20	-	
Rio de Janeiro (RJ)	40	40	-	
S. J. do V. R. Preto (RJ)	40	38	2	Acarape (CE) e S. Caetano do Sul (SP)
Ribeirão Branco (SP)	20	20	-	
Sumaré (SP)	40	40	-	
Guarulhos (SP)	40	38	2	São Paulo (SP)
São Paulo (SP)	60	59	1	Guarulhos (SP)
São Caetano do Sul (SP)	40	40	-	
Campinas (SP)	40	40	-	
Sudeste	460	429	31	
Brasília (DF)	20	20	-	

Município	Nº de crianças previstas	Nº de crianças obtidas	Nº de crianças substituídas	Cidade original
Goiânia (GO)	20	19	1	Brasília (DF)
Valparaíso (GO)	20	20	-	
Pirenópolis (GO)	20	20	-	
Centro-Oeste	80	79	1	
Total	1.200	1.166	34	

Nota: *Cidade original das crianças que foram substituídas

Como pode ser observado nessa tabela, houve poucas substituições de crianças (2,8% na amostra total) e a cidade onde ocorreu o maior número de substituições foi Esmeraldas (MG), com 43,3% de crianças substituídas por outras com as mesmas características de sexo e idade da cidade de Raposos (MG). A cidade de Raposos foi escolhida por ser do mesmo estado e possuir IDHM correspondente ao de Esmeraldas. Nos demais casos de substituição, também se buscou seguir esse critério, embora não tenha sido possível em todos.

O total de crianças não-escolares na amostra ficou próximo (19,2%) dos 23,8% previsto. E o quantitativo de escolares provenientes de escolas particulares foi 34,2%, não se desviando muito do esperado (cerca de 30%). Quanto às faixas de IDHM das cidades pesquisadas, verifica-se o resultado apresentado na próxima tabela.

Tabela 4.13 Quantitativo de crianças por faixa de IDHM das cidades

Faixas de IDHM	Frequência	%
<= 0,700	340	28,3
Entre 0,701 e 0,800	459	38,3
>=0,801	401	33,4
Total	1200	100

4.2 Instrumentos

(1) SON-R 2½-7[a] - Teste não-verbal de inteligência, de aplicação individual, baseado no SON-R 2½-7 (Tellegen & cols., 1998). Esse teste avalia dois fatores e é composto por quatro subtestes, a saber: testes de raciocínio (Categorias e Situações) e testes de raciocínio espacial (Mosaicos e Padrões).

(2) Questionário para os pais ou responsáveis. Esse instrumento tem como objetivo coletar variáveis de contexto e familiares dos participantes. Entre outras questões é solicitada a profissão dos pais/responsáveis e indicadores do nível socioeconômico, tais como: nível de escolaridade dos pais e quantidade de bens materiais na residência respondida livremente em uma escala de números inteiros iniciando no zero.

Juntamente com o SON-R 2½-7[a] foram aplicados, com parte da amostra, outros testes semelhantes a fim de verificar sua validade convergente:

(3) WPPSI-III (*Wechsler Preschool and Primary Scale of Intelligence – Third Edition*) (Wechsler, 2002a). Traduzido por Andrea Quesada e revisado por Rosana Tristão e Elaine Arima (manuscrito não publicado). Esse teste tem um total de 14 subtestes: 7 Verbais, 5 de Execução e 2 subtestes de Velocidade de Processamento. Os subtestes Verbais são: Informação, Vocabulário, Raciocínio Verbal, Compreensão, Semelhanças, Vocabulário Expressivo e Nomear Figuras. Os subtestes de Execução são: Cubos, Raciocínio Matricial, Categorias, Completar Figuras e Armar Objetos. Os subtestes Completar Símbolos e Códigos fornecem uma medida de velocidade de processamento.

O WPPSI-III é subdividido por faixa etária: para crianças com idade de 2;6-3;11 e para crianças com idade de 4;0-7;3. Há três tipos de subtestes para as duas faixas etárias: Essenciais, Suplementares e Opcionais. A faixa etária de 2;6-3;11 é avaliada por meio de 5 subtestes – Vocabulário expressivo, Informação, Cubos e Armar Objetos (Subtestes Essenciais) e Nomear Figuras (Subteste Suplementar). Para a faixa etária de 4;0-7;3 há 14

subtestes – Informação, Vocabulário, Raciocínio Verbal, Cubos, Raciocínio Matricial, Categorias e Códigos (subtestes Essenciais); Procurar Símbolos, Compreensão, Completar Figuras, Semelhanças e Armar Objetos (subtestes Suplementares) e Vocabulário Expressivo e Nomear Figuras (subtestes Opcionais).

Os subtestes Essenciais fornecem o QI geral do teste e devem ser usados quando se deseja os escores de QI Verbal, QI de Execução e QI Geral. A administração de subtestes Essenciais fornece escores úteis para situações clínicas. Os subtestes Suplementares provêm uma amostra mais ampla do funcionamento intelectual e cognitivo do que os subtestes Essenciais sozinhos. Os subtestes Suplementares são utilizados para substituir os subtestes Essenciais e para compor escores adicionais (por exemplo: Quociente de Velocidade de Processamento (QVP) para as idades de 4;0-7;3 ou o Escore de Linguagem Geral (ELG) para as idades de 2;6-3;11). Subtestes Opcionais são usados para compor escores opcionais (isto é, o ELG para as idades de 4;0-7;3), mas não pode ser usado para substituir qualquer subteste Essencial (Wechsler, 2002a).

(4) WISC-III (Escala de Inteligência Weschler para Crianças). É a terceira edição da Escala de David Wechsler para avaliar a inteligência de crianças. Esse teste foi desenvolvido para uso com crianças entre 6 anos e 0 mês e 16 anos e 11 meses. Foi traduzido e adaptado para o Brasil por Vera Figueiredo a partir de trabalhos realizados em escolas com 801 crianças, na região Sul do estado do Rio Grande do Sul. O teste é composto por 13 subtestes divididos em dois grupos: Verbal (Informação, Semelhanças, Aritmética, Vocabulário, Compreensão e Dígitos) e Execução (Completar Figuras, Código, Arranjo de Figuras, Cubos, Armar Objetos, Procurar Símbolos e Labirintos). Esse teste fornece três medidas compostas de inteligência, a saber: QI Verbal, QI de Execução e, a soma dos pontos ponderados nos subtestes verbais e de execução, produz os pontos ponderados em QI Total (Wechsler, 2002b).

(5) Matrizes Progressivas Coloridas de Raven (Angelini & cols., 1999). Trata-se de um teste não-verbal de inteligência que avalia o fator “g” proposto por Spearman. É indicado para crianças de 5 a 11 anos e meio e foi construída para avaliar com maior detalhamento os processos intelectuais de crianças nessa faixa etária, bem como deficientes mentais e pessoas idosas. O teste é dividido em três séries: A, Ab e B, sendo cada uma composta por 12 problemas, impressos com fundo colorido vivo.

(6) Escala de Maturidade Mental Columbia (Burgemeister, Blum & Lorge, 2001). É um teste de aplicação individual, indicado para crianças entre 3 anos e 6 meses e 9 anos e 11 meses. É composto por 92 itens de classificação pictóricos e figurativos que estão organizados em uma série de oito escalas ou níveis sobrepostos. As crianças submetidas ao teste executam o segmento do teste mais adequado à sua idade cronológica. Desta forma, cerca de 55 ou 66 itens são apresentados, dependendo do nível que está sendo administrado. Os itens são compostos por uma série de três a cinco desenhos impressos em um cartão de 6 por 9 polegadas. É solicitado para a criança que olhe para todas as figuras do cartão e selecione aquela que é diferente ou que não se relaciona com as demais.

4.3 Procedimento

Inicialmente o SON-R 2½-7 passou por um processo de validação das figuras do subteste Categorias. Esse subteste foi aplicado em cinco crianças com idades na faixa etária do teste. Tanto este quanto os demais subtestes foram revisados por três especialistas da área de avaliação psicológica que trabalham com crianças. Os problemas detectados nessa fase foram solucionados juntamente com os autores e a editora do teste para que o mesmo pudesse ser aplicado no estudo piloto. Na seção de resultados são relatadas as mudanças ocorridas no teste.

O estudo piloto foi realizado na cidade de Brasília em duas instituições que atendem crianças na faixa etária de interesse do teste. O manual utilizado no estudo piloto foi

previamente traduzido para o português e passou por várias revisões de conteúdo e por revisão de língua portuguesa. O tempo médio de aplicação do SON-R 2½-7[a] no estudo piloto foi de 50 min. Para a pesquisa nacional não houve alterações no teste, pois o estudo piloto indicou que este estava adequado para uso nas condições adotadas.

4.3.1 Procedimento utilizado na pesquisa nacional de normatização

Contato com as instituições

Em cada cidade buscou-se a lista das escolas com seus respectivos endereços e telefones e, quando necessário, buscou-se também o contato das secretarias de educação. Os demais locais de aplicação, como igrejas, pastoral da criança, residências etc. foram visitados e consultados pelos próprios aplicadores. No caso das escolas e secretarias de educação foi feito um contato inicialmente por telefone. Durante esse primeiro contato foram explicitados os objetivos, procedimentos, justificativa da pesquisa e foi solicitada a permissão da escola para sua realização. Para as escolas que concordaram em participar da pesquisa foram enviados pelo correio Termos de Consentimento Livre e Esclarecido (TCLE) para serem entregues aos pais, um plano amostral com a quantidade de crianças necessárias por sexo e faixa etária e um ofício para ser arquivado pela escola. Os pais que aceitaram que seus filhos participassem da pesquisa receberam um relatório com as seguintes informações: o desempenho da criança no teste e sugestão de algumas atividades pedagógicas que pudessem contribuir para o incremento dos tipos de raciocínio avaliados.

Os aplicadores responsáveis pela administração do SON-R 2½-7[a] nos municípios selecionados receberam, previamente para agendar a aplicação, o contato da escola que concordou participar da pesquisa. Eles também receberam da coordenação da pesquisa o plano amostral para a escola indicada e para a cidade como um todo – pois em quase todas as cidades há um percentual de crianças não-escolares. Os aplicadores ficaram responsáveis por

obter as informações socioeconômicas da criança testada por meio dos pais (pessoalmente ou por telefone) ou da escola.

Seleção das crianças

No processo de seleção das crianças um dos critérios mais importantes de seleção foi a idade, buscou-se garantir o máximo de homogeneidade em relação às idades dentro de cada faixa etária. Isso significa que, por exemplo, para as crianças da faixa etária mais nova a idade exata da criança poderia desviar o mínimo possível da idade de 3 anos e 3 meses. Esse critério é especialmente importante para as faixas etárias de crianças mais novas, pois nesses grupos um período de 3 meses, por exemplo, pode representar um grau significativo de desenvolvimento. Outro objetivo desse critério é evitar viés na seleção. Um viés na seleção poderia ocorrer se em uma determinada escola as crianças fossem escolhidas de acordo com as preferências dos professores e não de acordo com a idade. O critério da idade é muito importante para se obter uma amostra aleatória das crianças.

Um segundo critério importante de seleção foi o percentual de crianças que freqüenta escola em um determinado município. Se, por exemplo, apenas 60% das crianças de certa faixa etária freqüentam escola, 40% precisam ser encontradas fora da escola. Para selecionar as crianças que não freqüentam escola várias estratégias foram utilizadas: (1) encontrá-las em postos de saúde; (2) pastoral da criança; (3) Organizações Não Governamentais; (4) visitando bairros; (5) igrejas; (6) por meio de listas de espera por vaga de escolas públicas; e (7) com a ajuda de prefeituras que possuem lista de crianças que recebem assistência do município.

Ademais, foram determinados critérios de exclusão de crianças da amostra, a saber: crianças diagnosticadas com algum problema físico ou condição que possa afetar seu desempenho no teste (epilepsia, meningite, encefalite, tumor cerebral etc.); crianças com problema de visão; problema de audição; alguma deficiência motora que possa afetar seu

desempenho no teste; crianças que estejam fazendo uso de algum medicamento (ansiolítico, antidepressivo etc.) ou que estejam recebendo tratamento em hospitais psiquiátricos.

Aspectos éticos

O projeto de pesquisa passou pelo Comitê de Ética em Pesquisa da Faculdade de Medicina, da Universidade de Brasília, antes da realização do estudo piloto no Distrito Federal. Em algumas cidades, durante o processo de aplicação, foi solicitado aos pais ou responsáveis permissão para tirar fotos das crianças. Aqueles que concordaram, tiveram suas crianças fotografadas ao responderem o teste. Quatro dessas fotos foram utilizadas na presente tese.

Treinamento dos aplicadores

Os aplicadores foram treinados em Brasília, divididos em dois grupos: um com sete e outro com oito pessoas a fim de que se pudesse dispensar atenção suficiente para cada examinador. Durante o treinamento os seguintes materiais foram utilizados: o teste SON-R 2½-7[a], o manual em português, um cronômetro e folhas de resposta em português, as quais foram utilizadas na pesquisa nacional.

O treinamento para cada grupo teve duração de dois dias. O primeiro grupo começou o treinamento no início de maio de 2007. O treinamento incluiu uma parte teórica relacionada aos testes SON, a relevância da pesquisa de normatização e uma parte sobre amostragem. A parte prática incluiu o processo de testar crianças utilizando o teste SON. Durante a parte prática os aplicadores treinaram em duplas, invertendo os papéis entre criança e aplicador nos subtestes. Também foi apresentada por meio de DVD uma aplicação do teste SON-R 2½-7[a] com uma criança.

Durante a pesquisa dois aplicadores foram substituídos e um novo aplicador foi agregado à equipe, nesses casos o treinamento foi feito em Salvador-BA, seguindo os procedimentos adotados em Brasília-DF.

Examinadores

Ao todo, 18 aplicadores foram recrutados para a pesquisa de normatização do SON-R 2½-7[a] no Brasil. A Tabela 4.14 apresenta quantos e quais examinadores ficaram responsáveis pelos municípios.

Tabela 4.14 Examinadores e municípios nos quais eles aplicaram o SON-R 2½-7[a]

Nome dos examinadores	Municípios	Nº de crianças
1. Darlene	Belém (PA)	20
	Manaus (AM)	40
2. Shana	Porto Alegre (RS)	40
3. Luciana	Curitiba (PR)	40
4. William	São José dos Pinhais (PR)	20
	Dom Feliciano (RS)	40
	São Gabriel (RS)	20
5. Joice	Araguaína (TO)	20
6. Clara	Tocantinópolis (TO)	20
	Augustinópolis (TO)	20
	São Luiz (MA)	20
	Estreito (MA)	20
	São João do Paraíso (MA)	40
	Imperatriz (MA)	40
7. Ana	Salvador (BA)	20
8. Ana	Irará (BA)	40
9. Cinthia	Ipecaetá (BA)	40
	Fortaleza (CE)	40
	Chorozinho (CE)	40
	Acarape (CE)	40
	Sobral (CE)	40
10. Alessandra	Belo Horizonte (MG)	60
11. Ana Carolina	Esmeraldas (MG)	60
12. Tiago	Paulistas (MG)	20
13. Vilma	Rio de Janeiro (RJ)	40
	São José do Vale do Rio Preto (RJ)	40
14. Érica	Ribeirão Branco (SP)	20
15. Leticia	Sumaré (SP)	40
16. Iara	Guarulhos (SP)	40
	São Paulo (SP)	60

Nome dos examinadores	Municípios	Nº de crianças
	São Caetano do Sul (SP)	40
	Campinas (SP)	40
17. Camila	Brasília (DF)	20
18. Zuleide	Goiânia (GO)	20
	Valparaíso (GO)	20
	Pirenópolis (GO)	20
Total		1.200

4.4 Análise dos dados

Os dados foram analisados por meio de procedimentos estatísticos paramétricos de modelagem e análise de itens. Foram efetuadas correlações bivariadas no estudo de validade convergente, além da análise da dificuldade e discriminação dos itens por meio da Teoria Clássica dos Testes (TCT). Uma vez coletados todos os dados verificou-se a validade de construto do teste por meio da solução Schmid-Leiman, viabilizada por uma syntax do *Statistical Package for the Social Sciences* (SPSS), utilizando-se a versão 16.0. A Análise Fatorial de Informação Plena (FIFA) com os itens do teste foi realizada no Testfact 4.0. Foi executada também a Análise Fatorial Confirmatória (AFC) por meio do LISREL 8.80. Os itens foram analisados por meio da Teoria de Resposta ao Item (TRI), utilizando-se, para tanto, o *BILOG-MG 3.0*. A normatização foi executada por meio de uma syntax do SPSS programada pelo professor Peter J. Tellegen, na qual são efetuadas transformações lineares, regressão linear múltipla e regressão logística.

5. Resultados do estudo piloto

Nessa seção serão apresentados os resultados do estudo piloto realizado no Brasil com o SON-R 2½-7.

5.1 Resultados do estudo de validação de conteúdo

No processo de adaptação de testes verbais de uma cultura para outra deve-se dar atenção especial a problemas relacionados com a tradução. Os testes não verbais, por sua vez, também não estão imunes a problemas culturais, como apontam Tellegen e Laros (2004). Dessa forma, o primeiro passo, prévio à normatização e validação do SON-R 2½-7, foi proceder um estudo da adequação do conteúdo do teste para a realidade brasileira.

No estudo de validação de conteúdo do SON-R 2½-7 participaram cerca de dez especialistas da área de psicologia e neuropsicologia infantil e estudantes de graduação e pós-graduação em Psicologia. Nesse estudo o subtteste Categorias foi apontado como problemático. Dessa forma, esse subtteste foi aplicado em cinco crianças com idades na faixa etária do teste. Na primeira parte do subtteste alguns cartões apresentaram-se problemáticos, os quais sofreram modificações da versão do SON-R 2½-7, holandesa, para a versão do SON-R 2½-7[a], versão reduzida brasileira. A seguir são apresentadas as mudanças.

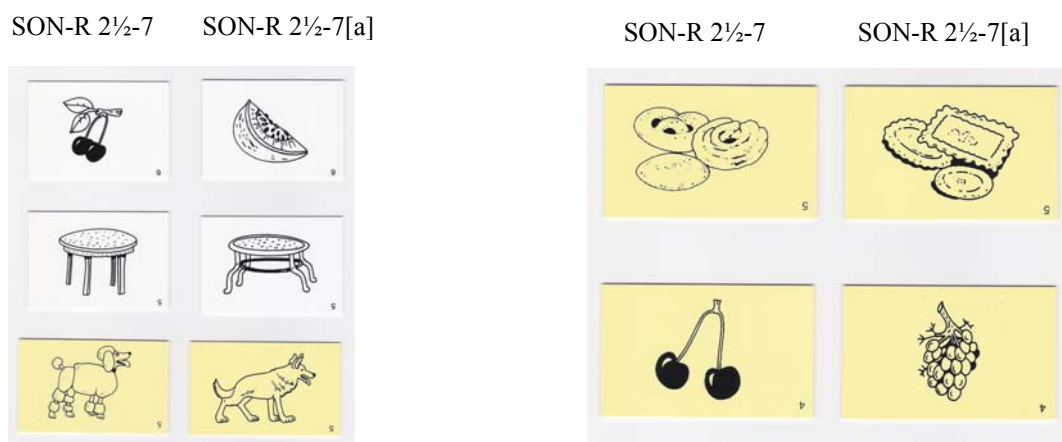


Figura 5.1 Mudanças do SON-R 2½-7 para o SON-R 2½-7[a]

Como pode ser verificado, nas figuras à esquerda, as cerejas, típicas da cultura holandesa, foram substituídas pela melancia, que é mais adequada para o contexto brasileiro. A segunda figura, a cadeira, foi identificada por algumas crianças como um banquinho para sentar, dessa forma, foi trocada por outra modalidade de mesa e o poodle foi substituído por uma raça de cachorro mais típica da cultura brasileira. Já nas figuras do lado direito, as rosquinhas – confundidas com botões – foram substituídas por biscoitos, e as cerejas, novamente, foram substituídas por uvas.

Na segunda parte do subtteste Categorias também foram identificados problemas com algumas figuras, os quais se encontram listados na sequência.

a) Item 10: da página com as alternativas, a terceira figura, o chapéu, foi confundido com sofá e com ferro de passar roupas. Também a quinta figura, da xícara, foi confundida com um pinico. Essas figuras foram substituídas na versão do SON-R 2½-7[a].

Item 10 do SON-R 2½-7



Item 10 do SON-R 2½-7[a]



Figura 5.2 Mudanças no item 10 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a]

b) item 11: na página dos exemplos a terceira figura, do patinador no gelo, mostrou-se problemática por não ser comum no contexto brasileiro e foi substituída, conforme apresentado a seguir.

Item 11 do SON-R 2½-7



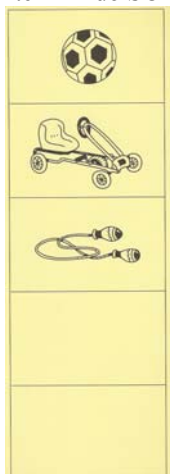
Item 11 do SON-R 2½-7[a]



Figura 5.3 Mudanças no item 11 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a]

c) Item 12: nesse item houve quatro modificações, a primeira figura (a bola) não foi reconhecida como um brinquedo na maioria dos casos, dessa forma, foi substituída pela figura de um ursinho. Por sua vez, a terceira figura, corda de pular, em alguns casos também não foi reconhecida como um brinquedo, assim, seu desenho sofreu modificações (ver Figura 5.4).

Item 12 do SON-R 2½-7



Item 12 do SON-R 2½-7[a]

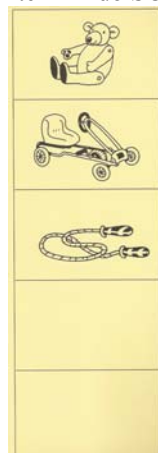
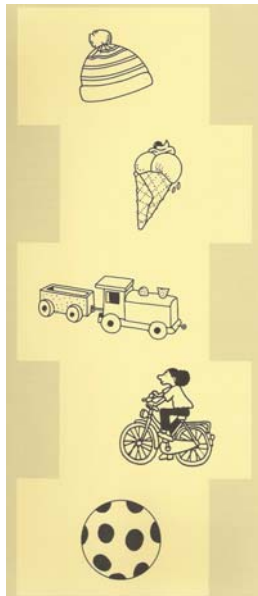


Figura 5.4 Mudanças no item 12 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a]

As duas outras mudanças realizadas no item 12 foram: a substituição da quarta figura, menina de bicicleta, pois nesse contexto a bicicleta pode ser confundida como um brinquedo e a quinta figura, a bola, foi substituída por um tamborzinho de brinquedo (ver Figura 5.5).

Item 12 do SON-R 2½-7



Item 12 do SON-R 2½-7[a]



Figura 5.5 Mudanças no item 13 (alternativas) do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a]

d) Item 13. Nesse item a quinta figura, luva de tomar banho utilizada na Holanda, não existe no Brasil e foi substituída por uma toalha, como pode ser visto na Figura 5.6.

Item 13 do SON-R 2½-7



Item 13 do SON-R 2½-7[a]

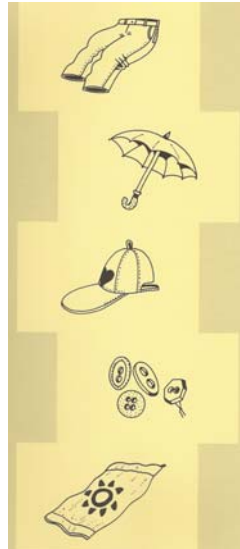


Figura 5.6 Mudanças no item 13 do subtteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a]

e) Item 14: a quarta figura se apresentou problemática pelo tamanho desproporcional da vasilha com a comida do garoto, portanto, foi substituída, conforme Figura 5.7.

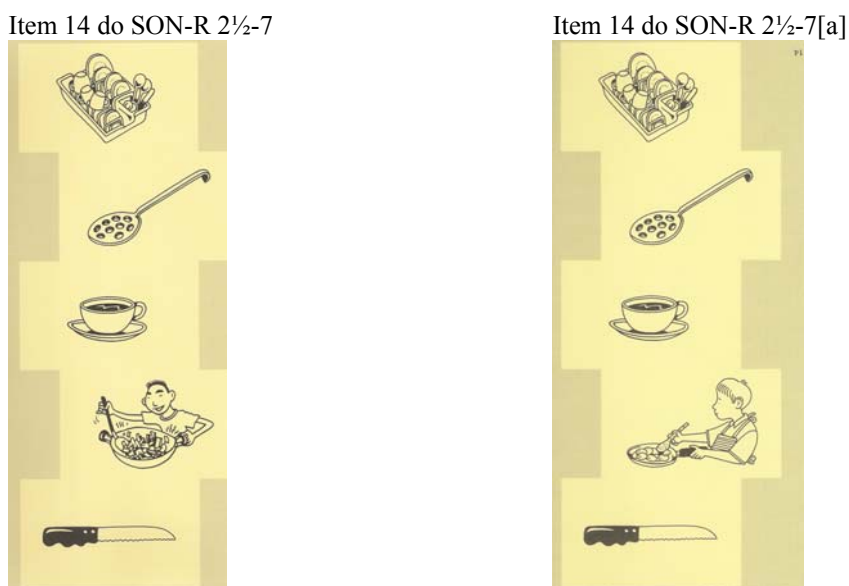


Figura 5.6 Mudanças no item 14 do subteste Categorias do SON-R 2½-7 para o SON-R 2½-7[a]

5.2 Resultados do estudo piloto do SON-R 2½-7[a]

O estudo piloto foi feito também com a versão do SON-R 2½-7[a] em duas instituições de ensino da cidade de Brasília, uma que oferece atendimento regular e a outra que oferece atendimento para crianças com problemas auditivos. Embora o estudo de normatização não contemple essas últimas crianças, o estudo piloto serviu para verificar a adequação das instruções não-verbais de aplicação constantes no manual em português.

Utilizando as normas holandesas foi calculado o QI médio das crianças da instituição de ensino regular (creche) e da que oferece atendimento especial (CEAL), bem como da amostra total. A tabela a seguir apresenta os resultados.

Tabela 5.1 QI de execução, raciocínio e total das crianças da creche e do CEAL (N=130)

Local	Estatísticas	QI		
		Execução	Raciocínio	Total
Creche (N=92)	Média	92,3	95,6	93
	Desvio Padrão	14	15,1	14,3
	Mínimo	57	51	55
	Máximo	128	130	125
CEAL (N=38)	Média	88,8	84,5	84,6
	Desvio Padrão	13	15,5	14,1
	Mínimo	62	52	53
	Máximo	123	116	116
Total	Média	91,3	92,3	90,5
	Desvio Padrão	23,8	16	14,7
	Mínimo	57	51	53
	Máximo	128	130	125

Com o objetivo de verificar a qualidade psicométrica do teste SON-R 2½-7[a] foi realizada uma análise por meio da Teoria Clássica dos Testes (TCT). Dessa forma, foi calculado o índice de dificuldade, ou percentual de acerto, e a correlação item-total corrigida dos itens de todos os subtestes, ou discriminação. Também foi verificada a fidedignidade de cada subteste utilizando o coeficiente Alfa de Cronbach. A Tabela 5.2 apresenta os resultados.

Tabela 5.2 Percentual de acerto, correlação item-total e fidedignidade dos subtestes do SON-R 2½-7[a] (N = 130)

Itens	Mosaicos		Categorias		Situações		Padrões	
	Dif.	r _{it}	Dif.	r _{it}	Dif.	r _{it}	Dif.	r _{it}
1	0,97	0,27	0,96	0,34	0,98	0,16	0,98	0,15
2	0,92	0,43	0,98	0,35	0,98	0,21	0,98	0,28
3	0,75	0,56	0,96	0,27	0,93	0,34	0,96	0,37
4	0,88	0,53	0,89	0,48	0,93	0,36	0,97	0,38
5	0,82	0,56	0,79	0,51	0,71	0,53	0,96	0,30
6	0,74	0,69	0,80	0,49	0,68	0,57	0,88	0,48
7	0,65	0,76	0,68	0,51	0,57	0,61	0,85	0,57
8	0,53	0,70	0,52	0,47	0,60	0,63	0,73	0,59
9	0,29	0,55	0,47	0,57	0,46	0,67	0,65	0,72
10	0,20	0,49	0,28	0,45	0,26	0,56	0,49	0,72
11	0,07	0,33	0,19	0,50	0,15	0,40	0,35	0,67
12	0,05	0,32	0,12	0,40	0,05	0,30	0,28	0,63
13	0,00*	0,00*	0,07	0,42	0,05	0,29	0,16	0,52
14	0,01	0,19	0,05	0,42	0,01	0,12	0,08	0,40
15	0,01	0,19	0,03	0,36	-	-	0,02	0,22
16	-	-	-	-	-	-	0,00*	0,00*
α de Cronbach	0,85		0,81		0,81		0,85	

Nota: r_{it}= correlação item-total. Dif=dificuldade ou % de acerto. * Item com variabilidade zero ou praticamente zero.

Uma análise da Tabela 5.2 mostra que todos os subtestes apresentam boa fidedignidade ($\alpha > 0,80$). Como o SON-R 2½-7[a] é aplicado seguindo um procedimento adaptativo é importante que o índice de dificuldade dos itens de cada subteste seja decrescente. Nesse sentido, vê-se que há poucos desvios. Quanto à correlação item-total corrigida, observa-se problemas apenas nos itens ou muito fáceis ou muito difíceis, devido à falta de variabilidade.

6. Normatização do SON-R 2½-7[a]

O presente estudo é o primeiro de normatização do SON-R 2½-7 em sua versão reduzida, SON-R 2½-7[a]. Nessa versão dois subtestes não foram incluídos: Analogias e Quebra-Cabeças.

Em geral, o primeiro passo para uma normatização é transformar os escores brutos dos subtestes para escores com distribuição normal ($M=0$ e $DP=1$), sendo essas transformações feitas separadamente para cada grupo de idade. A desvantagem desse método, entretanto, é que o número relativo de sujeitos em cada grupo de idade permite que fatores aleatórios tenham um papel importante nas transformações. No SON-R 2½-7[a], um método diferente, desenvolvido para a normatização do SON-R 5½-17, foi aplicado (ver Laros & Tellegen, 1991; Snijders, Tellegen & Laros, 1989).

Nesse método, as distribuições dos escores para todos os grupos de idade são ajustados simultaneamente como uma função contínua da idade. Isso é feito para cada subteste separadamente. A função provê uma estimativa, dependendo da idade, da distribuição dos escores na população. Com o procedimento de ajuste é feito um esforço para minimizar a diferença entre a distribuição observada e a distribuição populacional estimada, limitando assim o número de parâmetros da função.

Dentro da classificação de idade do modelo, duas condições devem ser satisfeitas:

1. Para cada idade, o escore padronizado deve aumentar se o escore bruto aumentar.
2. Para cada escore bruto, o escore padronizado deve diminuir se a idade aumenta.

Uma grande vantagem desse método é que o uso da informação de todos os grupos de idade simultaneamente torna a normatização mais precisa. Ademais, os escores normatizados podem ser calculados com base na idade exata. O modelo também permite fazer

extrapolações fora da faixa de idade na qual a pesquisa de normatização foi feita. No SON-R 2½-7[a] o modelo compreende a faixa dos 2 anos e 6 meses a 7 anos e 11 meses.

Frente ao exposto, o objetivo do processo de normatização descrito a seguir é transformar os escores brutos obtidos pelas crianças em cada nível de idade em distribuições normais. Na população a relação da distribuição de escore com a idade mudará gradualmente. Pelo presente método de normatização procura-se fazer essa estimação de forma que fique simples, por um lado (e não muito dependente de flutuações aleatórias), mas ainda bastante próxima aos dados, por outro. A utilização do método de regressão *stepwise* ao longo do processo de normatização tem como objetivo selecionar as variáveis que são mais relevantes.

A seguir será descrito o processo de normatização para o Brasil do SON-R 2½-7[a]. De modo geral, é o mesmo realizado na Holanda; não obstante, alguns aspectos foram um pouco simplificados ao longo do tempo, naturalmente.

6.1 O Modelo de Regressão Logística

Transformação da idade e dos escores brutos

No processo de normatização foi efetuada a transformação da idade e dos escores brutos das crianças da amostra. A idade foi transformada para uma variável chamada y , de forma que as idades entre 2 e 8 anos ficassem no intervalo entre -1 e +1, bem como os escores. Especialmente quando se trabalha com polinômios, como é o caso do processo de normatização, é mais fácil se os valores possíveis das variáveis independentes ficarem no intervalo entre -1 e +1. Os valores dos parâmetros ficam melhores para serem interpretados e é mais fácil computar derivadas das funções nos limites dos valores possíveis. As equações utilizadas para efetuar as transformações foram:

$$y = (\text{idade} - 5) \div 3 \quad [\text{Eq. 1}]$$

$$x = (\text{escore} - 1 - (\text{Max} - 1 \div 2)) \div (\text{Max} - 1 \div 2) \quad [\text{Eq. 2}]$$

Quando o máximo = 15 e o escore = 0 então $x = -8 \div 7$

Quando o escore = 15 então $x = 1$

Esse último cálculo é feito para o limite inferior, para o limite superior utiliza-se a seguinte fórmula:

$$X = (\text{escore} - (\text{Max} - 1) \div 2) \div ((\text{Max} - 1) \div 2) \quad [\text{Eq. 3}]$$

Quando o escore = 0 então $X = -1$
 Quando o escore = 15 então $X = 8/7$

É relevante ressaltar que, embora o intervalo de idade da amostra brasileira compreenda a faixa entre 3 e 8 anos – ao invés de 2 a 8 anos –, optou-se por utilizar a mesma fórmula de transformação empregada nos demais países, uma vez que o impacto desse procedimento não seria significativo.

Ajuste da relação entre idade e escore

A idade foi transformada de forma que a relação entre a idade e o escore transformados se tornasse linear. O modelo de normatização assume ordens superiores de idade, escore e termos de interação. Esse procedimento leva a distorções (por exemplo, funções que não são crescentes quando deveriam ser), devido a esse fato primeiramente a idade é transformada de tal forma que sua relação com os escores se torne linear. Assim, para a idade haverá termos de interação com escores de ordens superiores.

Por meio da regressão linear múltipla foi analisada a relação entre a idade e os escores, transformados conforme equações 1 e 2. Verificou-se uma relação quadrática, conforme pode ser visto na próxima tabela.

Tabela 6.1 Coeficientes da regressão linear múltipla entre o escore e a idade transformados

Escore transformado	R	Variáveis independentes	β (não-padronizado)	Significância
Mosaicos	0,77	y	0,821	0,000
		y ²	-0,291	0,000
Categorias	0,72	y	0,885	0,000
		y ²	-0,305	0,000
Situações	0,67	y	0,684	0,000
		y ²	-0,159	0,001
Padrões	0,79	y	0,798	0,000
		y ²	-0,264	0,000

O valor do parâmetro quadrático (y^2) é usado (com o sinal invertido) como a variável p_0 no processo de normalização (esse é o primeiro da lista de parâmetros de cada subteste).

Assim, para cada subteste a idade é transformada por meio da seguinte equação:

$$y = y - p_0(y * y - 1) \quad [\text{Eq. 4}]$$

Após a transformação efetuada sobre y utilizando a Equação 4 a correlação dos escores com y é praticamente igual ao índice R da regressão múltipla (combinação de y e y²), a saber: $r = 0,76$; $0,71$; $0,66$ e $0,78$, respectivamente.

Ajuste dos escores dicotomizados

Esse é o passo mais complicado e diferente, uma vez que normalmente o ajuste é feito tendo como base distribuições cumulativas, como foi no SON-R 5½-17. Não obstante, no processo de normalização do SON-R 2½-7 para a Holanda e do SON-R 2½-7[a] para o Brasil foi utilizado outro método, no qual não são necessárias distribuições cumulativas para os grupos de idade. Ao invés disso, cada escore individual foi traduzido, digitalmente, em um “pequeno pedaço” de uma distribuição cumulativa.

Esse procedimento foi efetuado porque quando se tem um grupo de pessoas pode-se falar em distribuição de escores e frequência cumulativa, ou seja, para cada escore é possível calcular a proporção de indivíduos que possui aquele escore ou outro. Por outro lado, quando

se trata de cada criança individualmente não é possível obter uma distribuição de escores. Assim, encontrou-se como solução dicotomizar a distribuição de escores (com apenas zeros e uns) para cada indivíduo. Para cada possibilidade de escore é feita a pergunta: Essa pessoa tem esse escore ou um escore inferior? Um exemplo é o seguinte: o escore mínimo possível no subteste Categorias é 0 e o máximo é 15. Dessa forma, se uma criança tem o escore 9, digitalizando seu escore ela recebe 0 para os escores de 0 a 8 e recebe 1 para os escores 9 a 15. Uma vez efetuado esse procedimento é possível utilizar a regressão logística com esses dados, como função da idade, do escore e da interação entre esses termos.

Análise de regressão logística

Durante o início da infância as crianças mostram o desenvolvimento inicial dos maiores conceitos e habilidades que são essenciais para o sucesso ocupacional e educacional posterior. Por exemplo, Carey (1978) estimou que as crianças aprendem uma média de seis palavras por dia durante o período de 18 meses até os 6 anos de idade.

Considerando as grandes mudanças cognitivas que ocorrem nos pré-escolares o processo de normatização do SON-R 2½-7[a] tem como objetivo final disponibilizar normas para cada idade exata, ao invés de faixas etárias. Dessa forma, o modelo de regressão logística é utilizado com o intuito de identificar parâmetros que viabilizem a transformação dos escores considerando toda a informação referente aos escores e à idade.

Dessa forma, foram executadas quatro análises de regressão logística, uma para cada subteste, com o método *forward stepwise*. A variável dependente foi o escore dicotomizado, com apenas zeros e uns. Cada criança tem 14 (Situações), 15 (Categorias e Mosaicos) ou 16 (Padrões) possibilidades de escore. As estatísticas descritivas dos escores digitalizados são apresentadas na tabela a seguir.

Tabela 6.2 Estatísticas descritivas dos escores digitalizados

Subteste	Escore máximo	Frequência	Mínimo	Máximo	Média	Desvio padrão
Mosaicos	15	18.000	0	1	0,47	0,50
Categorias	15	18.000	0	1	0,52	0,50
Situações	14	16.800	0	1	0,51	0,50
Padrões	16	19.200	0	1	0,60	0,49

As variáveis independentes utilizadas no processo de regressão logística foram potências da variável escore, x (elevado até a quinta potência), já transformada conforme equações 2 e 3, e da variável idade, y , transformada segundo a Equação 1, bem como a interação entre essas duas variáveis. As variáveis independentes foram calculadas por meio do SPSS 16.0. Dessa forma, as análises de regressão logística para cada subteste tiveram como variáveis independentes: x a x^5 , y , xy a x^5y . Os coeficientes de regressão obtidos estão apresentados na tabela a seguir.

Tabela 6.3 Coeficientes β da regressão logística dos escores digitalizados ajustados pela idade e o escore

Variáveis independentes	Parâmetro	Coeficiente β da regressão			
		Mosaicos	Categorias	Situações	Padrões
x	p_1	-7,389	-4,633	-5,010	-7,525
x^2	p_2	-5,547	0,000	-1,312	0,000
x^3	p_3	0,000	0,000	-1,140	3,030
x^4	p_4	1,466	-1,483	0,000	-1,039
x^5	p_5	-2,502	-0,746	0,000	-1,956
Y	p_6	4,890	4,027	3,374	5,410
Xy	p_7	0,000	0,000	0,000	0,000
x^2y	p_8	3,256	0,835	1,175	0,000
x^3y	p_9	4,972	0,000	0,000	0,000
x^4y	p_{10}	0,000	0,000	0,000	0,000
x^5y	p_{11}	0,000	0,000	0,000	0,000
Constante	p_{12}	-1,943	-1,288	-0,670	-0,526

Depois do procedimento *stepwise* de ajuste, o número de parâmetros selecionados nos subtestes variou de cinco a sete.

Correção da distribuição de escores

Depois do procedimento anterior o banco de dados voltou novamente para 1.200 casos. As distribuições padronizadas para a amostra total foram avaliadas. Pequenos desvios na média, desvio padrão, curtose e assimetria são corrigidos pelos parâmetros p_{13} a p_{16} . No caso dos dados do Brasil apenas p_{14} e p_{16} foram utilizados, pois houve problema somente com curtose. Os demais parâmetros, p_{13} e p_{15} , foram fixados com o valor 0 (zero) no processo de normalização.

Correlação entre os subtestes e cálculo dos escores normalizados

Ao utilizar escores z a correlação é o produto médio desses escores (covariância). Para os quatro subtestes os seis *cross*-produtos foram ajustados com funções de terceiro grau da idade. Ao invés de ajustar as covariâncias individuais como função da idade, por causa das relações pequenas e em geral insignificantes, para todas as correlações, foi utilizado o valor médio das covariâncias individuais. Devido ao fato de não haver relação clara nas correlações (covariâncias) com a idade, as covariâncias foram assumidas como constantes. Uma vantagem é que agora não são necessárias tabelas separadas para os escores totais; apenas tabelas por idade para os escores dos subtestes. Outra consequência é que a generalizabilidade para os escores totais é fixa nos grupos de idade.

Após a regressão logística e as correções, descritas nas seções anteriores, a proporção cumulativa na população na faixa de idade de 3 a 8 anos, pode então ser estimada para todas as combinações possíveis de idade e escore.

No processo de normalização os pesos da regressão (β) são utilizados como parâmetros p_1 - p_{12} . Assim, tendo como base esses parâmetros o escore z da distribuição normal foi computado para o intervalo definido pelos limites inferiores e superiores dos escores do subteste. O passo seguinte foi calcular as proporções de crianças com escores mais altos e

mais baixos dentro de um intervalo e fazer a correspondência com os valores da distribuição normal z . Devido ao fato de a distribuição não ter ficado completamente normal, foi efetuada uma pequena correção sobre a curtose. O escore resultante foi transformado linearmente para a métrica ($M=10$, $DP=3$), escores normatizados.

Os escores totais foram baseados na soma dos escores normatizados dividido pela raiz quadrada da variância esperada. Depois disso, foi aplicada uma pequena correção devido a assimetria.

Ajuste da Fidedignidade

A fidedignidade foi calculada por meio do λ_2 de Guttman (1945 citado em Cronbach, 1996) para cada subteste e grupo de idade. Da mesma forma que o λ_3 (Coeficiente alfa), o λ_2 é uma medida de consistência interna. Entretanto, o λ_2 é preferível quando o número de itens é limitado e se a covariância entre os itens não é constante (Berge & Zegers, 1978).

O primeiro passo para realizar o ajuste da fidedignidade de cada grupo de idade foi transformar a variável idade. Na sequência foi estimado o erro padrão de mensuração (e.p.m.). Os coeficientes de fidedignidade dos subtestes nos 10 grupos de idade, bem como o desvio padrão dos escores nos subtestes normatizados ($M=10$, $DP=3$) e dos escores de QI ($M=100$, $DP=15$) foram utilizados para calcular o e.p.m. para cada grupo.

O e.p.m. de cada subteste foi ajustado com a função de terceiro grau da idade transformada (y') utilizando regressão múltipla *stepwise*. Assim, a variável dependente na regressão foi o e.p.m. de cada subteste e as variáveis independentes foram y , y^2 e y^3 . A Tabela 6.4 apresenta os coeficientes da regressão obtidos.

Tabela 6.4 Coeficientes β do ajuste do e.p.m pela idade

Variáveis independentes	Coeficiente β			
	Mosaicos	Categorias	Situações	Padrões
Constante (p ₁₇)	1,253	1,186	1,358	1,332
Y ¹ (p ₁₈)	0,061	0,414	0,278	0,046
Y ² (p ₁₉)	1,229	-0,372	0,098	-0,147
Y ³ (p ₂₀)	-1,212	0,294	-0,096	0,323

Como pode ser visto nessa tabela, os parâmetros β obtidos na regressão correspondem ao e.p.m. populacional (parâmetros p₁₇-p₂₀). Tendo como base os valores esperados obtidos nessa regressão (equivalentes ao e.p.m. dos subtestes em cada grupo de idade), o coeficiente de fidedignidade ajustado para cada grupo de idade foi computado. Todos os coeficientes de fidedignidade e e.p.m. estão apresentados na seção sobre a validade de construto.

6.2 Os Escores Escalares

Os escores normatizados são apresentados de duas formas, como escores normatizados e como idades de referência. O escore normatizado (também chamado de escore-desvio) mostra quão bem ou mal a criança desempenha em relação à população de crianças da mesma idade. A idade de referência (também chamada idade mental ou idade de teste) mostra em qual idade 50% das crianças na população têm desempenho pior que o sujeito.

Escores normatizados

Os escores normatizados dos subtestes são apresentados em uma escala distribuída normalmente com média 10 e desvio padrão igual a 3. Esses tão conhecidos escores Wechsler variam de 1 a 19. Como resultado dos efeitos de “solo” e de “teto”, os escores mais extremos não ocorrerão em todos os grupos de idade. Os escores brutos dos subtestes são menos diferenciados do que os escores normatizados. Como resultado, somente alguns dos valores no intervalo de 1 a 19 são usados em cada grupo de idade. Entretanto, os valores mostram a posição na distribuição normal com mais precisão do que seria possível com uma escala

menos diferenciada. A soma dos escores normatizados dos quatro subtestes é a base do QI. Esse SON-QI tem média igual a 100 e desvio padrão igual a 15, varia de 50 a 150.

A soma dos escores normatizados de Mosaicos e Padrões é transformada na Escala de Execução (SON-EE) e a soma de Categorias e Situações forma a Escala de Raciocínio (SON-ER). Ambas as escalas, assim como a distribuição de QI, têm média 100 e desvio 15, podendo variar de 50 a 150.

Idade de referência

A idade de referência é derivada do(s) escore(s) bruto(s). A idade efetiva da criança não é importante. Para o intervalo de 3 a 8 anos a idade de referência é apresentada em anos e meses. A idade de referência para o escore total é a idade na qual uma criança com esse escore bruto receberia um escore em QI de 100. Essa idade é determinada iterativamente, com a ajuda de um programa de computador, disponibilizado juntamente com o teste, para o Escore Total, Escore de Execução e Escore de Raciocínio.

6.3 Resultados da Normatização

Tendo sido finalizado o processo de normatização do SON-R 2½-7[a] para o Brasil, foram solicitadas estatísticas descritivas dos escores normatizados e do QI da amostra, conforme a próxima tabela.

Tabela 6.5 Estatísticas descritivas dos escores normatizados e do QI do SON-R 2½-7[a] (N = 1.200)

Estatísticas	Escores normatizados				QI		
	Mosaicos	Categorias	Situações	Padrões	EE	ER	ET
Mínimo	1	1	1	1	53	54	54
Máximo	19	19	19	19	150	150	150
Média	10,02	10,02	9,99	10	100	100	99,99
Desvio padrão	3,02	2,99	2,99	3,02	14,99	14,99	14,99

Nota: EE = Escala de Execução; ER = Escala de Raciocínio; ET = Escore Total.

As tabelas de normas foram preparadas para cada grupo de idade, partindo de 2;6 a 7;11, a seguir é apresentado um exemplo com as idades 3;3 e 7;9, no apêndice há outras idades.

Tabela 6.6 Normas brasileiras para conversão dos escores brutos em escores normatizados (M=10, DP=3) do SON-R 2½-7[a] nos grupos de idade 3;3 e 7;9

Idade: 3;3		Escore Bruto																
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mos		4	7	10	11	12	13	14	15	17	19	19	19	19	19	19	19	19
Pad		4	6	6	7	8	10	11	13	14	16	17	19	19	19	19	19	19
Cat		7	9	9	10	11	12	13	14	15	16	17	18	19	19	19	19	19
Sit		5	7	8	9	10	11	13	14	15	16	18	19	19	19	19	19	19

Idade: 7;9		Escore Bruto																
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mos		1	1	1	1	1	2	3	5	6	8	10	11	13	14	15	17	17
Pad		1	1	1	1	1	1	1	1	2	4	5	7	8	10	11	13	16
Cat		1	1	1	1	2	3	4	5	6	7	8	9	11	12	14	17	17
Sit		1	1	1	1	2	3	5	6	7	8	10	12	13	15	17	17	17

Nota: Mos=Mosaicos; Pad=Padrões; Cat=Categorias; Sit=Situações.

A tabela de conversão dos escores normatizados em QI é apresentada com alguns exemplos a seguir e no apêndice podem-se verificar outros exemplos.

Tabela 6.7 Normas brasileiras para conversão dos escores normatizados em QI (M=100, DP=15) do SON-R 2½-7[a]

Soma Mos e Pad	QI EE	Soma Cat e Sit	QI ER	Soma Total	QI total	80%-Interv.	Pct.
2	52	2	52	4	50	47-62	1%
3	54	3	55	5	50	47-62	1%
4	57	4	57	6	51	48-63	1%
5	59	5	60	7	52	49-64	1%
6	62	6	62	8	53	50-65	1%
7	65	7	65	9	55	51-66	1%
8	67	8	67	10	56	53-68	1%

Nota: Mos=Mosaicos; Pad=Padrões; Cat=Categorias; Sit=Situações. EE=Escala de Execução; ER=Escala de Raciocínio; 80%-Interv.=Intervalo de 80%; Pct.=Percentil.

7. Características Psicométricas do SON-R 2½-7[a]

A amostra de normatização foi utilizada para efetuar estudos sobre as características psicométricas e a validade do SON-R 2½-7[a]. Considerando os dez grupos de idade, que foram a base inicial para extratificar a amostra, foi calculado o escore bruto médio das crianças. A próxima tabela apresenta os resultados.

Tabela 7.1 Média e variância dos escores brutos por grupo de idade (N = 1.200)

Grupo de Idade	Mosaicos		Categorias		Situações		Padrões	
	Média	Var.	Média	Var.	Média	Var.	Média	Var.
3;3	2,60	3,28	2,98	5,67	3,92	5,61	4,95	4,69
3;9	3,63	5,21	4,21	8,77	4,47	6,20	5,98	4,55
4;3	5,22	5,72	5,87	8,23	5,34	6,50	7,55	3,29
4;9	6,44	4,22	7,43	5,86	6,65	4,58	8,87	3,26
5;3	6,84	4,99	7,69	7,64	7,23	5,54	9,40	4,53
5;9	7,72	5,25	8,63	7,01	7,66	4,80	10,23	4,24
6;3	8,63	3,14	9,39	6,98	8,20	3,64	11,18	3,91
6;9	9,18	2,39	10,09	5,53	8,85	3,24	11,71	3,62
7;3	9,85	3,54	10,71	7,44	9,63	5,26	12,70	4,75
7;9	10,04	3,33	11,40	3,13	9,83	2,88	12,67	3,01

Nota: Var.=variância

Observa-se nessa tabela que o escore bruto médio aumenta com o avanço da idade, conforme esperado, uma vez que as crianças mais velhas devem obter médias superiores às das crianças mais novas. Verifica-se também uma tendência nos grupos de idade de 6;9 e 7;9 à menor variabilidade quando comparados aos demais grupos.

A proporção de acerto dos itens dos subtestes também foi analisada, a tabela a seguir apresenta os resultados.

Tabela 7.2 Proporção de acerto dos itens dos subtestes do SON-R 2½-7[a] (N = 1.200)

Item	Mosaicos	Categorias	Situações	Padrões
1	0,96	0,92	0,97	0,97
2	0,88	0,92	0,95	0,95
3	0,81	0,90	0,91	0,97
4	0,83	0,84	0,88	0,97
5	0,80	0,73	0,68	0,95
6	0,72	0,74	0,69	0,86
7	0,64	0,62	0,48	0,83
8	0,59	0,55	0,61	0,73
9	0,34	0,51	0,46	0,68
10	0,20	0,34	0,25	0,49
11	0,13	0,27	0,13	0,35
12	0,06	0,19	0,10	0,32
13	0,02	0,17	0,05	0,18
14	0,02	0,10	0,03	0,17
15	0,01	0,05	-	0,07
16	-	-	-	0,02

Nessa tabela verifica-se que, em geral, a dificuldade dos itens é crescente. Apenas o subteste Situações apresenta mais variações, mas que não comprometem o procedimento adaptativo.

7.1 Resultados do estudo de validação

Após o estudo de normatização do SON-R 2½-7[a] utilizou-se a amostra para verificar a validade de construto do teste. Nas próximas seções estão apresentados os resultados dos estudos que foram efetuados com o SON-R 2½-7[a] a fim de verificar sua validade. A comprovação da validade dos instrumentos é condição crucial na ciência psicológica, pois se trabalha com o conceito de traço latente e assim, deve-se demonstrar a correspondência entre este e sua representação física (o comportamento). A validade de construto é a forma mais fundamental de validade dos instrumentos psicológicos, uma vez que é a maneira direta de verificar a hipótese da legitimidade da representação comportamental dos traços latentes (Pasquali, 2003).

O objetivo da verificação da validade de construto do SON-R 2½-7[a] não é descobrir o construto a partir de uma representação existente, no caso o teste; mas sim descobrir se a representação (o teste) se constitui uma representação legítima, adequada do construto – nesse estudo a inteligência, definida em termos dos raciocínios abstrato, concreto e espacial (Pasquali, 2003).

A validade de construto de um teste pode ser verificada das seguintes formas: (1) análise da representação comportamental do construto, acessada via análise da consistência interna e análise fatorial; (2) análise por hipótese, que se fundamenta na capacidade do teste para discriminar ou prever um critério externo a ele mesmo, sendo verificada pela validação convergente-discriminante, idade e outros testes do mesmo construto e (3) curva de informação da TRI, operacionalizada via a Teoria de Resposta ao Item (Pasquali, 2003). Nesse estudo foram efetuados esses três tipos de análise, os quais se encontram descritos a seguir.

7.1.1 Análise da consistência interna

A análise da consistência interna, ou fidedignidade, foi feita por meio do SPSS 16.0 tendo como base a intercorrelação entre os itens. O coeficiente lambda 2 de Guttman foi considerado como medida da fidedignidade. Considerou-se também o erro padrão de mensuração (e.p.m.) ao se tratar da fidedignidade, uma vez que em toda medida psicológica é intrínseco o erro.

No presente estudo foram pesquisados 10 grupos de idade, os quais foram considerados no cálculo da fidedignidade. De acordo com Cronbach (1996) o e.p.m. não tende a mudar muito de um grupo para outro – a menos que o teste seja extremamente fácil ou extremamente difícil para o grupo; não obstante, os grupos têm variâncias diferentes, de modo que a fidedignidade provavelmente vai mudar conforme o grupo. Esse fato torna o e.p.m. mais fundamental do que o coeficiente de fidedignidade. Tendo em vista esse fato, além dos

coeficientes de fidedignidade, foi calculado o desvio padrão dos escores brutos, conforme a próxima tabela.

Tabela 7.3 Fidedignidade (lambda 2 de Guttman) e desvio padrão (DP) dos escores brutos

Idade	Fidedignidade							
	Mos	Cat	Sit	Pad	M	EE	ER	QI
3;3	0,75	0,85	0,81	0,81	0,81	0,84	0,89	0,91
3;9	0,80	0,88	0,82	0,80	0,83	0,87	0,91	0,93
4;3	0,84	0,85	0,83	0,78	0,83	0,89	0,90	0,94
4;9	0,79	0,78	0,75	0,79	0,78	0,87	0,84	0,90
5;3	0,82	0,84	0,81	0,83	0,82	0,89	0,88	0,93
5;9	0,82	0,81	0,75	0,80	0,79	0,88	0,86	0,91
6;3	0,76	0,81	0,71	0,80	0,77	0,87	0,84	0,91
6;9	0,68	0,77	0,65	0,77	0,71	0,83	0,81	0,88
7;3	0,80	0,82	0,80	0,80	0,80	0,89	0,88	0,93
7;9	0,77	0,59	0,63	0,71	0,67	0,82	0,74	0,85
M	0,78	0,80	0,76	0,79	0,78	0,86	0,86	0,91

Idade	Desvio padrão							
	Mos	Cat	Sit	Pad	M	EE	ER	QI
3;3	1,81	2,38	2,37	2,16	2,18	3,40	4,26	6,87
3;9	2,28	2,96	2,49	2,13	2,47	3,99	4,99	8,26
4;3	2,39	2,87	2,55	1,81	2,41	3,83	4,90	8,20
4;9	2,05	2,42	2,14	1,81	2,11	3,51	4,01	6,87
5;3	2,23	2,76	2,35	2,13	2,37	4,01	4,56	7,89
5;9	2,29	2,65	2,19	2,06	2,30	3,90	4,26	7,33
6;3	1,77	2,64	1,91	1,98	2,08	3,45	3,99	6,72
6;9	1,54	2,35	1,80	1,90	1,90	3,09	3,65	5,95
7;3	1,88	2,73	2,29	2,18	2,27	3,80	4,58	7,70
7;9	1,83	1,77	1,70	1,74	1,76	3,11	3,00	5,32
M	2,01	2,55	2,18	1,99	2,18	3,61	4,22	7,11

Nota: Mos = Mosáicos, Cat = Categorias, Sit = Situações, Pad = Padrões, M = Média, EE = Escala de Execução, ER = Escala de Raciocínio, QI = Escala Total.

Como pode ser verificado nessa tabela, a variabilidade dos escores brutos das crianças nos cinco grupos de idade é bastante diferenciada, o que influencia no cálculo do coeficiente de fidedignidade. Assim, foi feito um ajuste tendo como base o e.p.m. dos subtestes em cada grupo de idade. A tabela a seguir apresenta os resultados.

Tabela 7.4 Fidedignidade (λ^2 de Guttman) depois do ajuste e erro padrão de mensuração (e.p.m.) dos escores normalizados do SON-R 2½-7[a]

Fidedignidade								
Idade	Mos	Cat	Sit	Pad	M	EE	ER	QI
3;3	0,75	0,90	0,82	0,82	0,82	0,87	0,91	0,93
3;9	0,82	0,86	0,82	0,81	0,83	0,88	0,90	0,93
4;3	0,83	0,84	0,80	0,80	0,82	0,89	0,89	0,93
4;9	0,81	0,82	0,79	0,80	0,81	0,88	0,88	0,93
5;3	0,79	0,81	0,77	0,80	0,79	0,87	0,87	0,92
5;9	0,77	0,80	0,76	0,79	0,78	0,87	0,86	0,92
6;3	0,76	0,79	0,74	0,79	0,77	0,86	0,85	0,91
6;9	0,76	0,77	0,73	0,78	0,76	0,86	0,84	0,91
7;3	0,77	0,76	0,72	0,76	0,75	0,86	0,83	0,91
7;9	0,79	0,75	0,71	0,74	0,75	0,86	0,83	0,90
M	0,78	0,81	0,77	0,79	0,79	0,87	0,87	0,92
Erro padrão de mensuração								
Idade	Mos	Cat	Sit	Pad	M	EE	ER	QI
3;3	1,49	0,96	1,26	1,27	1,25	5,41	4,46	3,88
3;9	1,29	1,11	1,29	1,31	1,25	5,09	4,79	3,86
4;3	1,25	1,20	1,33	1,33	1,28	5,07	5,06	3,95
4;9	1,31	1,26	1,38	1,34	1,32	5,18	5,27	4,08
5;3	1,38	1,31	1,43	1,34	1,37	5,34	5,46	4,22
5;9	1,45	1,35	1,48	1,36	1,41	5,50	5,64	4,35
6;3	1,48	1,38	1,52	1,38	1,44	5,61	5,81	4,46
6;9	1,47	1,42	1,56	1,42	1,47	5,67	5,97	4,54
7;3	1,43	1,46	1,60	1,47	1,49	5,69	6,12	4,61
7;9	1,37	1,50	1,63	1,53	1,51	5,67	6,25	4,66
M	1,39	1,30	1,45	1,38	1,38	5,42	5,48	4,26

Nota: Mos = Mosaicos, Cat = Categorias, Sit = Situações, Pad = Padrões, M = Média, EE = Escala de Execução, ER = Escala de Raciocínio, QI = Escala Total.

Observa-se que a correção efetuada sobre os coeficientes de fidedignidade devido às diferenças na variabilidade dos escores dos cinco grupos aumentou os índices de fidedignidade inferiores a 0,70 e tornou mais confiáveis todos os coeficientes.

Pode-se observar na tabela anterior, Tabela 7.4, que há uma tendência à diminuição do índice de fidedignidade dos subtestes, com exceção de Mosaicos, com o aumento da idade. Também se verifica que o maior e o menor e.p.m. médio estão na Escala de Raciocínio, nos subtestes Categorias e Situações, respectivamente. O subteste Categorias, que apresentou problemas na validação de conteúdo, na versão do SON-R 2½-7[a] apresenta o maior índice médio de fidedignidade entre os subtestes.

Além dos coeficientes de fidedignidade também foi computada a generalizabilidade para cada grupo de idade na Escala de Execução, Raciocínio e Total. Os resultados podem ser vistos na próxima tabela.

Tabela 7.5 Generalizabilidade e erro padrão de estimação para os escores normatizados da Escala de Execução (EE), Escala de Raciocínio (ER) e Escala Total (ET) do SON-R 2½-7[a]

Idade	Generalizabilidade			Erro Padrão de Estimação		
	EE	ER	ET	EE	ER	ET
3;3	0,77	0,73	0,83	7,10	7,80	6,30
3;9	0,77	0,73	0,83	7,10	7,80	6,30
4;3	0,77	0,73	0,83	7,10	7,80	6,30
4;9	0,77	0,73	0,83	7,10	7,80	6,30
5;3	0,77	0,73	0,83	7,10	7,80	6,30
5;9	0,77	0,73	0,83	7,10	7,80	6,30
6;3	0,77	0,73	0,83	7,10	7,80	6,30
6;9	0,77	0,73	0,83	7,10	7,80	6,30
7;3	0,77	0,73	0,83	7,10	7,80	6,30
7;9	0,77	0,73	0,83	7,10	7,80	6,30

Conforme observado na tabela anterior, a generalizabilidade é fixa em todos os grupos de idade. Isso ocorre porque não houve relação clara entre as correlações (covariâncias) com a idade.

7.1.2 Análise Fatorial

7.1.2.1 Análise Fatorial de Informação Plena - FIFA

Inicialmente foi efetuada uma análise fatorial com todos os itens do SON-R 2½-7[a] a fim de verificar sua estrutura fatorial com um e dois fatores. Como a escala de resposta do teste é dicotômica, ou seja, é composta por apenas acertos (1) e erros (0), a análise utilizada foi a Análise Fatorial de Informação Plena (*Full Information Factor Analysis- FIFA*), que é mais adequada para esse tipo de dado. Para implementar essa análise foi utilizado o pacote estatístico Testfact 4.0, que executa os principais procedimentos de análise clássica do item, bem como métodos modernos de análise fatorial baseados na Teoria de Resposta ao Item (TRI) (ver Wilson, Wood & Gibons, 1991).

A FIFA utiliza o método da Máxima Verossimilhança Marginal (MML) para estimar as cargas fatoriais dos itens, bem como o algoritmo EM (*Expectation Maximization*) (Wilson, Wood & Gibons, 1991). Entre as vantagens dessa análise estão: (1) ela considera toda a informação da matriz de dados do teste e (2) provê alternativa para contornar os problemas da matriz não positiva-definida (Laros, Pasquali & Rodrigues, 2000).

Os resultados da FIFA realizada com os dados do SON-R 2½-7[a] podem ser vistos em duas estruturas: com um fator (ver Tabela 7.6) e dois fatores: Execução e Raciocínio (ver Tabela 7.7). O objetivo dessa análise é embasar a interpretação da escala com um e dois fatores.

Tabela 7.6 Cargas fatoriais dos itens do SON-R 2½-7[a] no fator único e a carga média por subtteste (N=1.200)

Item	Carga	Item	Carga	Item	Carga	Item	Carga
Mos1	0,46	Cat1	0,40	Sit1	0,43	Pad1	0,49
Mos2	0,50	Cat2	0,41	Sit 2	0,43	Pad 2	0,50
Mos3	0,50	Cat3	0,48	Sit 3	0,50	Pad 3	0,50
Mos4	0,42	Cat4	0,48	Sit 4	0,49	Pad 4	0,49
Mos5	0,47	Cat5	0,50	Sit 5	0,50	Pad 5	0,50
Mos6	0,45	Cat6	0,48	Sit 6	0,50	Pad 6	0,47
Mos7	0,42	Cat7	0,50	Sit 7	0,50	Pad 7	0,50
Mos8	0,40	Cat8	0,48	Sit 8	0,50	Pad 8	0,45
Mos9	0,43	Cat9	0,49	Sit 9	0,50	Pad 9	0,38
Mos10	0,42	Cat10	0,50	Sit 10	0,50	Pad 10	0,44
Mos11	0,42	Cat11	0,47	Sit 11	0,49	Pad 11	0,43
Mos12	0,45	Cat12	0,49	Sit 12	0,50	Pad 12	0,44
Mos13	0,47	Cat13	0,50	Sit 13	0,50	Pad 13	0,45
Mos14	0,36	Cat14	0,50	Sit 14	0,44	Pad 14	0,45
Mos15	0,33	Cat15	0,50			Pad 15	0,41
						Pad 16	0,32
Média	0,43		0,48		0,48		0,45

Nota: Análise fatorial de informação plena (*Full Information Factor Analysis - FIFA*).

Como pode ser observado, todos os 60 itens saturam no fator único com cargas fatoriais iguais ou superiores a 0,32, que corresponde a cerca de 10% de variância comum com o fator. De acordo com Tabachnick e Fidell (1996) esse valor é satisfatório. Também foi verificado o valor médio das cargas fatoriais por subtteste e Mosaicos e Padrões ficaram ao redor de 0,40; enquanto Categorias e Situações aproximadamente 0,50. Tais resultados dão suporte à utilização do SON-R 2½-7[a] com uma Escala Total, uma vez que existe um fator geral. Também se verificou a estrutura com dois fatores: Execução e Raciocínio, com rotação Promax. Os resultados são apresentados na Tabela 7.7.

Tabela 7.7 Cargas fatoriais dos itens do SON-R 2½-7[a] nos dois fatores e a carga média por fator (N=1.200)

Escala de Execução				Escala de Raciocínio			
Item	Carga	Item	Carga	Item	Carga	Item	Carga
Mos1	0,31	Pad1	0,42	Cat1	0,29	Sit1	0,32
Mos2	0,43	Pad2	0,43	Cat 2	0,27	Sit 2	0,33
Mos 3	0,36	Pad 3	0,38	Cat 3	0,39	Sit 3	0,36
Mos 4	0,30	Pad 4	0,36	Cat 4	0,35	Sit 4	0,42
Mos 5	0,35	Pad 5	0,56	Cat 5	0,39	Sit 5	0,38
Mos 6	0,33	Pad 6	0,39	Cat 6	0,38	Sit 6	0,43
Mos 7	0,29	Pad 7	0,48	Cat 7	0,37	Sit 7	0,42
Mos 8	0,28	Pad 8	0,35	Cat 8	0,38	Sit 8	0,37
Mos 9	0,26	Pad 9	0,26	Cat 9	0,38	Sit 9	0,46
Mos 10	0,30	Pad 10	0,33	Cat 10	0,44	Sit 10	0,40
Mos 11	0,25	Pad 11	0,31	Cat 11	0,37	Sit 11	0,43
Mos 12	0,34	Pad 12	0,33	Cat 12	0,40	Sit 12	0,39
Mos 13	0,35	Pad 13	0,36	Cat 13	0,43	Sit 13	0,40
Mos 14	0,24	Pad 14	0,36	Cat 14	0,47	Sit 14	0,31
Mos 15	0,21	Pad 15	0,28	Cat 15	0,45		
		Pad 16	0,21				
Média		0,34				0,38	

Nota: Análise fatorial de informação plena (*Full Information Factor Analysis- FIFIA*).

Conforme observado nessa tabela, em geral os dois fatores apresentam cargas fatoriais satisfatórias, a maioria dos itens possui carga fatorial superior a 0,32. Entre os itens com carga fatorial baixa, alguns são os últimos dos subtestes Mosaicos e Padrões, dessa forma, devem ter sido influenciados pela pouca variabilidade de resposta. A correlação encontrada entre os dois fatores foi alta, 0,80. Frente aos resultados, considera-se que a estrutura fatorial com dois fatores também se mostra válida.

7.1.2.2 Análise Simultânea dos Componentes Principais - SCA

No processo de validação do SON-R 2½-7 na Holanda foi efetuada uma análise fatorial com os escores normatizados dos subtestes (M=10 e DP=3) utilizando a Análise Simultânea dos Componentes Principais (SCA) (Millsap & Meredith, 1988; Kiers & Ten Berge, 1989). A análise que foi efetuada com os dados brasileiros teve como objetivo verificar se a estrutura interna do teste no Brasil também é similar em três grupos

diferenciados de idade, a saber: (1) 3;3 a 4;3; (2) 4;9 a 6;3 e (3) 6;9 a 7;9. A SCA baseia-se nas três matrizes de correlação dos escores normatizados dos subtestes. Por meio dessa análise foi testado se uma solução uniforme com pesos otimizados dos componentes nos três grupos explica significativamente menos da variância total do que os componentes que são otimizados nos grupos separados (Millsap & Meridith, 1988; Niesing, 1997). Os resultados indicaram que, com a extração de dois componentes, a solução uniforme foi ligeiramente menos adequada (80,1% vs. 80,2% da variância total explicada nos grupos separados). Assim, as características estruturais do SON-R 2½-7[a] demonstraram ser altamente independentes das características examinadas dos três grupos de idade. Outro resultado da análise SCA é que o uso de pesos simples (1 e 0) é quase tão bom quanto pesos otimizados (79,92% vs. 80,05%).

Não obstante a boa consistência da estrutura interna do SON-R 2½-7[a] nos diferentes grupos, pode-se observar que o segundo componente antes da rotação é relativamente menos importante: o autovalor do segundo componente foi cerca de 0,60. Entretanto, o autovalor do segundo componente tende a aumentar com o aumento da idade, conforme Tabela 7.8. É uma indicação de que a solução com dois fatores ficou ligeiramente mais adequada para o grupo de idade mais velho. Provavelmente, a precisão da diferença entre o fator espacial e o de raciocínio não é tão alta. Portanto, conclusões importantes não podem ser baseadas nessas diferenças. Mais detalhes da análise com o programa SCA (Kiers, 1990) podem ser vistos na Tabela 7.8.

Tabela 7.8 Autovalor e percentual de variância explicada por componente e grupo de idade dos dois componentes principais do SON-R 2½-7[a] via análise SCA

Componente	3;3-4;3anos (N=360)		4;9-6;3 anos (N=480)		6;9-7;9 anos (N=360)	
	Autov.	% var.	Autov.	% var.	Autov.	% var.
Componente 1	2,69	67,3	2,60	65,0	2,55	63,6
Componente 2	0,53	13,3	0,61	15,1	0,65	16,3

3;3 a 7;9 anos

Cargas fatoriais

	Componente 1	Componente 2
Mos	0,11	0,63
Cat	0,65	0,05
Sit	0,75	-0,06
Pad	-0,08	0,78

Nota: Autov.= Autovalor; Var.= variância. Componente 1=Raciocínio; Componente 2=Execução.

Além do percentual de variância explicada dos dois fatores nos três grupos de idade a tabela também apresenta as cargas fatoriais dos dois fatores no grupo total de idade. O resultado demonstra cargas expressivas em ambos os fatores. Dessa forma, os resultados apresentados dão suporte à utilização e interpretação do SON-R 2½-7[a] por meio de duas escalas em diferentes grupos de idade.

7.1.2.3 Análise Fatorial Confirmatória - CFA

A Análise Fatorial Confirmatória (CFA) é um procedimento estatístico viabilizado pelos Modelos de Equações Estruturais (Byrne, 2001; Kline, 1998; Pilati & Laros, 2007; Ullman, 2007). Devido ao seu caráter confirmatório, a CFA não é uma técnica adequada para efetuar explorações sobre a estrutura de um instrumento, mas confirmar uma estrutura teórica previamente estabelecida. Nos estudos internacionais (ver Tellegen & Cols., 1998) realizados com o SON-R 2½-7 é proposta uma estrutura com dois fatores e um fator geral.

Inicialmente, foi verificada se a estrutura interna do SON-R 2½-7[a] se confirmava com dois fatores: Execução e Raciocínio. Também buscou-se verificar a invariância dessa estrutura em quatro grupos diferentes de idade, a saber: 3;3 a 7;9; 3;3 a 4;3; 4;9 a 6;3 e 6;9 a 7;9. A CFA realizada nesse estudo foi executada por meio do pacote estatístico Lisrel 8.80, tendo como método de estimação ML (máxima verossimilhança). A tabela a seguir apresenta um resumo dos resultados.

Tabela 7.9 Coeficientes de determinação (cargas fatoriais), erro padrão e correlação múltipla ao quadrado (SMC) da AFC do SON-R 2½-7[a].

Subteste	3;3-7;9			3;3-4;3			4;9-6;3			6;9-7;9		
	F1	F2	SMC	F1	F2	SMC	F1	F2	SMC	F1	F2	SMC
Mosaicos	0,83 (0,03)		0,69	0,79 (0,05)		0,63	0,85 (0,04)		0,73	0,83 (0,05)		0,69
Padrões	0,75 (0,03)		0,56	0,73 (0,05)		0,53	0,76 (0,04)		0,58	0,75 (0,05)		0,56
Categorias		0,78 (0,03)	0,61		0,85 (0,05)	0,72		0,77 (0,05)	0,60		0,71 (0,05)	0,51
Situações		0,73 (0,03)	0,53		0,75 (0,05)	0,57		0,68 (0,05)	0,46		0,79 (0,05)	0,62

Nota: Em parênteses é apresentado o erro padrão. F1=Execução; F2=Raciocínio.

Como pode ser observado na Tabela 7.9, todas as cargas fatoriais são moderadas ou altas e o erro associado é bastante pequeno. Os índices de correlação múltipla ao quadrado (SMC) também indicam resultados muito satisfatórios. Foram verificadas também as estatísticas sobre o ajuste do modelo nos quatro grupos de idade. Byrne (2001) aponta alguns índices importantes para o ajuste do modelo, como o GFI (*Goodness of Fit Index*), que é uma medida da quantidade relativa de (co)variância na matriz de covariância observada que pode ser predita pelo modelo e o AGFI (*Adjusted Goodness of Fit Index*), que é semelhante ao GFI, mas com um ajuste pelo número de graus de liberdade relativo ao número de variáveis. Ambos possuem como valores desejáveis índices superiores a 0,90.

Ademais, faz parte das estatísticas de ajuste o RMSEA (*Root Mean Square Error of Approximation*), descrito como a discrepância por grau de liberdade – ou seja, a falta de ajuste dado um número particular de parâmetros no modelo testado. Esse índice foi reconhecido como um dos critérios mais informativos na estrutura do modelo de covariância. O RMSEA considera o erro de aproximação para a população e questiona o seguinte: “quão bem o modelo se ajusta à matriz de covariância da população, se ela estivesse disponível?”. Valores inferiores a 0,05 para RMSEA indicam um bom ajuste (Byrne, 2001; Kline, 1998; Pilati & Laros, 2007).

No modelo testado para os quatro grupos de idade verificou-se que tanto o GFI quanto o AGFI obtiveram valores superiores a 0,95, indicando um ajuste muito bom do modelo. Quanto ao RMSEA, com exceção do grupo com idade entre 3 anos e 3 meses e 4 anos e 3 meses, que obteve um índice de 0,07, os demais grupos apresentaram valores inferiores a 0,05. Esses resultados indicam que a estrutura do SON-R 2½-7[a] com dois fatores é robusta e foi confirmada nos quatro grupos de idade.

7.1.2.4 Análise Fatorial de Segunda Ordem

Além da AFC com dois fatores, também foi verificada a estrutura do SON-R 2½-7[a] com um fator de segunda ordem. Thompson (2005) argumenta que, quando os fatores estão correlacionados, pode-se extrair mais fatores da matriz. Esse autor assinala que os fatores extraídos das correlações entre as variáveis (ou outras estatísticas mensurando associações) são chamados fatores de primeira ordem. Por sua vez, os fatores extraídos das correlações entre os fatores de primeira ordem são chamados fatores de segunda ordem. No presente estudo foi verificada a estrutura do SON-R 2½-7[a] com dois fatores de primeira ordem: Execução e Raciocínio e um fator geral de segunda ordem, ou fator geral. Para tanto, inicialmente foi solicitada uma análise dos componentes principais (PC) com rotação *Promax* dos fatores. Utilizou-se a análise PC porque, segundo Thompson (2005), somente nessa

análise a correlação entre os fatores é igual à correlação entre os escores fatoriais. Os fatores de segunda ordem obtidos podem ser vistos na Tabela 7.10, a seguir.

Tabela 7.10 Coeficientes de determinação (cargas fatoriais) do SON-R 2½-7[a] nos fatores de primeira e segunda ordem.

Subteste	Segunda ordem	Primeira ordem	
		Fator I	Fator II
Mosaicos	0,84	0,79	0,14
Categorias	0,81	0,11	0,80
Situações	0,79	-0,07	0,94
Padrões	0,80	0,97	-0,08

Nota: Análise dos componentes principais com rotação *Promax*.

Conforme observado nessa tabela, tanto as cargas no fator de segunda ordem quanto nos fatores de primeira ordem são bastante elevadas. De acordo com Thompson (2005) a interpretação de fatores de ordens superiores traz consigo algumas dificuldades. Uma delas é que os fatores são abstrações das variáveis mensuradas. Por sua vez, os fatores de segunda ordem são abstrações de abstrações, igualmente obtidos por meio das variáveis mensuradas. Esse autor indica que, de alguma forma, se quer interpretar os fatores de segunda ordem em termos das variáveis observadas, ao invés de considerá-los como manifestação dos fatores dessas variáveis.

Nesse contexto, Thompson (2005) propõe a verificação do fator de segunda ordem por meio da solução Schmid-Leiman. Nessa solução é proposto um método elegante para expressar os fatores de primeira e segunda ordem em termos das variáveis mensuradas, removendo toda a variância nos fatores de primeira ordem presente nos fatores de segunda ordem. Esse procedimento permite ao pesquisador determinar qual variância, se há alguma, é única em um dado nível de análise ou perspectiva. Assim, a solução Schmid-Leiman não apenas preserva as características desejáveis de interpretação da solução oblíqua, mas também revela a estrutura hierárquica das variáveis. A Tabela 7.11 apresenta a solução Schmid-Leiman para o SON-R 2½-7[a].

Tabela 7.11 Solução Schmid-Leiman para o SON-R 2½-7[a] nos fatores de primeira e segunda ordem.

Subteste	Segunda ordem	Primeira ordem	
		Fator I	Fator II
Mosaicos	0,84	0,35	0,06
Categorias	0,81	0,05	0,35
Situações	0,79	-0,03	0,42
Padrões	0,80	0,43	-0,04

Como pode ser observado nessa tabela, as cargas nos fatores de primeira ordem são inferiores às encontradas na solução anterior apresentada na Tabela 7.10. Isso ocorre porque a variância aqui não é compartilhada entre os fatores, ou seja, é única. Verifica-se que Padrões e Situações apresentaram as maiores cargas. Não obstante a magnitude das cargas fatoriais, a estrutura hierárquica com um fator de segunda ordem e dois de primeira para o SON-R 2½-7[a] se sustenta na solução Schmid-Leiman.

7.1.3 Análise dos itens pela Teoria de Resposta ao Item

Os estudos no âmbito da inteligência realizados nas primeiras décadas do século 20 bem como as controvérsias que os acompanharam levaram ao desenvolvimento da Teoria Clássica dos Testes (TCT). De acordo com Baker e Kim (2004) muitos construtos hoje familiares tais como escore verdadeiro, fidedignidade e validade surgiram do trabalho de Spearman ao desenvolver um modelo matemático para sua teoria de inteligência. Esses autores apontam que a teoria estatística dominante da época baseava-se em conceitos correlacionais. Assim, o elemento básico nessa teoria era o escore no teste; os itens e suas características desempenhavam um papel irrelevante na estrutura da teoria. Ao longo dos anos tanto psicometristas quanto estudiosos do campo prático levantaram discussões postulando que seria mais razoável que uma teoria sobre testes começasse com as características dos itens que compõem um teste do que com o escore resultante.

A teoria dos testes baseada em itens tem suas origens no trabalho de Binet e Simon no início do século 20. Esses pesquisadores utilizaram uma apresentação tabular da relação funcional entre a proporção de resposta correta ao item e a idade cronológica. Por meio desse procedimento eles visavam alocar os itens no teste de inteligência que eles elaboraram. Utilizando essa mesma abordagem, outros pesquisadores plotaram curvas relacionando duas variáveis. Ou seja, eles estavam utilizando curvas características dos itens. Assim, por muitos anos a abordagem da curva característica foi considerada simplesmente como uma técnica alternativa de análise de item (Baker & Kim, 2004).

O marco inicial da teoria de testes baseada em itens foi com o trabalho de Lawley (1943). Em um artigo pioneiro na área o autor mostrou como obter estimativas de máxima verossimilhança dos parâmetros da curva característica do item, definiu o escore verdadeiro em termos dos itens de um teste e mostrou que o coeficiente de fidedignidade clássico pode também ser expresso como uma função desses parâmetros do item. Lord (1952) aprimorou o trabalho de Lawley, mostrando que uma grande quantidade de construtos adicionais da TCT pode ser expressa como funções dos parâmetros das curvas características dos itens do teste. O trabalho desses dois autores estabeleceu os conceitos básicos da teoria psicométrica baseada em itens, agora conhecida como Teoria de Resposta ao Item (TRI) (Baker & Kim, 2004).

Existem muitas possibilidades de modelos de resposta ao item, diferindo na forma matemática da função característica do item e/ou no número de parâmetros especificados no modelo. Todos os modelos de TRI contêm um ou mais parâmetros descrevendo o item e um ou mais parâmetros para descrever o examinando. Assim, o passo inicial em qualquer aplicação da TRI é estimar esses parâmetros (Hambleton, Swaminathan & Rogers, 1991). No presente estudo, foi utilizado o modelo de dois parâmetros.

Modelo de Dois Parâmetros

No modelo de dois parâmetros são consideradas duas informações acerca do item, a dificuldade e a discriminação, como pode ser visto na expressão a seguir:

$$P_j(\theta) = \frac{1}{1 + \exp[-Da_j(\theta - b_j)]}$$

$P_j(\theta)$: Probabilidade de um indivíduo com habilidade θ responder o item j corretamente

θ : Habilidade ou traço latente

b : Parâmetro de dificuldade do item j

a : Parâmetro de discriminação do item

D : Constante que pode assumir os valores: 1,7 quando se está ajustando o modelo logístico, ou 1 quando se está ajustando o modelo da ogiva normal.

De acordo com Crocker e Algina (1986), o modelo de um parâmetro pode ser visto como um caso especial do modelo de dois parâmetros, pois apresenta um parâmetro a menos. Na sequência é apresentado esse modelo.

Modelo de Um Parâmetro

No presente modelo todos os itens têm a mesma discriminação. A probabilidade de resposta é determinada apenas pela diferença entre o nível do traço do indivíduo e a dificuldade do item. O modelo de um parâmetro pode ser descrito da seguinte forma:

$$P_j(\theta) = \frac{1}{1 + \exp[-(\theta - b_j)]}$$

No presente estudo foi ajustado o modelo normal de 2 parâmetros para a análise dos 60 itens do SON-R 2½-7[a]. Buscou-se analisar a discriminação dos itens do teste a fim de

verificar sua qualidade psicométrica e a dificuldade com o intuito de verificar se o grau de dificuldade dos itens é crescente (ver Tabela 7.12).

Tabela 7.12 Parâmetros a e b dos 60 itens do SON-R 2½-7[a] (N = 1.200)

Item	a	b	Item	a	b
Mos1	0,88	-2,84	Sit1	2,01	-2,23
Mos2	1,27	-1,58	Sit2	2,00	-1,98
Mos3	1,12	-1,21	Sit3	1,44	-1,72
Mos4	2,20	-1,12	Sit4	1,41	-1,55
Mos5	1,86	-1,03	Sit5	1,15	-0,62
Mos6	2,02	-0,67	Sit6	1,02	-0,71
Mos7	2,12	-0,37	Sit7	0,98	0,12
Mos8	2,24	-0,20	Sit8	1,33	-0,32
Mos9	1,97	0,55	Sit9	0,94	0,21
Mos10	1,95	0,99	Sit10	1,04	0,98
Mos11	2,17	1,25	Sit11	0,98	1,65
Mos12	1,91	1,74	Sit12	1,30	1,64
Mos13	1,77	2,16	Sit13	1,40	1,99
Mos14	2,59	2,04	Sit14	1,98	2,05
Mos15	2,57	2,26	Pad1	1,58	-2,38
Cat1	2,29	-1,70	Pad2	1,46	-2,20
Cat2	2,48	-1,62	Pad3	1,36	-2,58
Cat3	1,78	-1,61	Pad4	1,52	-2,49
Cat4	1,71	-1,25	Pad5	1,17	-2,29
Cat5	1,15	-0,80	Pad6	1,82	-1,34
Cat6	1,48	-0,81	Pad7	1,43	-1,24
Cat7	1,13	-0,38	Pad8	2,06	-0,70
Cat8	1,48	-0,09	Pad9	2,68	-0,47
Cat9	1,50	0,04	Pad10	1,95	0,10
Cat10	1,21	0,59	Pad11	1,94	0,51
Cat11	1,52	0,79	Pad12	1,86	0,62
Cat12	1,41	1,14	Pad13	1,83	1,11
Cat13	1,26	1,27	Pad14	1,77	1,14
Cat14	1,23	1,66	Pad15	2,22	1,56
Cat15	1,25	2,14	Pad16	3,01	1,96

Frente ao exposto nessa tabela quanto à discriminação, parâmetro a , observa-se que quase todos os itens possuem índices satisfatórios. Muitos itens apresentam discriminação superior a 2,0 e o último item do subteste Padrões possui discriminação igual a 3,0, que é um valor bastante elevado para esse parâmetro. Todavia, considerando o fato de que apenas poucos sujeitos da amostra alcançam e acertam esse item, hipotetiza-se que problemas como

falta de informação tenham atrapalhado a estimação da TRI. Ademais, a pequena variabilidade de resposta em alguns itens influencia a estimativa da discriminação.

No que se refere à dificuldade, verifica-se que alguns itens possuem dificuldade inferior ou superior ao esperado para eles. Não obstante, para efetuar mudanças na ordem dos itens é necessário que se façam estudos mais aprofundados. Esse fato já ocorreu em outros estudos com os testes SON e, ao se mudar a ordem dos itens indicados, em outros estudos o mesmo ocorria com outros itens. Dessa forma, é necessário cautela quanto a esse aspecto. Ademais, no procedimento adaptativo os itens de entrada são: 1, 3 ou 5 e o critério de parada são 3 erros em todo o subteste, ou 2 consecutivos na segunda parte dos subtestes de Execução. Dessa forma, as diferenças encontradas na dificuldade não influenciam no resultado do teste.

7.2 Análise por hipótese

Juntamente com os coeficientes de correlação estão os intervalos de confiança. Miles e Shevlin (2001) assinalam que é útil verificar se um coeficiente de correlação é estatisticamente significativo. Não obstante, um teste de significância diz apenas a probabilidade deste ser ou não diferente de zero. Uma correlação é apenas outro meio de expressar a inclinação. O valor da probabilidade para uma correlação apenas indica que a correlação é maior ou igual a zero – nunca deve ser interpretado como um indicativo da força da associação, pois uma correlação pequena, por exemplo 0,10, poderá ser estatisticamente significativa (0,05 de significância) com uma amostra de 400 sujeitos, mesmo que uma correlação desse tamanho não possua qualquer significância prática.

Por sua vez, Rosnow e Rosenthal (1996) sugerem que sejam construídos intervalos de confiança de 95% para cada coeficiente de correlação obtido. Assim, no presente estudo pretende-se, por meio desse procedimento, verificar com esse nível de confiança, em qual intervalo encontra-se o verdadeiro coeficiente de correlação populacional.

Idade

A idade foi utilizada como um dos critérios para a validação de construto do SON-R 2½-7[a], pois a inteligência está intrinsecamente ligada a mudanças no desenvolvimento cognitivo dos indivíduos. Esperava-se encontrar correlações positivas e substanciais entre os escores brutos nos subtestes e a idade, pois crianças mais velhas devem ter maiores escores no teste do que crianças mais novas. A Tabela 7.13 apresenta os resultados.

Tabela 7.13 Coeficientes de correlação entre a idade e o escore bruto nos subtestes do SON-R 2½-7[a] (N = 1200)

		Mosaicos	Categorias	Situações	Padrões
Idade (meses)	<i>r</i>	0,76	0,71	0,66	0,78
	IC 95%	0,74-0,78	0,68-0,74	0,63-0,69	0,76-0,80

Como pode ser observado nessa tabela há correlação positiva e alta entre a idade e o escore bruto nos subtestes do SON-R 2½-7[a]. Os intervalos de confiança mostram que o menor limite inferior é no subteste Situações e o maior limite superior é no subteste Padrões. Tal resultado dá suporte à validade do SON-R 2½-7[a].

Estudo de validade convergente

Com o objetivo de verificar a validade convergente do SON-R 2½-7[a] foram realizadas análises de correlação *r* de Pearson entre os três tipos de QI do SON-R 2½-7[a] calculados com as normas brasileiras e outros testes de inteligência. Dessa forma, foram feitos três estudos de validação com outros testes de inteligência. Inicialmente, verificou-se a correlação entre as três escalas do SON-R 2½-7[a]: Total, Execução e Raciocínio e as escalas do WPPSI-III, avaliadas segundo as normas estadunidenses. Os resultados podem ser vistos na Tabela 7.14.

Tabela 7.14 Coeficientes de correlação entre as escalas do SON-R 2½-7[a] e do WPPSI-III e o intervalo de confiança de 95% das correlações (N = 49)

		SON QI Total	SON QI EE	SON QI ER
WPPSI-III FSIQ	<i>r</i>	0,75	0,59	0,78
	IC 95%	0,60-0,85	0,37-0,75	0,64-0,87
WPPSI-III VIQ	<i>r</i>	0,66	0,48	0,73
	IC 95%	0,47-0,80	0,23-0,67	0,57-0,84
WPPSI-III PIQ	<i>r</i>	0,73	0,61	0,70
	IC 95%	0,57-0,84	0,40-0,76	0,52-0,82
WPPSI-PSIQ	<i>r</i>	0,70	0,56	0,69
	IC 95%	0,52-0,82	0,33-0,73	0,51-0,82

Nota: FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; PSIQ = Processing Speed IQ.

Conforme observado nessa tabela, todas as correlações entre as escalas do SON-R 2½-7[a] e do WPPSI-III são altas, o que aponta para a validade convergente do SON-R 2½-7[a]. Os intervalos de confiança mostram que as menores correlações possuem intervalos menos precisos. Pode-se verificar também que o menor coeficiente de correlação observado é com o QI verbal do WPPSI-III e o maior com a escala total. Tal resultado é consistente com os encontrados na Austrália entre o SON-R 2½-7 e o WPPSI-R (Tellegen & Cols., 1998). A correlação na Austrália com o WPPSI-R foi $r = 0,75$ para a FSIQ, $r = 0,54$ para a VIQ e $r = 0,73$ para a PIQ, similar à relação forte entre o WPPSI-III e o SON-R 2½-7[a] encontrada no Brasil. Tal resultado corrobora o fato de que o SON-R 2½-7[a] é um teste que mede uma área ampla de inteligência, assim como o WPPSI-III. Este último afere um construto mais amplo de inteligência devido à parte verbal que está ausente no SON-R 2½-7[a].

A validade convergente do SON-R 2½-7[a] também foi verificada por meio da sua correlação com o WISC-III. A Tabela 7.15 apresenta os resultados.

Tabela 7.15 Coeficientes de correlação entre as escalas do SON-R 2½-7[a] e do WISC-III e o intervalo de confiança de 95% das correlações (N = 50)

		SON QI Total	SON QI EE	SON QI ER
WISC-III QI Total	<i>r</i>	0,67	0,56	0,59
	IC 95%	0,48-0,80	0,34-0,73	0,38-0,75
WISC-III QI Verbal	<i>r</i>	0,51	0,36	0,51
	IC 95%	0,27-0,69	0,09-0,58	0,44-0,78
WISC-III QI de Execução	<i>r</i>	0,64	0,53	0,56
	IC 95%	0,30-0,71	0,34-0,73	0,30-0,71
WISC-III QI de Compreensão Verbal	<i>r</i>	0,53	0,39	0,50
	IC 95%	0,30-0,71	0,13-0,61	0,26-0,69
WISC-III QI de Organização perceptual	<i>r</i>	0,68	0,61	0,54
	IC 95%	0,50-0,81	0,22-0,76	0,31-0,71

As correlações entre as escalas do SON-R 2½-7[a] e do WISC-III são moderadas ou altas, a menor correlação observada ($r=0,36$) é entre o QI da Escala de Execução do SON-R 2½-7[a] e a Escala Verbal do WISC-III. Os intervalos de confiança mostram que as correlações inferiores a 0,40 são pouco confiáveis.

Considerando as correlações do SON com a parte verbal do WISC-III pode-se inferir que estas foram as menores devido ao fato de que a parte verbal do WISC-III é muito sensível à escolarização. Por sua vez, as correlações com o WPPSI-III foram melhores, pois este é menos verbal que o WISC-III. Ademais, a correlação com o WISC-III é inferior à encontrada com o WPPSI-III devido, parcialmente, aos problemas na normatização brasileira do WISC-III. A amostra de normatização desse teste foi exclusivamente de crianças de uma cidade do interior do Rio Grande do Sul. Tal fato prejudicou a representatividade das normas para o Brasil.

Por fim, a validade convergente do SON-R 2½-7[a] foi testada por meio da sua correlação com as Matrizes Progressivas Coloridas de Raven (MPC) e com o Columbia (EMMC), conforme observado na Tabela 7.16.

Tabela 7.16 Coeficientes de correlação entre as escalas do SON-R 2½-7[a], as MPC de Raven e o EMMC (Colúmbia) e o intervalo de confiança de 95% das correlações (N = 120)

TESTES		SON QI Total	SON QI EE	SON QI ER
Raven	<i>r</i>	0,68	0,61	0,56
	IC 95%	0,57-0,80	0,49-0,71	0,43-0,67
Columbia	<i>r</i>	0,56	0,43	0,54
	IC 95%	0,43-0,67	0,27-0,57	0,40-0,66

Verifica-se nessa tabela que todas as correlações são moderadas ou altas, indicando a validade convergente do SON-R 2½-7[a]. As maiores correlações são com a escala completa e, no caso do Columbia, também com a Escala de Raciocínio. Considerando a natureza desses testes, é razoável esperar correlações mais altas com a Escala de Raciocínio.

8. Relação do SON-R 2 ½ -7[a] com variáveis externas

O objetivo desse capítulo é apresentar a relação do teste com outras variáveis obtidas na pesquisa de normatização, tais como: nível socioeconômico, duração do teste, influência do examinador, unidade da federação, sexo e a avaliação que os examinadores realizaram das crianças durante a testagem.

8.1 Nível socioeconômico (NSE)

Além da aplicação do SON-R 2½-7[a] foram solicitadas aos pais ou responsáveis pelas crianças informações acerca do contexto socioeconômico da família. Sobre a escolaridade foram encontrados os resultados apresentados na tabela seguinte.

Tabela 8.1 Grau de escolaridade dos pais ou responsáveis

Escolaridade do pai	Frequência	%
Analfabeto	31	2,6
1º grau incompleto	263	21,9
1ª grau completo	102	8,5
2º grau incompleto	39	3,3
2º grau completo	257	21,4
Superior incompleto	14	1,2
Superior Completo	107	8,9
Pós-graduação	4	0,3
Em branco	383	31,9
Escolaridade da mãe		
Analfabeta	26	2,2
1º grau incompleto	273	22,8
1ª grau completo	118	9,8
2º grau incompleto	65	5,4
2º grau completo	250	20,8
Superior incompleto	39	3,3
Superior Completo	125	10,4
Pós-graduação	3	0,3
Em branco	301	25,1

Os demais itens pesquisados, por meio de um questionário sobre a quantidade de bens presentes na residência, foram submetidos a uma análise dos componentes principais. Os

resultados indicaram um KMO superior a 0,80 e 48,3% de variância explicada para uma estrutura com um fator. As cargas fatoriais podem ser observadas na Tabela 8.2.

Tabela 8.2 Cargas fatoriais no fator nível socioeconômico (NSE)

Possui na residência	Carga fatorial
Automóveis	0,81
Computador	0,77
Banheiros	0,74
TV	0,72
Empregadas domésticas	0,67
Lavadora de roupas	0,58
Geladeira	0,53

Após a análise dos componentes principais foi calculado o escore socioeconômico de cada criança, obtido por meio da média dos itens respondidos. Esse procedimento foi efetuado a fim possibilitar a verificação do perfil socioeconômico da amostra. O escore variou de 0 a 3,43 e a média foi 0,81 (DP = 0,45). Devido à grande variabilidade de escores, estes foram transformados em postos percentílicos. O gráfico a seguir apresenta o resultado.

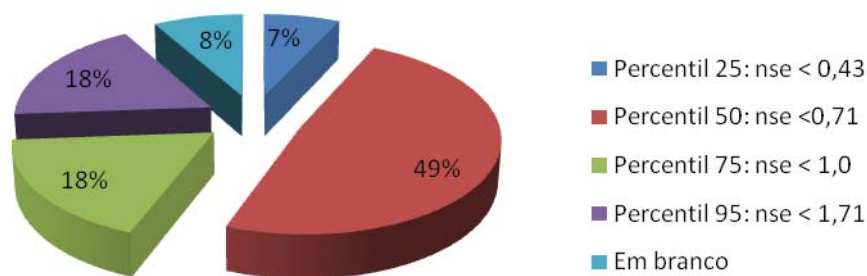


Figura 8.1. Distribuição da amostra segundo o nível socioeconômico

Considerando os escores normatizados obtidos pelas crianças no teste, foram verificadas algumas estatísticas acerca das crianças em cada posto percentílico, conforme a tabela a seguir.

Tabela 8.3 Comparação dos escores normatizados nos postos percentílicos do nse (N = 1.101*)

Postos	%	EE		IC 95%		ER		IC 95%		SON		IC 95%	
		M	DP	Inf.	Sup.	M	DP	Inf.	Sup.	M	DP	Inf.	Sup.
25	7,6	88,5	15,2	85,2	91,8	86,8	13	84,0	89,6	86,4	13,9	83,4	89,4
50	53,2	96	13,6	94,9	97,1	96,6	14,1	95,4	97,8	95,9	13,5	94,8	97,0
75	19,6	103,9	13,8	102,0	105,8	102,5	12,8	100,8	104,2	103,5	12,7	101,8	105,2
95	19,5	108,3	13,5	106,5	110,1	108,8	13,1	107,0	110,6	109,5	13,2	107,7	111,3

Nota: *Foram observados 99 casos com dados faltosos no NSE. EE=Escala de Execução; ER=Escala de Raciocínio; SON=QI Total; IC 95%= intervalo de confiança de 95% para a média.

Como pode ser verificado, à medida que aumenta a pontuação no nível socioeconômico, também aumenta o QI médio das crianças. Embora o foco do SON-R 2½-7[a] seja a inteligência fluida, o potencial para aprender, fatores socioeconômicos conseguem diferenciar grupos por faixas de desempenho. Verifica-se que todos os intervalos de confiança das médias são distintos, ou seja, não têm valores convergentes, o que ratifica a diferença entre os grupos.

8.2 Duração da testagem

Em geral o teste foi administrado em apenas uma sessão na pesquisa de normatização e quando necessário foi dado um intervalo para a criança. Considerando o tempo total de testagem das crianças da amostra, verificou-se uma média geral de cerca de 31 minutos (DP=9). Para as crianças mais novas (3;3 a 4;3) essa média foi 27 minutos (DP=10), as da faixa intermediária 4;9 e 6;3 completaram em média com 32 minutos (DP=9) e as mais velhas (6;9 a 7;9) tiveram tempo médio de 33 minutos (DP=8).

Ao se fazer uma análise de correlação r de Pearson entre o tempo gasto pelas crianças para completar o teste e o QI total, verificou-se uma relação significativa nos três grupos de idade: no grupo mais novo (faixa de 3;3 a 4;3) a correlação foi $r=0,59$, no grupo intermediário (4;9 a 6;3) esse índice foi $r=0,46$ e no grupo mais velho (6;9 a 7;9) foi $r=0,28$. Essa relação pode ser explicada pelo fato de as crianças em cada grupo de idade, que foram melhores no teste, completaram mais itens, obtendo assim um QI mais elevado. É interessante observar que essa relação é mais alta no grupo mais novo, pois é o grupo em que se espera que as crianças completem um número menor de itens. Assim, as que levam mais tempo respondendo o teste vão mais longe e obtêm um QI mais elevado. A tabela a seguir apresenta o percentual de crianças em faixas de tempo bem como o tempo médio de duração dos subtestes.

Tabela 8.4 Duração da aplicação por faixa de idade, de tempo e por subteste (N = 1.083*)

Minutos	Percentual				Subteste	Minutos	
	3 a 8 anos	3;3-4;3	4;9-6;3	6;9-7;9		M	DP
≤ 15	4,2	9,1	2,5	1,5	Mosaicos	9,7	4,1
De 16 a 20	9,2	18,5	5,7	5,0	Categorias	8,1	3,0
De 21 a 25	15,1	18,2	15,6	11,5	Situações	6,3	2,6
De 26 a 30	23,0	20,1	23,1	25,7	Padrões	6,7	3,2
De 31 a 35	20,4	14,1	23,6	22,3			
De 36 a 40	14,2	10,3	15,4	16,4			
De 41 a 45	9,0	6,6	8,6	12,1			
> 45	4,8	3,1	5,4	5,6			

Nota: *Foram retirados 117 casos com informações inconsistentes no relato do tempo de duração. M= média; DP= desvio padrão.

8.3 Influência do examinador

Grande parte das crianças da amostra (91%) foi testada por 16 examinadores. Estes testaram de 36 a 140 crianças. Foi realizado um teste de comparação das médias das crianças testadas por diferentes examinadores, controlando-se pelo efeito do sexo e nível socioeconômico. Os resultados indicaram um efeito significativo de 0,27 ($F[15, 1029]=4,3$;

$p>0,01$). Entretanto, outras características como variabilidade nos grupos de idade, local de moradia e flutuações no perfil da amostra devem ter influenciado esse resultado. A seguir encontra-se uma tabela com o QI médio das crianças testadas por diferentes examinadores.

Tabela 8.5 Estatísticas descritivas do QI das crianças por examinador (N=1.047*)

Examinador	M	DP	N	Dev.
A	98,8	15,0	52	-0,8
B	98,7	12,0	38	-0,8
C	107,1	15,4	39	7,6
D	101,2	15,4	80	1,7
E	98,0	15,5	87	-1,5
F	98,5	15,7	104	-1,0
G	93,1	14,1	43	-6,4
H	97,8	12,1	103	-1,7
I	98,2	13,1	46	-1,3
J	92,5	15,7	140	-7,0
K	102,9	9,9	38	3,4
L	96,6	13,6	83	-2,9
M	109,0	11,7	31	9,5
N	100,1	13,3	44	0,6
O	109,4	13,9	69	9,9
P	108,8	16,6	50	9,3
Total	99,5	15,1	1.047	

Nota: *Foram retirados 153 casos, testados por examinadores que aplicaram o teste em menos de 30 crianças. Dev.=desvio em relação à média geral de 99,5; M= média; DP= desvio padrão.

8.4 Diferenças regionais

Todas as regiões brasileiras foram contempladas na amostra e dentro destas diferentes estados e cidades com diversos IDH. Foi solicitado o QI médio das crianças por região e por localidade (capital e interior). A Tabela 8.6 apresenta os resultados.

Tabela 8.6 Média, desvio padrão e intervalo de confiança da média para o QI das crianças por região e localidade

Região	M	DP	N	e.p.m	IC 95%		Local	M	DP	N	e.p.m	IC 95%	
					Inf.	Sup.						Inf.	Sup.
Norte	96,1	13,2	120	1,2	93,7	98,5	Capital	104,1	14,3	381	0,7	102,6	105,6
Nordeste	96,7	15,7	381	0,8	95,1	98,3	Interior	98,1	14,9	819	0,5	97,1	99,1
Sudeste	103,4	13,8	459	0,6	102,1	104,7							
Sul	98,1	14,2	160	1,1	95,9	100,3							
Centro-Oeste	105,7	16,5	80	1,8	102,0	109,4							

Nota: e.p.m.=erro padrão de mensuração; IC 95%= intervalo de confiança de 95% para a média; M= média; DP= desvio padrão.

As diferenças nas médias dos grupos são significativas ($p < 0,05$), mas outros fatores podem estar influenciando esse resultado, como nível socioeconômico e escolaridade. Os intervalos de confiança das médias mostram que, em geral, os valores não se distanciam muito do valor médio.

8.5 Influência do gênero

Na próxima tabela (Tabela 8.7) é apresentado o resultado médio de meninos ($N = 600$) e meninas ($N = 600$) nos escores normalizados por subteste e nos QIs. Foi realizado um teste t com o objetivo de verificar quais diferenças entre as médias são estatisticamente significativas entre meninos e meninas. Apenas no subteste Categorias e na Escala de Raciocínio foram encontradas diferenças significativas.

Tabela 8.7 Média, desvio padrão e estatísticas do teste *t* para a comparação de médias por sexo

Escore	Masculino		Feminino		Diferença	<i>t</i>	<i>p</i>
	M	DP	M	DP			
Mosaicos	10,0	3,2	10,1	2,9	-0,1	-0,35	0,724
Categorias	9,8	3,0	10,3	2,9	-0,5	-2,82	0,005
Situações	9,9	3,0	10,0	3,0	-0,1	-0,65	0,518
Padrões	9,9	3,1	10,1	2,9	-0,3	-1,45	0,146
SON-EE	99,6	15,6	100,4	14,4	-0,8	-0,95	0,342
SON-ER	99,1	15,1	100,9	14,8	-1,7	-1,98	0,048
SON-QI	99,3	15,3	100,7	14,6	-1,4	-1,61	0,107

Nota: M= média; DP= desvio padrão; *t*= estatística do teste *t* de *student*; *p*= *p-value*.

8.6 Avaliação realizada pelos examinadores

Ao final da administração do teste o comportamento da criança durante a testagem foi avaliada pelo examinador em quatro aspectos: motivação, concentração, cooperação com o examinador e compreensão das instruções. Para tanto, foi utilizada a seguinte escala: **1**=boa, **2**=variável, **3**=mediana e **4**=muito baixa. Com base nos dados coletados foram efetuadas análises a fim de verificar a média no teste das crianças avaliadas sob diferentes condições. As categorias de resposta 2 e 3 foram agregadas nas análises. Os resultados da relação entre a avaliação feita pelo examinador e o QI das crianças está apresentada na Tabela 8.8, a seguir.

Tabela 8.8 Percentual, média e desvio padrão por situação de testagem e grupo de idade

Característica	Avaliação	3;3-4;3			4;9-6;3			6;9-7;9		
		%	M	DP	%	M	DP	%	M	DP
Motivação	Boa	65,8	102,4	14,6	80,5	101,6	14,7	91,8	100,6	14,6
	Var./mediana	29,2	95,6	14,5	18,4	95,0	15,3	7,6	95,2	15,0
	Muito baixa	5,0	90,1	16,8	1,0	79,6	19,0	0,6	101,5	9,2
Concentração	Boa	50,3	104,8	14,6	77,0	102,7	14,0	87,9	101,9	14,0
	Var./mediana	41,9	95,7	13,4	21,1	92,8	15,6	11,6	88,2	14,2
	Muito baixa	7,8	89,1	15,4	1,9	79,4	16,5	0,6	84,0	1,4
Cooperação	Boa	74,4	101,9	15,0	86,6	101,4	14,5	93,8	101,0	14,3
	Var./mediana	21,7	93,6	13,1	12,3	92,3	18,0	6,2	88,1	14,8
	Muito baixa	3,9	93,9	17,5	1,0	93,0	9,1	-	-	-
Compreensão	Boa	51,1	107,3	12,6	82,0	103,2	13,7	91,5	101,9	13,5
	Var./mediana	36,7	94,9	12,2	16,5	87,7	13,2	7,9	81,9	13,9
	Muito baixa	12,2	82,7	13,1	1,5	72,1	13,8	0,6	75,0	1,4

Nota: M= média; DP= desvio padrão.

O número de crianças em cada situação avaliada por grupo de idade variou de zero a 860, o que impossibilita comparações por meio de testes estatísticos, pois o número de casos é muito discrepante de um grupo para outro. Os percentuais em cada situação de testagem e grupo de idade indicam a grande variabilidade quanto ao número de crianças. Dessa forma, foi efetuada uma análise de correlação entre a situação de testagem, avaliada na escala de 1 (Boa) a 4 (Muito baixa), e o QI total. Os resultados indicaram maior correlação com compreensão das instruções ($r = -0,44$, $p < 0,001$) e concentração ($r = -0,30$, $p < 0,001$), indicando que esses são aspectos importantes na testagem.

9. Considerações Finais

Entre os testes utilizados por psicólogos e outros profissionais, particularmente por aqueles que trabalham com crianças e adolescentes em escolas, os testes de inteligência são os de uso mais freqüente. A aferição da inteligência geralmente é realizada com o objetivo de avaliação diagnóstica e, os seus resultados têm consequências importantes tanto para a vida escolar como para a formulação de recomendações sobre a criação de programas especiais de educação e tratamento das desordens identificadas (Hutz & Bandeira, 1993; International Test Commission, 2003).

A presente tese teve como objetivo principal estabelecer normas nacionais e obter evidências sobre a validade do SON-R 2½-7[a]. O processo de normatização e validação de testes psicológicos é frequente no âmbito da psicologia. Não obstante, em geral esses estudos são realizados tendo como base dados coletados em grandes centros urbanos ou em uma cidade específica (ver Rosa & Alves, 2000; Wechsler, 2002b). Uma das contribuições mais importantes da presente pesquisa é que na mesma buscou-se representar a diversidade da realidade socioeconômica e cultural brasileira na amostra. Dessa forma, a coleta de dados realizada nas cinco regiões brasileiras e com participantes de 37 cidades com diferentes Índices de Desenvolvimento Humano, possibilitou a obtenção de grande variabilidade de participantes.

Quanto ao processo de normatização, o método utilizado também foi inovador, pois permite a obtenção de normas para a idade da criança em anos, meses e dias. Embora nesse estudo estejam apresentadas, em parte, apenas as tabelas de conversão de escores mês a mês, o *software* do teste disponibilizará as normas para cada idade exata. Esse *software* será disponibilizado juntamente com o teste para aqueles que o adquirirem. Ademais, alguns resultados encontrados na normatização brasileira superam aqueles da Holanda (ver Tellegen & cols., 1998). Por exemplo, os índices de fidedignidade, em geral, foram superiores na

versão brasileira. Esse resultado foi impactado diretamente pela grande variabilidade do desempenho das crianças brasileiras.

No que se refere ao estudo de validade de construto do SON-R 2½-7[a], este teve como objetivo validar a estrutura do teste com um fator geral e dois fatores de primeira ordem. Nesse sentido, não é de hoje que pesquisadores e estudiosos da área de inteligência procuram evidências sobre a arquitetura intelectual humana. Spearman (1904), por exemplo, propôs um fator geral de inteligência, fator *g*; outros estudiosos, como Cattell (1963, 1967) e Horn (1994), assinalam que a inteligência pode ser explicada em termos de dois fatores: inteligência fluida e cristalizada. Carroll (1993), por sua vez, descreve a estrutura da inteligência como uma pirâmide, sendo o equivalente ao *g* o topo da pirâmide, o centro equivalente às inteligências fluida e cristalizada e a base estaria composta por fatores específicos como raciocínio quantitativo. Gomes e Borges (2007) testaram o modelo de Carroll com uma amostra brasileira e identificaram os três níveis propostos por esse autor.

Na presente tese a validade do SON-R 2½-7[a] foi verificada por meio da validade convergente, da Análise Fatorial Confirmatória e da solução Schmid-Leiman para verificação de um fator de segunda ordem. O estudo de validade convergente do SON-R 2½-7[a] com o WPPSI-III, WISC-III, Matrizes Progressivas Coloridas de Raven e a Escala de Maturidade Mental Columbia indicou que o SON-R 2½-7[a] possui correlações altas com todos esses testes, o que evidencia sua validade.

A correlação elevada com o WPPSI-III ($r = 0,75$), igual ao resultado encontrado na Austrália ($r = 0,75$) (Tellegen & cols., 1998), indica que o SON-R 2½-7[a] avalia um campo amplo da inteligência. O WPPSI-III, como é destinado para crianças mais novas do que o WISC-III, é menos verbal que este, portanto, é razoável que o SON-R 2½-7[a] se assemelhe mais ao WPPSI-III do que ao WISC-III, embora ainda não se tenha normas brasileiras do WPPSI-III. Também se pode ressaltar o fato de que as normas brasileiras do WISC-III foram

obtidas tendo como base uma amostra pouco representativa da realidade brasileira, com crianças apenas do Rio Grande do Sul (ver Wechsler, 2002b). Ademais, as correlações com o Columbia ($r = 0,56$) e o Raven ($r = 0,68$) denotam que o SON-R 2½-7[a] avalia uma parte do construto de inteligência que esses dois testes não avaliam.

Os resultados da Análise Fatorial Confirmatória encontrados na presente tese indicam que o SON-R 2½-7[a] possui uma estrutura robusta com dois fatores: Execução e Raciocínio, em diferentes grupos de idade. Sua estrutura é invariante tanto no grupo total de crianças, com idade entre 3;3 a 7;9, quanto dividindo-se a amostra em três grupos distintos de idade. O ajuste da estrutura fatorial do SON-R 2½-7[a] com dois fatores foi muito bom em todas as condições testadas.

Os resultados da solução Schmid-Leiman com um fator de segunda ordem para o SON-R 2½-7[a] indicaram a existência de dois fatores de primeira ordem, a saber: Execução e Raciocínio, assim como um fator geral de segunda ordem. Tal resultado embasa o uso do SON-R 2½-7[a] com uma escala geral e com duas escalas específicas.

De maneira geral, os estudos de validade realizados nessa tese indicam que o SON-R 2½-7[a] é um teste válido, fidedigno e robusto em sua estrutura interna. Os estudos realizados apontam ainda, que se trata de um teste com alta qualidade psicométrica e com índices de confiabilidade adequados para uso nas diferentes faixas etárias contempladas pelo teste.

Por fim, considerando variáveis externas ao SON-R 2½-7[a], tais como nível socioeconômico e situação de testagem, foi verificado se havia diferença nos escores das crianças, pois Spearman (1904), por exemplo, verificou que a habilidade de discriminação sensorial, que ele equivocadamente acreditava ser o coração da inteligência, era influenciada pelo meio (associado ao nível socioeconômico). E considerando-se o teste de Binet, constatou-se que também havia diferenças no desempenho entre crianças com diferentes níveis socioeconômicos (Davidson & Downing, 2000). Na presente tese também foram

verificadas diferenças no desempenho médio das crianças com menor nível socioeconômico (QI = 89) daquelas com nível socioeconômico intermediário (QI = 96) e alto (QI = 104 e 108). Hipotetiza-se que a estimulação precoce, mais comum em lares de pessoas com maior nível socioeconômico e cultural, tenha impactado nesse resultado.

Quanto à situação de testagem avaliada pelo examinador, durante a aplicação, verificou-se a relação entre esta e o resultado no SON-R 2½-7[a]. Os resultados indicaram que os fatores mais importantes durante a aplicação do teste são a compreensão das instruções por parte das crianças e a concentração. Dessa forma, esses fatores devem ser observados e incentivados durante o processo de testagem.

Frente aos resultados encontrados, a presente tese disponibiliza um teste não-verbal de inteligência válido e com normas brasileiras, expandidas para a faixa entre 2 anos e 6 meses e 7 anos e 11 meses – recomenda-se cautela no uso das normas nas idades extremas dessa faixa etária. É uma contribuição para a área de avaliação psicológica infantil, uma vez que essa é uma faixa etária muito importante, especialmente até os 5 anos de idade, para a intervenção precoce (Campbell & Ramey, 1995).

Ademais, como a amostra contempla a diversidade socioeconômica e cultural brasileira, configura-se como um estudo pioneiro na área de validação e normatização de testes no âmbito da psicologia nacional. Como o SON-R 2½-7[a] tem como foco a inteligência fluida – que pode ser vista como a habilidade de compreender relações complexas e resolver problemas novos (Martinez, 2000) –, a qual é predita fortemente pela memória de trabalho (Yuan e cols., 2006; Kane, Hambrick, & Conway, 2005), é relevante para o diagnóstico de atrasos ou déficits na aprendizagem.

10. Limitações e agenda de pesquisa

Assim como outros estudos, o presente também conta com algumas limitações, por exemplo, as amostras utilizadas nos estudos de validade convergente. Uma vez que se trata de aplicação individual, a maior amostra utilizada foi 120 crianças, todas do Distrito Federal. Dessa forma, outros estudos de validade convergente com o SON-R 2½-7[a] poderiam ser realizados com amostras maiores e com outros testes de inteligências validados para o contexto brasileiro.

Não obstante a amostra de normatização do SON-R 2½-7[a] tenha contemplado a realidade nacional por meio da aplicação em todas as regiões brasileiras, em diferentes estados e cidades, não é considerada estatisticamente representativa. Tal fato, porém, embora seja uma limitação da amostra, é o viável na atual conjectura. A seleção de uma amostra estatisticamente representativa em pesquisas com seres humanos – assim como com animais, teria que se enquadrar nas normas do Comitê de Ética em Pesquisa, que possui como um requisito básico para participação o voluntariado e a desistência em qualquer momento. Dessa forma, se, por exemplo, ao ser realizada uma pesquisa com seres humanos, fosse considerada uma amostra probabilística, as instituições sorteadas teriam que participar; ou, se fosse uma amostra com reposição, a cada declínio, uma nova instituição seria contactada. E quando a coleta de dados fosse realizada com crianças não institucionalizadas, quantos declínios e reposições não seriam necessários!

11. Referências

- Alchieri, J. C., Noronha, A. P. P. & Primi, R. (2003). *Guia de referência: testes psicológicos comercializados no Brasil*. São Paulo: Casa do Psicólogo – FAPESP.
- Angelini, A. L., Alves, I. C. B., Custório, D. M., Duarte, W. F. & Duarte, J. L. M. (1999). *Matrizes Progressivas Coloridas de Raven: Escala especial*. São Paulo: Centro Editor de Testes e Pesquisas em Psicologia.
- Baker, F. B. & Kim, S. (2004). *Item response theory: parameter estimation techniques*. Nova York: Marcel Dekker.
- Berge, J. M. F. ten & Zegers, R. E. (1978). A series of lower bounds to the reliability of a test. *Psychometrika*, 4, 575-579.
- Bock, R. D., Gibbons, R. D. & Muraki, E. (1988). Full Information item Factor Analysis. *Applied Psychological Measurement*, 12, 261-280.
- Boring, E. G. (1923). Intelligence as the tests test it. *New Republic*, 35, 35-37.
- Brody, N. (2000). History of theories and measurements of intelligence. Em R. J. Sternberg. (Org.), *Handbook of intelligence*, pp. 16-33. New York: Cambridge university press.
- Bugg, J. M., Zook, N. A., DeLosh, E. L., Davalos, D. B. & Davis, H. P. (2006). Age differences in fluid intelligence: Contributions of general slowing and frontal decline. *Brain and Cognition*, 62, 9-16
- Burgemeister, B. B., Blum, L. H. & Lorge, I. (2001). *Escala de Maturidade Mental Colúmbia: manual para aplicação e interpretação*. Revisão técnica Irai Cristina Boccato Alves, José Luciano Miranda Duarte; Tradução Heloisa da Costa Marques Faria – 3ª. Ed. São Paulo: Casa do Psicólogo.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum Associates.
- Campbell, F. A. & Ramey, C. T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: positive effects of early intervention. *American Educational Research Journal*, 32(4), 743-772.
- Carey, S. (1978). The child as word learner. Em M. Halle, J. Bresnan e G. A. Miller (Orgs.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.
- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40, 153-193.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Cattell, R. B. (1967). The theory of fluid and crystallized general intelligence checked at the 5-6 year-old level. *British Journal of Educational Psychology*, 37, 209-224.
- Cattell, R.B. (1950). *Handbook for the individual of group Culture Fair Intelligence Test. Scale I*. Champaign, Ill: I.P.A.T.

- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. 2^a Ed. Hillsdale, NJ: Erlbaum.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando: Harcourt Brace Jovanovich College Publishers.
- Cronbach, L. J. (1996). *Fundamentos da testagem psicológica*. 5^a Ed. Porto Alegre: Artes Médicas.
- Davidson, J. E. & Downing, C. L. (2000). Contemporary models of intelligence. Em R. J. Sternberg. (Org.), *Handbook of intelligence*, pp. 16-33. New York: Cambridge university press.
- Drunen, P. V. (2003). *Tussen traditie em vernieuwing – zestig jaar SON-tests*. Amsterdam: Boom test uitgevers.
- Evers, A., Van Vliet-Mulder, J. C. & Groot, J. (2000). *Documentation of tests and test research in the Netherlands by the Committee on Test Evaluation (COTAN)*. Assen: Van Gorcum/NIP.
- Feuerstein, R. (1980). *Instrumental enrichment: an intervention program for cognitive modifiability*. Baltimore, MD: University Park Press.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: what IQ tests really measure. *Psychological Bulletin*, 2, 171-191.
- Galton, F. (1869). *Hereditary genius: an inquiry into its laws and consequences*. London: Macmillan.
- Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York: Basic.
- Gomes, C. M. A. & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira. *Avaliação Psicológica*, 6(2), 167-179.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Guilford, J. P. (1981). Higher-order structure-of-intellect abilities. *Multivariate Behavioral Research*, 16, 411-435.
- Guilford, P. J. (1985). The structure-of-intellect model. Em B. B. Wolman (Org.), *Handbook of intelligence: theories, measurements, and applications*, pp. 225-266. New York: Wiley.
- Hakstian, A. R. & Cattell, R. B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology*, 70, 657-669.
- Halstead, W. C. (1951). Biological intelligence. *Journal of Personality*, 20, 118-130.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications.
- Hebb, D. O. (1942). The effects of early and late brain injury upon test scores and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society*, 85, 275-292.
- Horn, J. L. (1985). Remodeling old models of intelligence. Em B. B. Woman (Org.), *Handbook of intelligence: theories, measurements, and applications*, pp. 267-300. New York: Wiley.
- Horn, J. L. (1994). Theory of fluid and crystallized intelligence. Em R. J. Sternberg (Org.), *Encyclopedia of human intelligence* (pp. 443-451). New York: Macmillan.

- Horn, J. L., Donaldson, G. & Engstrom, R. (1981). Apprehension, memory, and fluid intelligence decline in adulthood. *Research on aging*, 3, 33-84.
- Hu, S. & Oakland, T. (1991). Global and regional perspectives on testing children and youth: an empirical study. *International Study of Psychology*, 26, 329-344.
- Hunt, E. B. (1978). Mechanics of verbal ability. *Psychological Review*, 85, 109-130.
- Hutz, C. & Bandeira, D. (1993). Tendências contemporâneas no uso de testes, uma análise da literatura brasileira e internacional. *Psicologia: Reflexão e Crítica*, 6, 85-101.
- IBGE (2000). *Atlas do Desenvolvimento Humano no Brasil*. Brasília: IBGE.
- IBGE (2007). Síntese de indicadores sociais: uma análise das condições de vida da população brasileira. *Estudos & Pesquisas*, nº. 21.
- International Test Commission (2003). *Diretrizes para o uso de testes: International Test Commission (ITC)*. Versão em Português. Retirado em 21/01/2007 do website do IBAP: www.ibapnet.org.br/diretrizesITC.pdf.
- Jesus, G. R. de & Laros, J. A. (2006). *Performance of Brazilian and Dutch children on the SON-R Categories subtest*. Trabalho apresentado na 5ª Conferência do International Test Commission, Brussels, Belgium. Retirado em 22/01/2007 do website da International Test Commission (ITC), www.intestcom.org/Downloads/ITC2006Brussels/Session3.2.1RibeiroandLaros.pdf.
- Kane, M. J., Hambrick, D. Z. & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 66-71.
- Kiers, H.A.L. (1990). *SCA: a program for simultaneous component analysis*. Groningen: IEC, ProGamma.
- Kiers, H.A.L. & ten Berge, J.M.F. (1989). Alternating least squares algorithms for simultaneous component analysis with equal weight matrices in two or more populations. *Psychometrika*, 54, 467-473.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.
- Laros, J. A.; Pasquali, L. & Rodrigues, M. M. M. (2000). *Análise da unidimensionalidade das provas do Saeb*. Universidade de Brasília, Instituto de Psicologia, Centro de Pesquisa em Avaliação Educacional.
- Laros, J. A. & Tellegen, P. J. (1991). *Construction and validation of the SON-R 5½-17, the Snijders-Oomen non-verbal intelligence test*. Groningen: Wolters-Noordhoff.
- Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh*, 61A, 273-287.
- Lynn, R. & Hampson, S. (1986). The rise of national intelligence: evidence from Britain, Japan and the U.S.A. *Personality and Individual Differences*, 1, 23-32.
- Lord (1952). A theory of test scores. *Psychometric Monographs*, 7.
- Luria, A. R. (1973). *The working brain*. New York: Basic.
- Martinez, M. E. (2000). *Education as the cultivation of intelligence*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Miles, J. & Shevlin, M. (2001). *Applying regression & correlation – a guide for students and researchers*. London: Sage Publications.
- Millsap & R.E. & Meredith, W.M. (1988). Component analysis in cross-sectional and longitudinal data. *Psychometrika*, 53, 123-124.
- Neisser, U. & cols. (1996). Intelligence: knowns and unknowns. *American psychologist*, 51(2), 77-101.
- Neisser, U. (1979). The concept of intelligence. Em R. J. Sternberg & D. K. Detterman (Orgs.), *Human intelligence: perspectives on its theory and measurement*. Norwood, NJ: Ablex.
- Niesing, J. (1997). *Simultaneous component and factor analysis methods for two or more groups: a comparative study*. M & T Series, 30. Leiden: The Netherlands, DSWO Press.
- Pasquali, L. (2005). *Manual Técnico e de Aplicação do Teste Não-Verbal de Raciocínio para Crianças - TNVRI*. 1. ed. São Paulo: Vetor Editora.
- Pasquali, L. (2003). *Psicometria – Teoria dos testes na psicologia e na educação*. Petrópolis, RJ: Vozes.
- Piaget, J. (1972). *The psychology of intelligence*. Totowa, NJ: Littlefield Adams.
- Pilati, R. & Laros, J. A. (2007). Modelos de equações estruturais em psicologia: conceitos e aplicações. *Psicologia: Teoria e Pesquisa*, 23(2), 205-216.
- Primi, R. (2001). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence*, 30, 41-70.
- Quesada, A. A. (2007). *Avaliação neuropsicológica verbal versus não-verbal de crianças com epilepsia*. Dissertação de mestrado. Universidade de Brasília.
- Raven, J.C. (1938). *Progressive Matrices: A perceptual test of intelligence*. London: Lewis.
- Rosa, H.R. & Alves, I.C.B. (2000). *R-2: Teste Não-Verbal de Inteligência para Crianças*. Manual. São Paulo; Vetor Editora.
- Rosnow, R. L. & Rosenthal, R. (1996). Computing contrasts, effect sizes and counternulls on other people's published data: general procedures for research consumers. *Psychological Methods*, 1(4), 331-340.
- Sijtsma, K. (1993). Kaf en koren onder Nederlandse tests. *De Psycholoog*, 28(12), 502-503.
- Simon, H. A. (1976). Identifying basic abilities underlying intelligent performance of complex tasks. Em L. B. Resnick (Org.), *The nature of intelligence*, pp. 65-98. Hillsdale, NJ: Erlbaum.
- Snijders, J. Th. & Snijders-Oomen, N. (1958). *Snijders-Oomen niet-verbale intelligentieschaal SON-'58*. Groningen: Wolters-Noordhoff.
- Snijders, J. Th. & Snijders-Oomen, N. (1976). *Snijders-Oomen Non-verbal Intelligence Scale, SON 2½-7*. Groningen: Tjeenk Willink BV.
- Snijders, J. Th., Tellegen, P. J. & Laros, J. A. (1989). *Snijders-Oomen non-verbal intelligence test, SON-R 5½-17. Manual and research report*. Groningen: Wolters-Noordhoff.
- Snijders-Oomen, N. (1943). *Intelligentieonderzoek van doofstomme kinderen* [The examination of intelligence with deaf-mute children]. Nijmegen: Berkhout.

- Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C. (1927). *The abilities of man*. London: Macmillan.
- Starren, J. (1975). SSON 7-17. *De ontwikkeling van een nieuwe versie van de SON voor 7-17 jarigen. Verantwoording en hanleiding*. Groningen: Wolters-Noordhoff.
- Sternberg, R. J. (2000). The concept of intelligence. Em R. J. Sternberg. (Org.), *Handbook of intelligence*, pp. 3-15. New York: Cambridge university press.
- Sternberg, R. J. (1990). *Metaphors of mind: conceptions of the nature of intelligence*. New York: Cambridge University Press.
- Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: the componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.
- Sternberg, R. J. (1997). *Successful intelligence*. New York: Plume.
- Tabachnick, B. G., & Fidell, L.S. (1996). *Using multivariate statistics* (3^a ed.). Northridge, CA: HarperCollins College.
- 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*, 20(2), 103-111.
- Tellegen, P.J., Winkel, M., Wijnberg-Williams, B. & Laros, J. A. (1998). *Snijders-Oomen Nonverbal Intelligence Test SON-R 2½-7: Manual and research report*. The Netherlands: Swets Test Publishers.
- Tellegen, P.J. & Laros, J.A. (1993). The Snijders-Oomen nonverbal intelligence tests: general intelligence tests or tests for learning potential? Em Hamers, J.H.M. & Sijtsma, K. & Ruijsenaars, A.J.J.M. *Learning Potential Assessment. Theoretical, Methodological and Practical Issues*. Amsterdam: Swets & Zeitlinger.
- Tellegen, P.J. & Laros, J.A. (2005). *Fair assessment of children from cultural minorities: a description of the SON-R non-verbal intelligence tests*. Slovakia: Paper presented at the UNESCO seminar.
- Thompson, B. (2005). *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, DC: American Psychological Association.
- Thorndike, R.L., Hagen, E.P. & Sattler J.M. (1986). *The Stanford-Binet intelligence scale: Fourth edition technical manual*. Chicago: The Riverside Publishing Company.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Thurstone, L. L. & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.
- Ullman, J. B. (2007). Structural equation modeling. Em B. G. Tabachnick & L. S. Fidell (Orgs.), *Using multivariate* (5^a ed.). Boston: Pearson Education.
- Van de Vijver, F. J. R. & Poortinga, Y. H. (1997). Testing in culturally heterogeneous populations: when are cultural loadings undesirable? *European Journal of Psychological Assessment*, 8, 17-24.
- Vernon, P. E. (1971). *The structure of human abilities*. London: Methuen.
- Wadsworth, B. R. (1997). *Inteligência e afetividade da criança na teoria de Piaget*. São Paulo: Pioneira.

- Wechsler, D. (2002a). *WPPSI-III – Technical and interpretive manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2002b). *WISC-III – Escala de Inteligência Wechsler para Crianças: Manual/David Wechsler, 3ª. Ed.; Vera Lúcia Marques de Figueiredo*. São Paulo: Casa do Psicólogo.
- Weiss, D.J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement, 6*, 473-492.
- Wilson, D. T., Wood, R. & Gibbons, R. (1991). *Testfact: test scoring, item statistics, and item factor analysis*. Chicago: Scientific Software.
- Yuan, K., Steedle, J., Shavelson, R., Alonzo, A. & Oppezzo, M. (2006). Working memory, fluid intelligence, and science learning. *Educational Research Review, 1*, 83-98.

Apêndice 1: Roteiro de treinamento dos examinadores

Treinamento dos examinadores para a pesquisa nacional com o Teste não-verbal de Inteligência SON-R 2,5-7 anos

Professor Responsável: Jacob Arie Laros
Alunos: Girlene R. de Jesus (doutoranda)
Cleide Sousa (mestre)
Camila Karino (Bacharel)

Datas: 20 a 22 de maio e 27 a 29 de maio de 2007.

Local: Universidade de Brasília – CAEP (Sala 5)

Programação

1º dia

09:00 – Boas vindas e apresentação da equipe
09:30 – Apresentação do teste não-verbal de inteligência SON (Prof. Laros)
10:30 – Intervalo
11:00 – Apresentação da pesquisa e delineamento (Girlene e Prof. Laros)
12:00 – Intervalo para almoço
14:00 – Demonstração de aplicação do teste SON-R 2½-7[a] anos, acompanhando o manual
15:30 – Intervalo
15:50 – Demonstração de aplicação do teste SON-R 2½-7[a] anos, acompanhando o manual (continuação)
18:00 – Encerramento

2º dia

09:00 – Treino de aplicação do teste SON-R 2½-7[a] anos em duplas, com idades pré-determinadas para as duplas.
10:10 – Intervalo
10:30 – Palestra sobre desenvolvimento infantil e aplicação de testes em crianças (Profa. Rosana)
12:00 – Intervalo para almoço
14:00 – Treino de aplicação do teste SON-R 2½-7[a] anos em duplas, com idades pré-determinadas para as duplas (continuação)
15:30 – Intervalo
15:50 – Correção do teste SON-R 2½-7[a] anos
18:00 – Encerramento

3º dia

09:00 – Demonstração pelos examinadores da aplicação do teste SON-R 2½-7[a] anos. Cada dupla aplica um subteste para demonstrar para o grupo.
10:30 – Intervalo
10:50 – Demonstração pelos examinadores da correção do teste SON-R 2½-7[a] anos
12:00 – Intervalo para almoço
14:00 – Feedbacks e Dúvidas finais sobre a aplicação do teste SON-R 2½-7[a] anos.
16:00 – Encerramento

Apêndice 2: Folha de resposta do SON-R 2½-7 [a]



SON-R 2¹/₂-7 [a]

Teste Não-verbal de Inteligência

Formulário de Registro

Informações Gerais

Nome da criança: _____ Código: _____

Data de aplicação do teste: ____/____/____ (dd/mm/aaaa)

Data de nascimento da criança: ____/____/____ (dd/mm/aaaa)

Idade: _____ (em anos e meses)

Sexo: () M () F

Nome do local de aplicação: _____

Local da aplicação: () Creche () Pré-escola () Escola () Outro: _____

Série: () Não frequenta escola () Pré-escola () 1ª série () Outro: _____

Código do Examinador: _____ Nome do Examinador: _____

Presença de outros: () Mãe () Pai () Mãe e Pai () Outros () Ninguém

Questionário Contextual

1. Cidade: _____ Estado: _____

2. Grau de escolaridade do pai ou responsável pela criança: _____

3. Grau de escolaridade da mãe ou responsável pela criança: _____

4. Ocupação do pai ou responsável pela criança: _____

5. Ocupação da mãe ou responsável pela criança: _____

6. Quantidade de televisões em casa []

7. Quantidade de computadores em casa []

8. Quantidade de automóveis próprios []

9. Quantidade de geladeiras em casa []

10. Quantidade de máquinas de lavar roupa em casa []

11. Quantidade de banheiros em casa []

12. Quantidade de empregadas domésticas em casa []

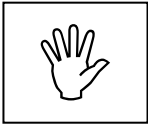
13. Grupo étnico da criança:

() Branco () Pardo/Mulato () Negro () Amarelo () Indígena

Início →

Item 1 = primeiro item a ser aplicado para crianças de dois a três anos
Item 3 = primeiro item a ser aplicado para crianças de quatro e cinco anos
Item 5 = primeiro item a ser aplicado para crianças de seis anos para cima

Quando o primeiro ou o segundo item administrado for completado incorretamente, os dois itens anteriores que não haviam sido aplicados devem ser administrados.

 Um subtteste é interrompido quando há um total de 3 respostas incorretas. Não é necessário que os erros sejam consecutivos.

Os subttestes de execução (Mosaicos e Padrões) são, ainda, interrompidos quando há dois erros consecutivos na parte II destes subttestes.

Instruções para a Aplicação dos Subttestes

1. MOSAICOS

INÍCIO: ___ h ___ min. TÉRMINO: ___ h ___ min. DURAÇÃO DO TESTE: ___ min.

Parte I: item 1-6

Somente os quadrados vermelhos são utilizados. O número de quadrados varia por item, e é dado somente o número preciso de quadrados necessários para o item. A “inversão” da figura é contada como sendo correta.

Itens 1-3: A moldura é utilizada tanto pelo examinador quanto pela criança. **Começando com o item 4,** o examinador não utiliza a sua moldura e tampouco demonstra o item. **Item-de-entrada 5:** somente uma moldura para a criança.

Parte II: Exemplo A

Inclua os quadrados amarelos. Demonstre o exemplo A na moldura da criança. **Iniciando com o exemplo A** apresente os quadrados na caixa e permaneça com o terceiro compartimento fechado.

Parte II: item 7-15

Tempo-limite: 2½ minutos por item.

Na parte II, o subtteste é, também, descontinuado, quando dois erros consecutivos são feitos. Apresente os quadrados amarelos/vermelhos antes de iniciar o **item 9**, quando os três compartimentos são abertos.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Score	

A

2. CATEGORIAS

INÍCIO: ___ h ___ min. TÉRMINO: ___ h ___ min. DURAÇÃO DO TESTE: ___ min.

Parte I : itens 1-7

Os cartões com os números 1 e 2 são feitos pelo examinador no **item 1**. A criança coloca os cartões numerados do 3 ao 6. Os cartões com número 1 e 2 são feitos pelo examinador nos itens **2 ao 5**. A criança coloca os cartões numerados do 3 ao 8. **A partir do item 6**, o examinador não mais demonstra os cartões.

Parte II: exemplo A

Demonstre o exemplo A primeiro e depois deixe a criança fazer o item sozinha.

Parte II: itens 8-15

Dirija a atenção da criança para a página à sua direita e, depois, coloque os cartões numerados do 1 ao 5 na página à sua esquerda. Ambas as alternativas devem estar corretas. Use a cartolina amarela para cobrir nos **itens 10 e 11**, agora a criança tem de apontar as alternativas. Ambas as figuras devem estar corretas. A cartolina amarela não é mais utilizada **a partir do item 12**.

1	2	3	4	5	6	7		8	9	10	11	12	13	14	15		Escore	
							A											

3. SITUAÇÕES

INÍCIO: ___ h ___ min. TÉRMINO: ___ h ___ min. DURAÇÃO DO TESTE: ___ min.

Parte I : itens 1-6

Coloque o cartão 1 no lugar correto e deixe a criança colocar os cartões 2-4 nos lugares restantes.

Parte II: exemplo A

Dirija a atenção da criança para a página à sua direita. Coloque os cartões do 1 ao 5 na página à sua esquerda. Demonstre o exemplo e também deixe a criança fazê-lo.

Parte II: itens 7-14

Nos **itens 7 e 8**, os cartões são posicionados na página à sua esquerda e a criança escolhe. Nos **itens 9 e 10**, as alternativas são, inicialmente, encobertas com a cartolina amarela, dirija a atenção da criança para o desenho e, depois, mostre as alternativas. **Iniciando com o item 11**, a cartolina amarela não mais é utilizada. **Iniciando com o item 12**, ambas as figuras devem estar corretas, o correto posicionamento na figura maior não é necessário.

1	2	3	4	5	6		7	8	9	10	11	12	13	14		Escore
						A										

Apêndice 3: Tabelas de conversão de escores brutos em escores normatizados e QI

Tabela A1. Conversão dos escores brutos em escores normalizados**2;10 anos**

Escores normalizados dos subtestes																	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mos	5	9	11	13	14	15	16	17	18	19	19	19	19	19	19	19	.
Cat	8	10	11	12	13	13	14	15	16	17	18	19	19	19	19	19	.
Sit	6	8	9	10	11	13	14	15	16	17	18	19	19	19	19	.	.
Pad	5	7	8	9	10	12	13	14	16	17	19	19	19	19	19	19	19

2;11 anos

Escores normalizados dos subtestes																	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mos	5	8	11	13	14	15	15	17	18	19	19	19	19	19	19	19	.
Cat	7	10	11	11	12	13	14	15	16	17	18	19	19	19	19	19	.
Sit	6	8	9	10	11	12	13	15	16	17	18	19	19	19	19	.	.
Pad	5	7	8	9	10	11	13	14	16	17	19	19	19	19	19	19	19

3;0 anos

Escores normalizados dos subtestes																	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mos	4	8	11	12	13	14	15	16	18	19	19	19	19	19	19	19	.
Cat	7	10	10	11	12	13	14	15	16	17	18	19	19	19	19	19	.
Sit	5	8	9	10	11	12	13	14	16	17	18	19	19	19	19	.	.
Pad	5	7	7	8	9	11	12	14	15	17	18	19	19	19	19	19	19

3;1 anos

Escores normalizados dos subtestes																	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mos	4	8	10	12	13	14	15	16	17	19	19	19	19	19	19	19	.
Cat	7	9	10	11	12	13	14	15	16	17	17	18	19	19	19	19	.
Sit	5	7	9	10	11	12	13	14	15	17	18	19	19	19	19	.	.
Pad	4	6	7	8	9	10	12	13	15	17	18	19	19	19	19	19	19

Tabela A2. Conversão dos escores normalizados em QI

Soma 16	PS	Soma 25	RS	Soma 1256	QI	80%-int	pct	Sum 1256	QI	80%-int	pct
2	52	2	52	4	50	47-62	1%	40	100	92-107	50%
3	54	3	55	5	50	47-62	1%	41	101	94-109	53%
4	57	4	57	6	51	48-63	1%	42	103	95-110	58%
5	59	5	60	7	52	49-64	1%	43	104	96-112	61%
6	62	6	62	8	53	50-65	1%	44	106	98-113	66%
7	65	7	65	9	55	51-66	1%	45	108	99-114	70%
8	67	8	67	10	56	53-68	1%	46	109	101-116	73%
9	70	9	70	11	57	54-69	1%	47	111	102-117	77%
10	73	10	72	12	59	55-70	1%	48	112	104-119	79%
11	75	11	75	13	60	56-71	1%	49	114	105-120	82%
12	78	12	78	14	62	58-73	1%	50	115	107-122	84%
13	81	13	80	15	63	59-74	1%	51	117	108-123	87%
14	83	14	83	16	64	60-75	1%	52	119	109-125	90%
15	86	15	86	17	66	61-76	1%	53	120	111-126	91%
16	89	16	89	18	67	63-78	1%	54	122	112-128	93%
17	92	17	91	19	69	64-79	2%	55	124	114-129	95%
18	94	18	94	20	70	65-80	2%	56	125	115-131	95%
19	97	19	97	21	71	66-82	3%	57	127	117-132	96%
20	100	20	100	22	73	68-83	4%	58	129	118-134	97%
21	103	21	103	23	74	69-84	4%	59	130	120-135	98%
22	105	22	105	24	76	70-85	5%	60	132	121-137	98%
23	108	23	108	25	77	72-87	6%	61	134	123-138	99%
24	111	24	111	26	79	73-88	8%	62	135	125-140	99%
25	114	25	114	27	80	74-89	9%	63	137	126-141	99%
26	117	26	117	28	82	76-91	12%	64	139	128-143	99%
27	119	27	120	29	83	77-92	13%	65	140	129-144	99%
28	122	28	123	30	85	78-93	16%	66	142	131-146	99%
29	125	29	126	31	86	80-95	18%	67	144	132-147	99%
30	128	30	129	32	87	81-96	19%	68	146	134-149	99%
31	131	31	132	33	89	82-98	23%	69	147	135-151	99%
32	134	32	135	34	91	84-99	27%	70	149	137-152	99%
33	137	33	138	35	92	85-100	30%	71	150	138-153	99%
34	140	34	141	36	94	87-102	34%	72	150	138-153	99%
35	142	35	145	37	95	88-103	37%	73	150	138-153	99%
36	145	36	148	38	97	89-104	42%	74	150	138-153	99%
37	148	37	150	39	98	91-106	45%	75	150	138-153	99%
38	150	38	150	40	100	92-107	50%	76	150	138-153	99%