



Universidade de Brasília

Instituto de Psicologia

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

Linha de Pesquisa: Avaliação e Medidas em Psicologia

TESTE NÃO VERBAL DE INTELIGÊNCIA SON-R 6-40:

VALIDAÇÃO E NORMATIZAÇÃO PARA O BRASIL

Renata Manuely Feitosa de Lima

Orientador: Prof. Dr. Jacob Arie Laros

Brasília-DF

Junho de 2018



Universidade de Brasília

Instituto de Psicologia

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

Linha de Pesquisa: Avaliação e Medidas em Psicologia

**TESTE NÃO VERBAL DE INTELIGÊNCIA SON-R 6-40:
VALIDAÇÃO E NORMATIZAÇÃO PARA O BRASIL**

Renata Manuely Feitosa de Lima

Tese apresentada ao Programa de Pós-graduação em Psicologia Social, do Trabalho e das Organizações como requisito parcial para a obtenção de grau de doutorado em Psicologia Social, do Trabalho e das Organizações, sob orientação do Professor Jacob Arie Laros.

Orientador: Prof. Dr. Jacob Arie Laros

Brasília-DF

Junho de 2018

**TESTE NÃO VERBAL DE INTELIGÊNCIA SON-R 6-40:
VALIDAÇÃO E NORMATIZAÇÃO PARA O BRASIL**

Tese de doutorado avaliada pela banca constituída por:

Prof. Dr. Jacob Arie Laros (Presidente)

Universidade de Brasília

Prof. Dra. Elizabeth do Nascimento (Membro)

Universidade Federal de Minas Gerais

Prof. Dr. Ricardo José de Moura (Membro)

Universidade de Brasília

Prof. Dr. Josemberg Moura de Andrade (Membro)

Universidade de Brasília

Profa. Dra. Cristiane Faiad de Moura (Suplente)

Universidade de Brasília

Agradecimentos

Primeiramente, agradeço a Deus, fonte de amor e sabedoria, que nos sustenta frente aos desafios.

De forma muito especial, agradeço ao meu orientador, prof. Laros, por me acompanhar nesse processo (desde a graduação!!) e marcar positivamente um momento muito feliz da minha vida! Sou grata pelas discussões, considerações, indicações de leituras, paciência, pela relação de respeito e confiança que estabelecemos ao longo desses anos. O senhor assumiu seu papel de maneira integral: foi um excelente pesquisador e professor, soube orientar, dirigir e guiar todos os passos dessa pesquisa. Muito obrigada por tudo!

Registro também um agradecimento super especial ao professor Felipe Valentini, que considero um grande mentor, por me auxiliar em diferentes momentos na minha jornada na pós-graduação. Obrigada por estar sempre disponível e por compartilhar todo o seu conhecimento. Você sempre contribuiu muito com meu aprendizado!

Um agradecimento especial as minhas grandes amigas Leticia e Girlene: de diferentes formas, vocês contribuíram de modo crucial para eu ter chegado até aqui! Também agradeço aos meus amigos queridos, Kaká, Joyce, Milton, Fred, Wlad, Tiago e Fernanda, pelo ânimo e pela torcida durante essa jornada.

Ao Cebraspe, especialmente à equipe da Coordenação de Ensino, Pesquisa e Estudos Avançados, por contribuírem muito com meu aprendizado e crescimento profissional!

Ao meu querido e amado esposo, Júnior, por ser meu porto seguro, com quem posso compartilhar minha carreira e minha vida. Sem a sua ajuda, com certeza eu não teria conseguido finalizar este trabalho.

À todos os meus familiares, especialmente minha mãe, irmãos e cunhada pelo apoio, compreensão e incentivo para a realização deste trabalho!

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Fundo SON pelo apoio financeiro.

Finalmente, meu agradecimento sincero e especial aos participantes da pesquisa pela paciência e disponibilidade. Esse trabalho não seria possível sem a participação de vocês!!

Agradecimento aos pesquisadores

É importante destacar que um projeto deste escopo somente poderia ser concluído, com sucesso, com o esforço e a dedicação de uma equipe de pesquisadores. Quando iniciei minhas atividades no laboratório Meta, em janeiro de 2012, estava em andamento um grande projeto que tratava sobre a normatização do teste SON-R 6-40 para o Brasil. Vários projetos foram desenvolvidos por diferentes pesquisadores nos anos posteriores para que conseguíssemos concluir esta grande pesquisa com sucesso.

Em primeiro lugar, temos que mencionar o professor Dr. Felipe Valentini que esteve integrado ao projeto de validação e normatização do SON-R 6-40 durante o período de 2009 a 2012. Ele foi responsável pelo processo de amostragem e pela coordenação da coleta de dados para estabelecer as normas do SON-R 6-40 até 2012. Suas orientações contribuíram imensamente para o sucesso dessa pesquisa.

Registramos também um agradecimento especial à professora Dra. Tatiana Pontrelli Mecca que contribuiu significativamente com a pesquisa, sendo a responsável pela aplicação do teste SON-R 6-40 nas crianças com transtorno do espectro do autismo. Reconhecemos, do mesmo modo, a valiosa ajuda da professora Dra. Lisandra Borges Vieira Lima, que foi a responsável pela aplicação do teste SON-R 6-40 nas crianças e adolescentes com deficiência auditiva.

Agradecimento aos aplicadores

A realização da coleta dos dados só foi possível devido à significativa participação e empenho dos psicólogos listados abaixo. A eles, os nossos sinceros agradecimentos:

Ana Cláudia Fernandes

Andressa Tonini

Annelise Carminatti

Beatriz Marinho

Daniela Carcavila

Eliete Neves

Elisângela Alencar

Elisete Cardoso

Éverson Meireles

Fábio Alves

Hunayara Tavares

Janaína Nogueira

Juliana Damásio

Karoline Ribeiro

Larissa Lemos

Lisandra Borges

Luana Janssen Teixeira

Lucas Donato

Ludiana Moreno

Márcia Santos

Marjorie Silva

Paloma Vilhena

Silvânia Onça

Suiany Ponte

Thelmara Bezerra

Agradecimento às instituições

Nossos agradecimentos às instituições escolares que, gentilmente, receberam a equipe de pesquisa e disponibilizaram tempo e espaço físico para a coleta dos dados.

Rio de Janeiro -RJ

Escola Municipal Padre Hugo
Escola Municipal Benjamin Constant
Escola Estadual Heitor Lira
CEFAP – Escola PMERJ

Belford Roxo – RJ

Escola Municipal Padre Hugo
Escola Estadual Gustavo Barroso

São Gonçalo – RJ

Pericar
Centro Educacional Carmem Aguiar

Santo André – SP

Escola Estadual Ennio Mario Bassalho de Andrade
Escola Estadual Professor José Augusto de Azevedo Antunes

São Caetano – SP

Escola Estadual Alfredo Bukart

Guarulhos – SP

Colégio Farias Guarulhos
Colégio CEAF
Escola Estadual João Crispiniano Soares

Campinas – SP

Escola Municipal Prefeito Humberto Alencar Castelo Branco
Escola Estadual Guido Segalho
Escola Estadual Nely Helena
Escola Estadual Deputado Jamil Gadia
Escola Estadual Dr. Tomas Alves
Escola Estadual Julio de Mesquita
Colégio Integração

Sumaré – SP

Colégio Objetivo
Colégio Adventista
Escola Tancredo Neves

Valinhos – SP

EMEB Jorge Bierrenbach de Castro
Escola Antônio Manoni
Centro Educacional SESI
Escola Estadual Sonia Maria

São Paulo – SP

Colégio Poly Master
Colégio Terra
Escola Estadual Padre Sabóia de Medeiros

Carapicuíba – SP

Instituto Educacional Evolução

Santa Teresa – ES

EMEF Professor Ethevaldo Damazio
EEEFM José Pinto Coelho

Vitória – ES

CMEI Ana Maria Chaves Colares
EMEF Eliane Rodrigues dos Santos
Instituto Federal do Espírito Santo

Canoas – RS

Colégio Guarani
Colégio Max Oderich

Florianópolis – SC

EBM Fernando B. Viegas de Amorim

Belém – PA

Escola Estadual José Alves Maia
Escola Castelo Branco
Escola Dom Pedro

Forquilha – CE

Creche Francisco Figueredo
Escola Antônio José de Loiola Filho
Escola Deputado José Parente Prado

Sobral – CE

Escola Ermílio de Moraes
Associação Perpétuo de Socorro
Colégio Santo Antônio
IDT

Natal – RN

Escola Estadual Berilo Wanderley
Escola Municipal Professor Carlos Bello
Moreno

Cruz do Espírito Santo – PB

Escola Cobé
Escola Fazenda Santa Luzia
Escola Deputado Fernando Milanez
Associação Centro Rural de Formação

João Pessoa – PB

Escola Aruanda
Universidade Federal da Paraíba

Cruz das Almas – BA

Escola Municipal Recanto Feliz
Escola Municipal José Conrado
Colégio Estadual Dr. Lauro Passos

Santo Antônio de Jesus – BA

Escola Cefira Bailão Diniz
Colégio Estadual Renato Machado
Colégio Estadual Prof. Rômulo Almeida

Maragogipe – BA

Escola Municipalizada Hilderico Pinheiro

Escola Municipalizada Luis Eduardo
Margalhães

Salvador – BA

Escola Suzana Wesley
Instituto Municipal de Educação
Prof. José Arapiraca
Colégio Estadual Dalva Matos

Abadiânia – GO

Escola Municipal Os Pequeninos
Colégio Estadual Ozório Rodrigues
Camargo

Anápolis – GO

Escola Walter Bezé
Colégio Virgínio Santillo

Brasília – DF

Escola Classe Octogonal Quadra 8

Patrocínio – MG

Escola Municipal João Beraldo
Escola Estadual Irmã Gislene
Escola Projeto Guri – EJA

Uberlândia – MG

Escola Estadual Frei Egídio Parisi

Belo Horizonte – MG

Colégio Beerlar São Pascoal
Centro Universitário UMA

Sumário

Lista de Figuras	10
Lista de Tabelas	11
Lista de Abreviações	13
Lista de Anexos	14
Resumo	15
<i>Abstract</i>	16
Apresentação geral	17
1. Introdução	19
2. História dos testes SON	29
3. Pesquisa de normatização no Brasil	60
4. Características psicométricas	80
5. Evidências de validade do SON-R 6-40	107
Considerações finais	125
Referências	129
Anexos	140

Lista de Figuras

Figura 1.1 Ilustração do modelo de Spearman e de Thurstone	20
Figura 1.2 Modelo de <i>Gf</i> e <i>Gc</i> de Cattell	21
Figura 1.3 Modelo dos três estratos de Carroll	22
Figura 1.4 Comparação entre os modelos de Carroll e o modelo CHC	24
Figura 1.5 Representação do estrato 2 e estrato 3 do modelo CHC atual	25
Figura 2.1 Autores dos testes SON	29
Figura 2.2 Diferentes versões dos testes SON	30
Figura 2.3 Figuras substituídas no subtteste Categorias com viés cultural	42
Figura 2.4 Figuras do SON-R 5½-17 e as alternativas do SON-R 6-40	48
Figura 2.5 A dificuldade dos itens do teste SON-R 5½-17 ($N = 1.350$)	51
Figura 2.6 Dificuldade dos itens do teste SON-R 6-40 ($N = 1.933$)	52
Figura 2.7 Exemplo de item do subtteste Analogias	56
Figura 2.8 Exemplo de item do subtteste Mosaicos	57
Figura 2.9 Exemplo de item do subtteste Categorias	57
Figura 2.10 Exemplo de item do subtteste Padrões	58
Figura 3.1 Evolução do IDHM no Brasil	61
Figura 3.2 Passos adotados no processo de normatização dos escores brutos	76
Figura 4.1 Escore médio bruto por subtteste em cada faixa etária	90
Figura 4.2 Escore médio bruto por faixa etária	92
Figura 4.3 Modelo estrutural com as estimativas padronizadas	106
Figura 5.1 Distribuição da amostra segundo o nível socioeconômico	112

Lista de Tabelas

Tabela 1.1 Definição das dimensões do segundo estrato do modelo CHC	26
Tabela 2.1 A escala do SON-R 6-40 e sugestão para interpretação	59
Tabela 3.1 Municípios selecionados em cada região do país	62
Tabela 3.2 Grupos de idade da amostra de normatização	63
Tabela 3.3 Exemplo de um elemento da pesquisa de normatização	65
Tabela 3.4 Plano amostral ajustado	66
Tabela 3.5 Perfil dos respondentes da pesquisa de normatização	67
Tabela 3.6 Quantidade de respondentes por municípios	68
Tabela 3.7 Nome dos examinadores e municípios	70
Tabela 3.8 Informações dos respondentes da pesquisa de normatização	71
Tabela 3.9 Estrutura geral dos grupos da amostra normativa	72
Tabela 3.10 Características da administração do SON-R 6-40	73
Tabela 4.1 Dificuldade dos itens dos subtestes do SON-R 6-40	81
Tabela 4.2 Correlação item-total dos itens dos subtestes do SON-R 6-40	83
Tabela 4.3 Estatísticas de ajuste dos modelos de TRI do conjunto de 124 itens dos quatro subtestes do SON-R 6-40	84
Tabela 4.4 Parâmetro b dos itens do SON-R 6-40 segundo a TRI	85
Tabela 4.5 Parâmetro a dos itens do SON-R 6-40 segundo a TRI	87
Tabela 4.6 Estatísticas descritivas do escore bruto por subteste em cada faixa etária (N=1.360)	89
Tabela 4.7 Características do escore total bruto em cada faixa etária (N=1.360)	91
Tabela 4.8 Correlações dos escores brutos com a idade	93
Tabela 4.9 Características dos escores normatizados dos subtestes e escore total (SON-QI) na amostra total e nos três grupos de idade	94

Tabela 4.10 Características do SON-QI nas 17 faixas etárias	95
Tabela 4.11 Valores mínimo e máximo dos escores normatizados dos subtestes e do SON-QI por faixa de idade	96
Tabela 4.12 Fidedignidade dos escores brutos dos quatro subtestes e SON-QI	97
Tabela 4.13 Fidedignidade dos escores normatizados por subteste e SON-QI	98
Tabela 4.14 Médias, desvios-padrão e diferença entre as médias dos escores normatizados no estudo de fidedignidade teste-reteste ($N=77$)	100
Tabela 4.15 Resultado das análises da dimensionalidade do SON-R 6-40	102
Tabela 4.16 Resultado da análise fatorial usando <i>MRFA</i>	103
Tabela 4.17 Índices de ajuste pelo método de estimação da máxima verossimilhança e cargas fatoriais padronizadas ($N=1.360$)	105
Tabela 5.1 Estatísticas descritivas do SON-QI médio dos respondentes por região e localidade	108
Tabela 5.2 Médias nos subtestes e no SON-QI por sexo	109
Tabela 5.3 Grau de escolaridade dos pais ou responsáveis	110
Tabela 5.4 Cargas fatoriais do questionário de nível socioeconômico (NSE)	111
Tabela 5.5 Estatísticas descritivas do SON-QI por examinador ($N=1.237$)	113
Tabela 5.6 Avaliação do examinador e desempenho no teste	114
Tabela 5.7 Tempo de administração do teste completo e por subteste	115
Tabela 5.8 Coeficientes de correlação entre os subtestes do SON-R 6-40 e da Leiter-R e entre o QI de ambos os instrumentos	117
Tabela 5.9 Coeficientes de correlação entre o escore de QI do teste SON-R 6-40 e os escores de outros testes de inteligência normatizados no Brasil	119
Tabela 5.10 Tipo de deficiência auditiva e tipo de comunicação utilizada pelos participantes ($N=21$)	120
Tabela 5.11 Diferenças entre os escores normatizados do grupo com deficiência auditiva e da amostra de normatização brasileira do SON-R 6-40	121
Tabela 5.12 Diferenças entre os escores normatizados do grupo com TEA e da amostra de normatização brasileira do SON-R 6-40	123

Lista de Abreviações

Fator <i>g</i>	Fator geral
<i>Gf</i>	Inteligência fluida
<i>Gc</i>	Inteligência cristalizada
<i>CHC</i>	Modelo Cattell-Horn-Carroll
SON-EE	Escala de Execução do SON
SON-ER	Escala de Raciocínio do SON
SON-QI	QI geral do SON
ANA	Subteste Analogias
MOS	Subteste Mosaicos
CAT	Subteste Categorias
PAD	Subteste Padrões
IDHM	Índice de Desenvolvimento Humano Municipal
TCT	Teoria Clássica dos Testes
TRI	Teoria de Resposta ao Item
MRFA	Minimum Rank Factor Analysis
CFA	Análise Fatorial Confirmatória
AIC	Akaike Information Criterion
Parâmetro <i>b</i>	Parâmetro de dificuldade da TRI
Parâmetro <i>a</i>	Parâmetro de discriminação da TRI
<i>d</i>	Tamanho de efeito <i>d</i> de Cohen
Standards	Standards for Educational and Psychological Testing
NSE	Nível socioeconômico
TEA	Transtorno do Espectro do Autismo

Lista de Anexos

Anexo 1 – Carta de apresentação da pesquisa SON-R 6-40 para escolas	141
Anexo 2 – Carta do comitê de ética	142
Anexo 3 – Termo de informação aos pais ou responsáveis	143
Anexo 4 – Termo de consentimento	144
Anexo 5 – Folha de resposta do teste SON-R 6-40	145
Anexo 6 – Questionário contextual	147

RESUMO

Os objetivos principais desta tese foram estabelecer normas brasileiras para o teste não-verbal de inteligência SON-R 6-40 e obter evidências de validade dos escores do teste no Brasil. Para alcançar os objetivos, foram executados vários estudos com o SON-R 6-40, originalmente desenvolvido na Holanda. A pesquisa maior, tanto em termos de tempo quanto em termos de número de sujeitos envolvidos, foi a de normatização. Para desenvolver normas brasileiras, o teste foi administrado individualmente em uma amostra de 1.360 pessoas com idades entre 6 e 37 anos. As pessoas que participaram do estudo eram residentes de 40 municípios de 13 estados diferentes de todas as cinco regiões brasileiras. Foram estabelecidos 17 grupos de idade e cada grupo foi representado por uma amostra de 80 pessoas, que foi estratificada segundo o sexo, a situação educacional e variáveis demográficas. Foi utilizado o método de normatização contínua para calcular as normas onde as distribuições dos escores para todos os grupos de idade foram ajustadas simultaneamente como uma função contínua da idade. As principais vantagens desse método são que os escores normatizados podem ser calculados com base na idade exata e que o uso da informação de todos os grupos de idade simultaneamente torna a normatização mais precisa. Após esta fase, o coeficiente de fidedignidade dos subtestes foi estimado, apresentado valores médios entre 0,88 e 0,90. A dimensionalidade do SON-R 6-40 foi examinada por meio de quatro métodos: Método *Hull*, Análise Paralela *Optimal Implementation*, Análise Paralela *Classical Implementation* e Schwarz's *Bayesian Information Criterion Dimensionality Test*. Todos os índices de dimensionalidade dos quatro métodos utilizados indicaram unidimensionalidade. A estrutura fatorial foi examinada por meio de Análise Fatorial Exploratória e Confirmatória: os resultados obtidos convergiram com o estudo original realizado na Holanda indicando que o teste possui um único fator geral. Simultaneamente com a pesquisa de normatização, foram realizados outros estudos. O primeiro estudo ($N=77$) foi para investigar a fidedignidade teste-reteste do SON-R 6-40: os resultados indicaram alta fidedignidade teste-reteste ($r=0,89$). O segundo estudo ($N=21$) foi realizado com pessoas com deficiência auditiva e os resultados indicaram que a diferença média do QI deste grupo em comparação com o grupo de normatização foi ao redor de um desvio-padrão. O terceiro estudo ($N=37$) foi conduzido com pessoas com transtorno do espectro do autismo: os resultados indicaram que o SON-R 6-40 é atrativo para pessoas que são difíceis de avaliar. No quarto estudo ($N=21$), o teste foi administrado juntamente com a Leiter-R, um teste não-verbal de inteligência, e a alta correlação ($r=0,93$) entre os dois testes sugere forte evidência de validade convergente do SON-R 6-40. Finalmente, a tese apresenta, brevemente, estudos desenvolvidos em uma fase anterior com o SON-R 6-40 no Brasil, a fim de reunir informações psicométrica necessárias para o usuário. Após a conclusão dos diversos estudos, a presente tese fornece aos psicólogos brasileiros um instrumento não-verbal com normas brasileiras recentes e com amplas evidências de validade para avaliar as habilidades cognitivas de crianças e adultos.

Palavras-chave: normatização, validade, SON-R 6-40, teste de inteligência, teste não-verbal.

ABSTRACT

The main objectives of this thesis were to establish Brazilian norms for the SON-R 6-40, a non-verbal intelligence test originally developed in the Netherlands, and to obtain validity evidence of the scores on this instrument in Brazil. To achieve these objectives, several studies were carried out with this instrument. The most extensive research, both in terms of time and in terms of the number of subjects involved, was the standardization study. To develop Brazilian norms, the test was applied individually to a representative sample of 1,360 people with ages varying between 6 years and 37 years. The subjects of the standardization sample were residents of 40 municipalities from 13 Brazilian states and from all five Brazilian regions. Seventeen age groups were established and each group was represented by a sample of 80 people, which was stratified according to sex, educational situation and demographic variables. A continuous method was used to calculate norms in which the distributions of scores for all age groups were adjusted simultaneously as a continuous function of age. The main advantages of this method are that the norms can be calculated on basis of the exact age and that the norms are more precise. After this phase, the reliability coefficients of the subtests were estimated, presenting mean values between .88 and .90. The dimensionality of the SON-R 6-40 was examined using four methods: Hull Method, Parallel Analysis Optimal Implementation, Parallel Analysis Classical Implementation and Schwarz's Bayesian Information Criterion Dimensionality Test. All dimensionality indices of the four methods indicated unidimensionality. The factor structure was examined by Exploratory and Confirmatory Factor Analysis: the obtained results converge with the original Dutch study indicating that the test measures a single general factor. Simultaneously with the standardization research various other studies were conducted. The first of these ($N=77$) was carried out to investigate the test-retest reliability of SON-R 6-40: the results indicated a high test-retest reliability ($r=.89$). The second study ($N=21$) was carried out with hearing impaired subjects: the results indicated that their mean IQ-score was about one standard deviation below that of the standardization sample. The third study ($N=37$) was conducted with people with an autism spectrum disorder: the results indicated that the SON-R 6-40 is an attractive instrument for this group that is difficult to test. In the fourth study ($N= 21$) the test was administered together with the Leiter-R, a non-verbal intelligence test, and the very high correlation ($r =.93$) between the two tests suggested strong evidence of the convergent validity of the SON-R 6-40. Finally, the thesis presents, briefly, a number of studies performed in an earlier phase with the SON-R 6-40 in Brazil, in order to gather necessary psychometric information for the user. After concluding the above mentioned studies, the present thesis provides Brazilian psychologists with a non-verbal instrument with recent Brazilian norms and with ample validity evidence to assess the cognitive abilities of children and adults.

Key words: standardization, validity, SON-R 6-40, intelligence test, non-verbal test.

Apresentação geral

O objetivo desta pesquisa foi normatizar o teste de inteligência SON-R 6-40 para o contexto brasileiro e reunir evidências de validade. O teste, composto por quatro subtestes, é destinado para avaliação das habilidades cognitivas de pessoas entre 6 e 40 anos de idade e pode ser aplicado sem instruções verbais, o que facilita a avaliação de pessoas não ouvintes ou com dificuldades de comunicação.

Normatizar refere-se à definição de critérios ou padrões de interpretação do escore que o indivíduo recebe ao responder um teste e também à avaliação da fidedignidade (Pasquali, 2011; Kittrell, 1999). Assim, um escore bruto precisa ser interpretado, contextualizado para alcançar um significado, um sentido. Geralmente, três padrões são aplicados no ato de normatização de um teste: (1) normas de desenvolvimento; (2) normas intragrupo e (3) normas referentes ao critério. Nessa pesquisa, foram adotadas as normas intragrupo, isto é, o escore bruto toma sentido em relação aos resultados dos demais do grupo de referência utilizado. No contexto de testes de inteligência, o grupo de referência são as pessoas com a mesma idade.

Para organizar e facilitar a leitura e apresentação dos estudos realizados, esse material foi dividido em cinco capítulos. O primeiro capítulo, situa o leitor quanto às concepções e modelos teóricos da inteligência. O segundo apresenta a história do desenvolvimento dos testes SON e os passos que guiaram o desenvolvimento do teste SON-R 6-40.

O capítulo três apresenta o delineamento aplicado no estudo de normatização no Brasil, descrevendo os critérios utilizados para a seleção dos participantes, treinamento dos aplicadores, características gerais do grupo normativo e o modelo de normatização. O capítulo quatro refere-se à apresentação das características psicométricas do teste. Neste capítulo são apresentadas as análises dos itens, as análises de dimensionalidade,

fidedignidade, estrutura fatorial e dados sobre os escores normatizados. Por fim, o objetivo do capítulo cinco é analisar a relação dos escores normatizados do teste SON-R 6-40 com outras variáveis obtidas na pesquisa de normatização, por exemplo, grau de escolaridade dos pais, duração da testagem, diferenças regionais, e apresentar os estudos de validação realizados no Brasil.

Espera-se que a presente pesquisa contribua com a área da avaliação psicológica no Brasil, que passou por uma grande mudança nos últimos quinze anos após a publicação da Resolução 002/2003 (Primi & Nunes, 2010). Além disso, espera-se que o atual trabalho contribua com a formulação de novas pesquisas na área de normatização de testes psicológicos.

1. Introdução

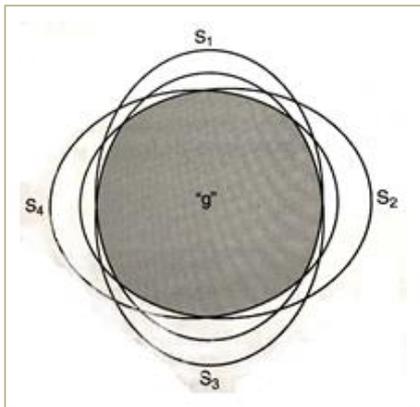
A inteligência é o construto mais pesquisado na história da Psicologia e apresenta amplo poder explicativo de diferentes comportamentos (Flanagan & Harrison, 2012; Neisser et al., 1996). Quando bem realizada, a mensuração da inteligência é capaz de fornecer dados acerca das potencialidades e fragilidades que podem ser utilizados em intervenções para ajudar as pessoas a alcançarem seu potencial de forma eficaz.

A história da avaliação psicológica indica Binet e Simon como os primeiros autores de um teste de inteligência que, interessados em detectar atraso intelectual, desenvolveram um teste composto por 30 itens que visavam avaliar o nível de inteligência de crianças e adultos (Pasquali, 2011). A tentativa de Binet e Simon era elaborar um teste de conteúdo mais cognitivo (não sensorial) e cobrindo funções mais amplas (não específicas) e holísticas, tais como empregar palavras, encontrar conexões e apreender o significado (Pasquali, 2011; Hogan, 2006).

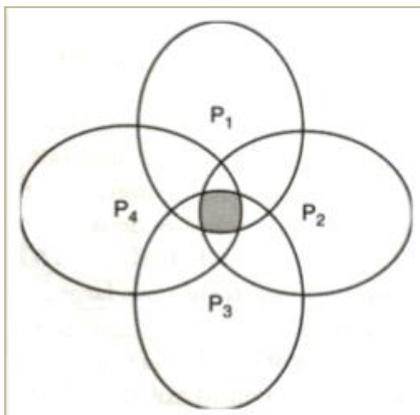
A conceituação da inteligência sofreu alterações e refinamentos com o passar do anos. Wechsler (1958) definiu inteligência como a capacidade do indivíduo agir propositalmente, pensar racionalmente e lidar efetivamente com seu ambiente (Wechsler, 1958 in Flanagan & Harrison, 2012). Almeida (1994) afirma que a inteligência está associada à capacidade para aprender relações, utilizando conhecimentos prévios ou apenas raciocínio. Antonio, Mecca e Macedo, (2012) definem a inteligência como a capacidade de pensar abstratamente, resolver problemas e aprender com a experiência.

Quanto às teorias fatoriais da inteligência, o primeiro modelo teórico foi concebido por Spearman (1927), que entendia que toda atividade intelectual era derivada de um fator geral (*g*), que é parcialmente utilizado por todas as outras capacidades mentais, e fatores específicos (*s*). A suposição deste autor era que o fator *g*, e não *s*, proporcionava uma melhor previsão de inteligência global (Cohen, Swerdlik, & Sturman, 2014).

Em contraste com a visão de Spearman, muitos modelos multifatoriais da inteligência foram propostos. O modelo das capacidades mentais primárias de Thurstone (1938), afirmava que a inteligência era composta de sete capacidades primárias e que as correlações entre esses fatores eram baixas. A Figura 1.1 apresenta uma representação da Teoria de Spearman e de Thurstone.



Modelo de Spearman



Modelo de Thurstone

Figura 1.1 Ilustração do modelo de Spearman e de Thurstone. Fonte: Hogan (2006)

Na Figura 1.1, a letra “P” significa *Primary Factor*, que é um fator que é independente dos demais. É possível observar na ilustração que, enquanto Spearman afirmava que as correlações entre os diferentes testes eram altas o suficiente para se pensar que elas estavam mensurando um mesmo fator comum, Thurstone acreditava que essas correlações eram baixas o suficiente para se pensar que elas mensuraram diversos fatores bastante independentes (Hogan, 2006).

O terceiro modelo acerca da estrutura da inteligência que obteve bastante influência sobre os demais é o chamado modelo da inteligência fluida (*Gf*) e cristalizada (*Gc*), proposto inicialmente por Cattell (1941), mas depois modificado por Horn (1985). Os autores do modelo afirmavam que tanto *Gf* quanto *Gc* eram fatores correlacionados, mas distintos, compostos por diferentes fatores mais específicos e que não existia um fator geral acima das capacidades *Gf* e *Gc* (Seabra, Laros, Macedo, & Abreu, 2014; Hogan 2006; Schelini, 2006; Horn & Noll, 1997). A Figura 1.2 apresenta uma representação desse modelo.

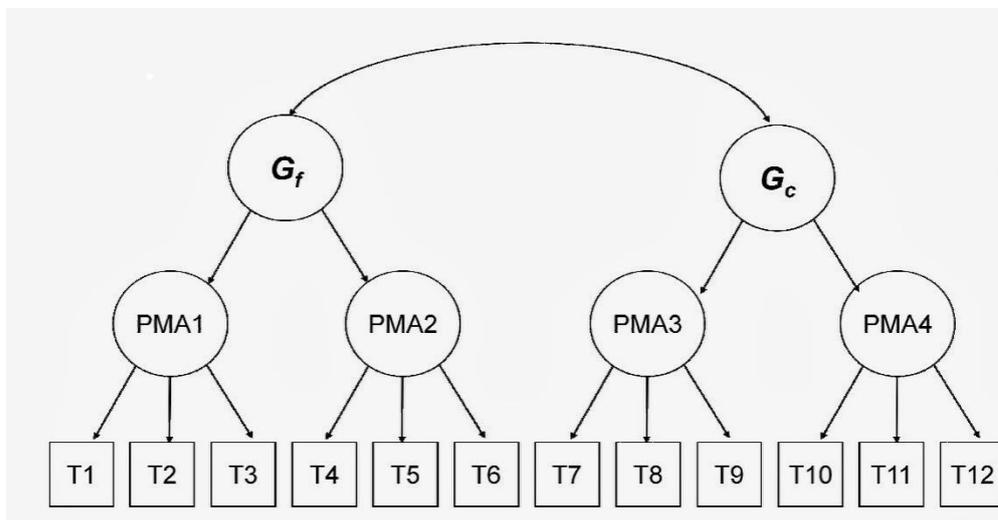


Figura 1.2 Modelo de *Gf* e *Gc* de Cattell.

Cattell (1971) e Horn (1976) descreveram a inteligência fluida como uma capacidade pouco dependente de conhecimentos adquiridos previamente, onde as operações mentais não podem ser executadas automaticamente, e que está associada a componentes não-verbais (Flanagan & Harrison, 2012). Em contraste, a inteligência cristalizada refere-se ao conjunto de conhecimentos adquiridos aprendidos por transmissão cultural ou pela escola; inclui habilidades como compreensão verbal ou de reconhecimento do significado das palavras (Seabra, Laros, Macedo, & Abreu, 2014).

Levando em consideração o trabalho realizado por Cattell e Horn, John B. Carroll propôs um modelo hierárquico das habilidades cognitivas, acrescentando o fator de inteligência geral no estrato mais alto e incluindo outras habilidades mais amplas. A Figura 1.3 apresenta uma representação do modelo.

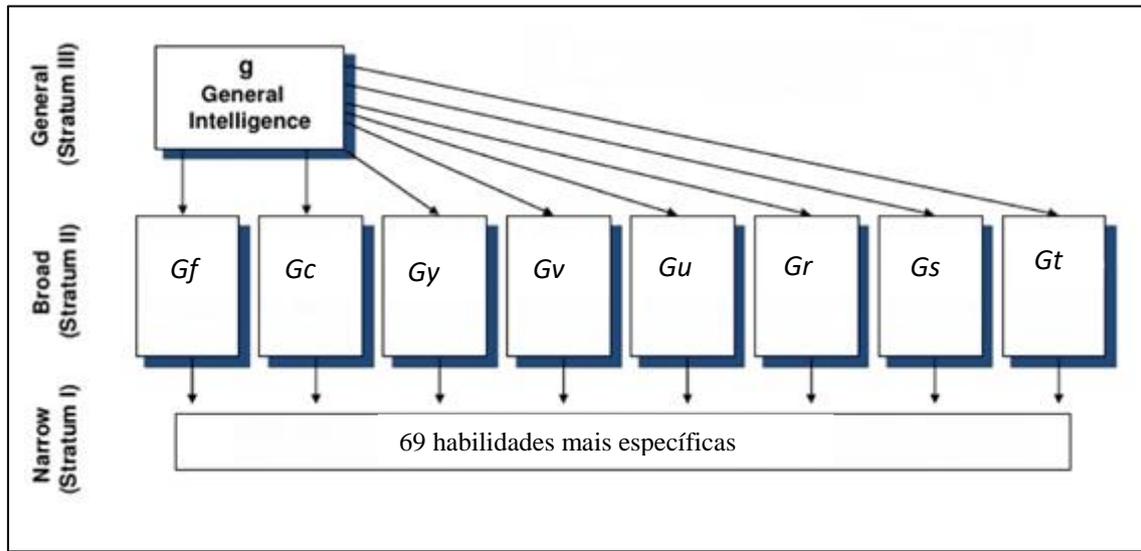


Figura 1.3 Modelo dos três estratos de Carroll

O modelo dos três estratos de Carroll (1993), organizou os diferentes fatores da inteligência em três níveis hierárquicos: o estrato mais alto é composto pelo fator *g*. O segundo estrato agrupa oito aptidões mais amplas: inteligência fluida (*Gf*), inteligência cristalizada (*Gc*), memória geral e aprendizagem (*Gy*), percepção visual abrangente (*Gv*), percepção auditiva abrangente (*Gu*), habilidade de retenção abrangente (*Gr*), rapidez cognitiva abrangente (*Gs*) e velocidade de processamento (*Gt*). Por fim, o estrato mais baixo era composto por cerca de 69 fatores mais específicos, chamados frequentemente de habilidades mentais primárias (Carroll, 1993; Flanagan & Harrison, 2012). Flanagan e Harrison (2012) afirmam que, apesar de algumas das habilidades localizadas no segundo estrato serem um pouco mais relacionadas a determinantes genéticos do que outras, não há uma clara distinção entre as consequências dos fatores genéticos e fatores ambientais, pois esses fatores se misturam ao longo do desenvolvimento.

O modelo da Carroll contribuiu de forma significativa com a organização e a compreensão das dimensões da inteligência e sintetizou os resultados mais importantes da psicometria (Flanagan & Harrison, 2012; Seabra, Laros, Macedo, & Abreu, 2014). Entretanto, apesar da enorme contribuição para a área, o modelo de Carroll possuía algumas limitações, por exemplo, as inferências de Carroll sobre as relações entre diferentes fatores de diferentes níveis (estratos) emergiram de dados derivados de um conjunto diversificado de estudos e amostras amplamente independentes (McGrew, 2005). Nenhum dos conjuntos de dados de Carroll incluiu a amplitude necessária de variáveis para avaliar, em uma única análise, a estrutura geral de seu modelo proposto de três estratos. Nesse contexto, o modelo Cattell-Horn-Carroll (CHC) foi proposto no intuito de ampliar a teoria dos estratos de Carroll. Na realidade, o modelo CHC, proposto por Flanagan e McGrew e (1998), integra os dois modelos teóricos mais proeminentes da psicometria: o modelo de Cattell e Horn e o modelo de Carroll (McGrew 2005; 2009).

Nos últimos 20 anos, a teoria CHC alcançou um status dentro da psicometria de melhor modelo atual da estrutura da inteligência humana (Kaufman, 2009; Keith & Reynolds, 2010). O modelo também é organizado em três estratos. Porém, uma das diferenças básicas entre o modelo CHC e o modelo de Carroll, foi a adição de novos fatores no segundo estrato (Flanagan & Harrison, 2012). A Figura 1.4 apresenta uma ilustração do modelo.

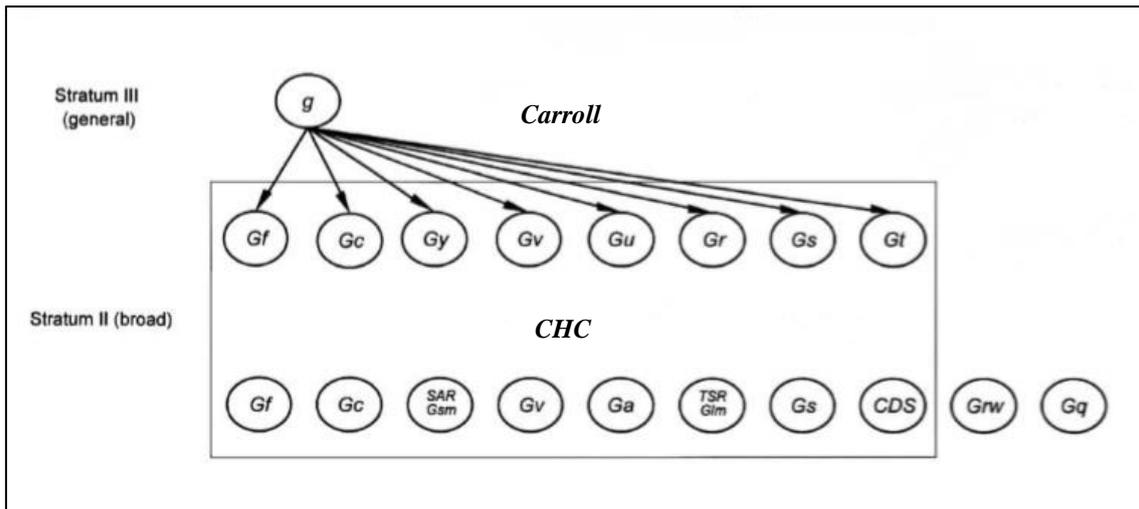


Figura 1.4 Comparação entre os modelos de Carroll e o modelo CHC

A Figura 1.4 destaca a diferença no segundo estrato entre os dois modelos. O segundo estrato do modelo CHC era composto, inicialmente, por 10 dimensões mais amplas: inteligência fluida (*Gf*), inteligência cristalizada (*Gc*), memória de curto prazo (*Gsm*), processamento visual (*Gv*), processamento auditivo (*Ga*), memória de longo prazo (*Glm*), rapidez cognitiva (*Gs*), velocidade de decisão correta (*CDS*), leitura/escrita (*Grw*) e conhecimento quantitativo (*Gq*). Entretanto, após uma revisão do modelo, o segundo estrato possui agora 16 dimensões (Schneider & Newman, 2015; McGrew, 2009). A Figura 1.5 apresenta uma representação do modelo atual.

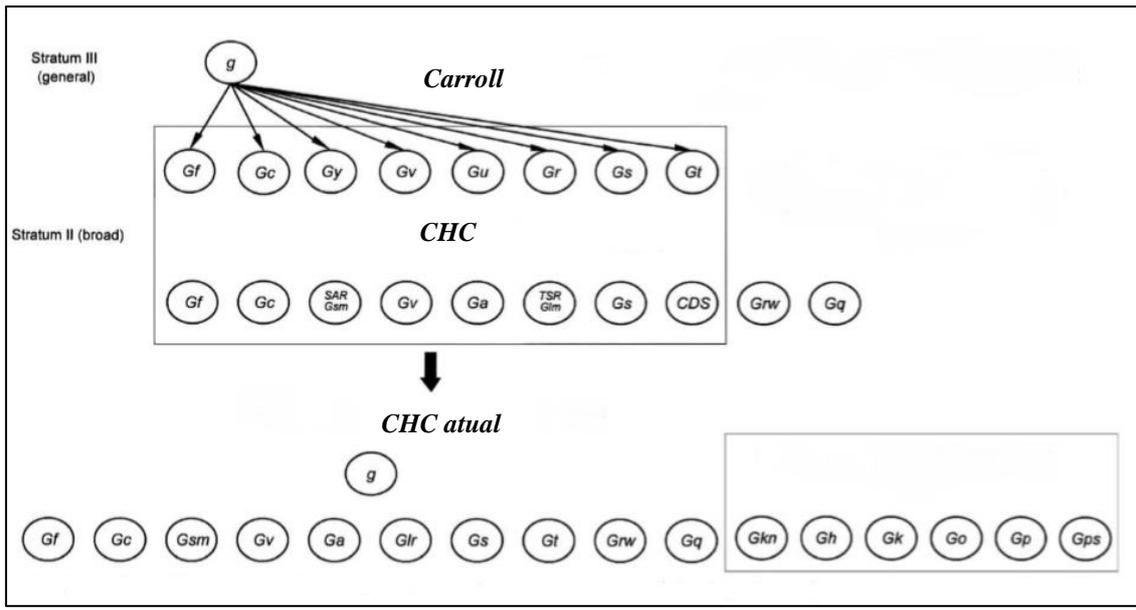


Figura 1.5 Representação do estrato 2 e estrato 3 do modelo CHC atual.
 Notas: O estrato I, que contempla mais de 70 habilidades, não é representado aqui.

No modelo atual, o segundo estrato possui as seguintes dimensões: inteligência fluida (*Gf*), compreensão/conhecimento (*Gc*), memória de curto prazo (*Gsm*), processamento visual (*Gv*), processamento auditivo (*Ga*), memória de longo prazo (*Glr*), rapidez cognitiva (*Gs*), velocidade de reação e decisão (*Gt*), leitura/escrita (*Grw*), conhecimento quantitativo (*Gq*), domínio/conhecimento específico (*Gkn*), habilidades táteis (*Gh*), habilidades cinestésicas (*Gk*), habilidades olfativas (*Go*), habilidades psicomotoras (*Gp*) e velocidade psicomotora (*Gps*). A Tabela 1.1 apresenta as definições de cada dimensão.

Tabela 1.1*Definição das dimensões do segundo estrato do modelo CHC*

Dimensão	Símbolo	Definição
1. Inteligência fluida	<i>Gf</i>	Capacidade para resolver problemas novos que não podem ser executados automaticamente. Inclue tarefas relacionadas à indução, formação de conceitos, identificação de relações, compreensão de implicações, resolução de problemas e transformação de informações. A <i>Gf</i> está associada às seguintes habilidades específicas (estrato I): de indução, raciocínio sequencial geral e raciocínio quantitativo.
2. Compreensão/conhecimento	<i>Gc</i>	É descrita como a capacidade que é desenvolvida/incorporada durante a educação formal e informal e pelas experiências da vida. Está associada ao desenvolvimento da linguagem, informação verbal geral, conhecimento lexical, compreensão de um discurso, entre outros.
3. Memória de curto prazo	<i>Gsm</i>	É descrita como a habilidade para reter informações por um curto momento (eventos que ocorreram no último minuto, por exemplo). Está associada às seguintes habilidades específicas: tempo de memória e memória de trabalho.
4. Processamento visual	<i>Gv</i>	Capacidade de gerar, armazenar, recuperar e transformar imagens visuais. Pode ser definida como a habilidade para manipular mentalmente imagens ou figuras. Está associada às seguintes habilidades específicas: visualização, relações espaciais, memória visual, entre outros.
5. Processamento auditivo	<i>Ga</i>	Habilidade de processar informações auditivas. Envolve uma ampla gama de habilidades envolvidas na interpretação e organização de sons. Está relacionada às seguintes habilidades específicas: codificação fonética, discriminação dos sons da fala, julgamento e manutenção do ritmo, entre outros.
6. Memória de longo prazo	<i>Glr</i>	Capacidade de armazenar, consolidar e recuperar informações continuamente ao longo do tempo, por exemplo, recuperar conceitos, ideias, nomes, etc. Está associada às seguintes habilidades específicas: memória associativa, memória de significados, fluência de ideias, criatividade, entre outros.
7. Rapidez cognitiva	<i>Gs</i>	Capacidade de executar de maneira rápida e eficiente tarefas cognitivas simples ou repetitivas. Está associada às seguintes habilidades específicas: velocidade perceptual, velocidade de raciocínio, velocidade de leitura, entre outros.
8. Velocidade de reação e decisão	<i>Gt</i>	Capacidade de tomar decisões simples de forma rápida. Está associada às seguintes habilidades específicas: tempo de reação, velocidade de processamento semântico, velocidade de comparação mental, entre outros.
9. Leitura/escrita	<i>Grw</i>	Capacidade de ler e compreender um texto, além de escrever uma história complexa conectada. Está associada às seguintes habilidades específicas: leitura de decodificação, compreensão da leitura, capacidade de escrita, entre outros.

Tabela 1.1 (continuação)

Dimensão	Símbolo	Definição
10. Conhecimento quantitativo	<i>Gq</i>	Está associado ao conhecimento matemático. Está associada às seguintes habilidades específicas: conhecimento matemático e desempenho em matemática.
11. Domínio/conhecimento específico	<i>Gkn</i>	Refere-se à domínios de conhecimentos especializados e profundos desenvolvidos através da prática regular e esforço continuado. Está associada às seguintes habilidades específicas: proficiência em língua estrangeira, habilidade em leitura labial, conhecimento mecânico, entre outros.
12. Habilidades táteis	<i>Gh</i>	Refere-se ao processamento de informações por meio do tato. Os aspectos cognitivos e perceptivos deste domínio ainda não foram amplamente investigados. Envolve sensibilidade tátil.
13. Habilidades cinestésicas	<i>Gk</i>	Capacidade de localização corporal, coordenação de movimentos corporais, incluindo o falar, o andar, expressões faciais, gestos e postura. Os aspectos cognitivos e perceptivos deste domínio ainda não foram amplamente investigados. Envolve sensibilidade cinestésica.
14. Habilidades olfativas	<i>Go</i>	Refere-se à capacidade de perceber odores. Os aspectos cognitivos e perceptivos deste domínio ainda não foram amplamente investigados. Envolve memória olfativa e sensibilidade olfativa.
15. Habilidades psicomotoras	<i>Gp</i>	Capacidade de realizar movimentos corporais com precisão. Está associada às seguintes habilidades específicas: força estática, destreza manual, destreza dos dedos, mira, entre outros.
16. Velocidade psicomotora	<i>Gps</i>	Capacidade de executar movimentos corporais com rapidez e facilidade. Está associada às seguintes habilidades específicas: velocidade de escrita, tempo de movimento, entre outros.

Notas. Fonte: Flanagan e Harrison (2012); Seabra, Laros, Macedo, & Abreu, 2014.

Em relação ao fator de inteligência fluida (*Gf*), na teoria CHC, *Gf* é uma habilidade associada aos processos de atenção e memória (*Gsm*, *Gs*, *Glr* e *Gf*), apresentando uma relação próxima com a memória de trabalho devido ao fato que o raciocínio indutivo é facilitado quando mais informações são recuperadas rapidamente (Schneider & Newman, 2015). Há estudos que destacam a importância de mensurar esse fator por ele prever resultados importantes, por exemplo, apresentar uma forte relação com o fator *g* (Carroll, 2003; Floyd, Evans, & McGrew, 2003; Gustafsson, 1984). Floyd et al., (2003) indica que

Gf é um ótimo indicador de habilidades matemáticas. E outros estudos indicam que *Gf* está presente em tarefas que envolvem a aquisição de conhecimento e habilidades em diferentes domínios (Valentin Kvist & Gustafsson, 2008).

Levando em consideração esses resultados e visando contribuir com a área da avaliação psicológica brasileira, especificamente, na área da mensuração da inteligência, o objetivo geral da tese foi estabelecer normas brasileiras para o teste não-verbal de inteligência SON-R 6-40 e obter evidências de validade dos escores do teste. O teste SON-R 6-40 busca mensurar a denominada inteligência fluida (*Gf*) e os resultados desta pesquisa indicam que tanto o processamento visual (*Gv*) quanto o fator memória de longo prazo (*Glr*) desempenham um papel importante nos quatro subtestes do teste SON-R 6-40, apresentando consonância com a literatura da área (Schneider & Newman, 2015).

Por fim, retornando a questão sobre a estrutura da inteligência, o atual estado da ciência indica que a inteligência humana é uma mistura de muitas habilidades inter-relacionadas de várias maneiras (Schneider & Newman, 2015; Flanagan & Harrison, 2012). As habilidades e suas inter-relações são determinadas por muitas influências endógenas (genéticas, fisiológicas, neurológicas) e exógenas (experenciais, nutricionais e higiênicas). Além disso, essas influências podem ser mais ou menos potentes em alguns períodos de desenvolvimento do que em outros (Flanagan & Harrison, 2012).

2. História dos Testes SON

Os testes de inteligência não-verbal Snijders-Oomen são instrumentos de avaliação individual da inteligência onde o uso da linguagem falada e escrita não são necessários. São testes publicados primeiramente na Holanda e ficaram sob a responsabilidade dos professores Nan Snijders-Oomen e Jan Snijders por um período de cinquenta anos. Atualmente, os professores Jacob Arie Laros e Peter Tellegen são os pesquisadores responsáveis pelo desenvolvimento dos testes SON.



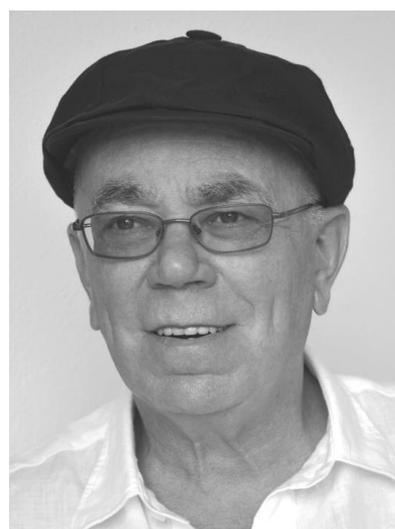
Nan Snijders-Oomen



Jan Snijders



Jacob Arie Laros



Peter Tellegen

Figura 2.1 Autores dos testes SON

Diferentes versões dos testes SON foram surgindo com o passar dos anos. A Figura 2.2 apresenta todas as versões dos testes que foram desenvolvidas nos últimos setenta anos.

SON (1943) Normas para crianças surdas com idade entre 4 e 14 anos Snijders-Oomen	
SON-'58 (1958) Normas para crianças surdas e ouvintes com idade entre 4 e 16 anos Snijders & Snijders-Oomen	
SON 2½-7 – Pré-escolar (1975) Normas para crianças surdas e ouvintes com idade entre 3 e 7 anos Snijders & Snijders-Oomen	SSON (1975) Normas para pessoas surdas e ouvintes com idade entre 7 e 17 anos Starren
SON-R 2½-7 (1998) Normas gerais para crianças com idade entre 2½ e 7 anos Tellegen, Winkel, Wijnberg-Williams & Laros	SON-R 5½-17 (1988) Normas gerais para pessoas com idade entre 5½ e 17 anos Snijders, Tellegen & Laros
SON-R 2–8 (2017) Normas holandesas para crianças com idade entre 2 e 8 anos Tellegen & Laros	SON-R 6-40 (2011) Normas gerais para pessoas com idade entre 6 e 40 anos Tellegen & Laros
SON-R 2½-7[a] (2015) Normas gerais brasileiras para crianças com idade entre 2½-7 anos Laros, Tellegen, Jesus & Karino	SON-R 6-40 (2018) Normas gerais brasileiras para pessoas com idade entre 6 e 40 anos Laros, Tellegen & Lima

Figura 2.2 Diferentes versões dos testes SON.

Notas. A última linha mostra as versões dos testes SON com normas brasileiras e com evidências de validade para Brasil.

A primeira versão dos testes SON foi publicada em 1943 pela autora Nan Snijders-Oomen e era voltado somente para crianças surdas (Snijders-Oomen, 1943). Nan Snijders-Oomen trabalhou no instituto para crianças surdas em Sint Michielgestel na Holanda. Ao utilizar os testes disponíveis para testar crianças surdas, ela chegou à conclusão que esses testes não estavam adequados. Nessa época poucos testes estavam disponíveis: uma adaptação holandesa do teste do Binet de 1905 (o Herderschêe-Binet) e uma adaptação do teste para Desenvolvimento de Raciocínio originalmente construído

nos Estados Unidos. A adaptação holandesa do teste do Binet foi muito criticada: as principais críticas foram que o teste era muito verbal, fazia grande apelo para o conhecimento escolar e colocava crianças de níveis socioeconômicos baixos em uma posição de desvantagem.

Assim, Snijders-Oomen se propôs a desenvolver um teste cujo objetivo era prover uma indicação clara da habilidade de aprendizagem da criança surda e sua perspectiva de sucesso no sistema educacional. Um requisito do teste era que o resultado deveria ser influenciado o mínimo possível pelo conhecimento e educação da criança.

Para desenvolver o primeiro teste SON os seguintes critérios foram estabelecidos (Snijders-Oomen, 1943):

- (1) Para aproximar a versatilidade de inteligência o teste deveria examinar uma ampla variedade de funções intelectuais;
- (2) Dentro de cada subteste tinha que ser estabelecido um grau crescente de dificuldade;
- (3) As tarefas no teste tinham que ser interessantes para as crianças, para evitar o tédio e a relutância que podem impedir a criança mostrar do que ela é capaz;
- (4) Os resultados nas tarefas precisavam apresentar uma diferença notável entre os grupos etários;
- (5) Educação e os resultados da educação deveriam afetar minimamente o desempenho no teste.

Snijders-Oomen deliberou que para poder fazer afirmações sobre a aptidão escolar das crianças, não seria suficiente ter uma avaliação global da inteligência: um teste também deveria ser capaz de informar sobre talentos especiais e sobre fraquezas das crianças (Drunen, 2003). Desta forma, ela identificou quatro aspectos que são importantes para a investigação do funcionamento cognitivo de crianças surdas: (1) *Percepção de*

Formas; (2) Compreensão da Coerência Concreta, (3) Capacidade de Abstração e (4) a Concentração de Atenção e a Memória de Curto Prazo.

No SON-'43 um subteste de memória foi incluído devido ao fato de que a memória é um aspecto importante no processo de aprendizagem. Entretanto, o subteste de memória mostrou uma relação tão fraca com a inteligência geral que este não foi incluído na determinação da idade mental (Snijders-Oomen, 1943).

O que principalmente norteou o trabalho de Nan Snijders-Oomen não eram considerações teóricas ou metodológicas, mas uma orientação clínica da questão como poderia ser obtida uma boa indicação de oportunidades de aprendizagem das crianças surdas e suas oportunidades de fazer uma trajetória escolar bem-sucedida. A orientação clínica de Snijders-Oomen foi refletida na sua ênfase à importância de observação durante a administração do teste. Dentro da tradição de testagem nos Estados Unidos, a observação desempenhou um papel mínimo: toda a ênfase estava na determinação objetiva do resultado do teste. Snijders-Oomen escreveu no manual do primeiro teste SON: "Inteligência não é o único fator que determina o desempenho em um teste de inteligência. Timidez, sentimento de inferioridade, fraqueza e imprudência podem fazer com que uma pessoa apresente um desempenho abaixo dos limites de suas capacidades" (Snijders-Oomen, 1943, p. 67).

Então, durante a administração do primeiro teste SON, foram observados os seguintes aspectos das atitudes de comportamento de crianças: (1) sociabilidade; (2) autoconfiança; (3) atenção; e (4) a maneira de trabalhar. Na folha de registro do teste foi reservado um espaço para anotar as observações nestas quatro categorias.

O primeiro teste SON pode ser descrito como um teste diagnóstico-clínico. Com base nos quatro aspectos cognitivos básicos foram desenvolvidos os seguintes subtestes para o SON-'43: (1) Percepção de forma (2) Coerência concreta, (3) Ordenamento lógico,

(4) Abstração e (5) Atenção e memória instantânea. Os escores nos primeiros quatro subtestes foram convertidos em uma idade mental média e o escore no último subteste foi convertido em idade de desenvolvimento para a atenção. Não foi calculado um escore de QI uma vez que Snijders-Oomen considerou que o perfil dos quatro subtestes cognitivos forneceu uma visão muito mais clara do que um escore geral de QI.

Apesar das inovações propostas no SON-'43, a normatização da primeira versão do teste SON apresentou limitações: a amostra total de normatização consistiu de apenas 165 crianças surdas. Snijders-Oomen escreveu a este respeito: "Para todos os subtestes foram obtidas normas mais ou menos confiáveis na medida do possível, considerando o tamanho reduzido de amostra" (Snijders-Oomen, 1943, p. 57).

Em 1958, a primeira revisão do teste SON foi publicada: o SON-'58. O número de subtestes foi estendido de 5 a 8 e houve também um aumento do número de itens por subteste. Da mesma forma como no SON-'43, não havia tempo limites rigorosos nos subtestes, uma vez que a experiência mostrou que estes interferem no exame das crianças surdas.

O SON-'58 foi composto dos seguintes oito subtestes: dois testes pertencentes ao aspecto de inteligência *Percepção de Formas*: Mosaicos e Desenhar (semelhante aos subtestes Mosaicos e Padrões do SON-R 2½-7), dois subtestes pertencentes ao aspecto de inteligência *Compreensão da Coerência Concreta*: Combinar e Complementar (semelhante aos subtestes Puzzles e Situações do SON-R 2½-7), dois testes pertencentes ao aspecto de inteligência *Capacidade de Abstração*: Analogias e Classificar (semelhantes aos subtestes Analogias e Categorias do SON-R 2½-7) e dois testes pertencentes ao aspecto de inteligência *Concentração de Atenção e a Memória de Curto Prazo*: Memória de Imagens e Blocos de Knox (não são incluídos subtestes comparáveis

no SON-R 2½-7). A representação de dois subtestes dentro de cada aspecto de inteligência tornou possível distinguir duas formas paralelas de quatro subtestes.

As instruções de aplicação do SON-'58 ofereciam encorajamento e aprovação em caso de uma resposta correta e, em caso de uma resposta errada, forneciam ajuda para a criança, de tal forma que ela ainda poderia ter a impressão que ela mesma tivesse cooperado na solução. A lista de observação que foi incluído na folha de registro do SON-'43 não foi retida na folha de registro do SON-R '58.

O SON-'58 foi normatizado tanto para crianças ouvintes quanto para crianças surdas com idades entre de 3 a 17 anos (Snijders & Snijders-Oomen, 1958). A amostra de crianças ouvintes foi composta de 1.355 crianças e a amostra de crianças surdas foi constituída por todos os 1.054 alunos dos institutos holandeses para surdos.

Na segunda revisão dos testes SON, duas baterias separadas foram desenvolvidas: (a) o SON 2½-7 (Snijders & Snijders-Oomen, 1975) para crianças entre 2½ e 7 anos de idade, também conhecido como SON Pré-escolar e (b) o SSON (Starren, 1975) para pessoas entre 7 a 17 anos de idade. A principal razão para a divisão em dois testes separados foi que a transição para um diferente tipo de tarefa parecia mais apropriado em todos os subtestes do SON-'58 a partir de 7 anos de idade (Snijders & Snijders-Oomen, 1975).

Na revisão do SON-'58 foi necessário rever a distinção teórica em quatro aspectos de inteligência que serviu de base para a construção do SON-'43 e do SON-'58. Na pesquisa de normatização do SON-'58, os subtestes mostravam correlações entre eles muito fortes, e usando análise fatorial, não foram encontradas evidências que apoiavam a distinção nos quatro aspectos de inteligência Percepção de Formas, Compreensão da Coerência Concreta, Capacidade de Abstração e Concentração de Atenção e Memória de Curto Prazo (Winkel, 1999). Por esta razão foi decidido não mais usar os quatros aspectos

de inteligência na revisão do SON-'58. Também foi decidido não mais tentar construir duas formas paralelas do teste.

O SON 2½-7 foi composto dos seguintes cinco subtestes: Classificar, Mosaicos, Combinar, Desenhar e Memória. No SON 2½-7 não foram calculados escores normatizados no nível de subtestes, apenas um escore geral de QI. Assim, não era possível calcular um escore de QI quando um dos subtestes não era administrado.

O SON 2½-7 parecia-se muito com o SON-'58: a estrutura geral e os pontos de partida foram semelhantes aos do SON-'58. Entretanto, os seguintes princípios podem ser distinguidos na construção do SON 2½-7: a aspiração de criar uma situação de pesquisa mais natural possível que não seria ameaçadora para a criança, utilização de tarefas atraentes, onde o elemento de desempenho não estava em primeiro plano, sem o uso de tempo limites rigorosos, e onde o aplicador tinha que estimular a criança e ajuda-la em caso de dificuldades. A possibilidade de observar a criança teve um papel importante. O teste deveria oferecer uma oportunidade de obter uma impressão do comportamento da criança em situações variadas.

O SSON consistiu dos seguintes sete subtestes: Montar Figuras, Histórias, Analogias, Refletir Figuras, Situações, Classificar Imagens e Memória para fotos. No SSON não foram calculados escores no nível de subtestes. Como no SON 2½-7, foram calculados apenas escores de QI geral. O conteúdo do teste foi muito parecido com o SON-'58, com a grande diferença que todos os subtestes estavam em formato de múltipla escolha. Comparado com o SSON, o SON-'58 continha tarefas que estimulavam mais e que ofereciam maior oportunidade para observar o comportamento das pessoas que respondiam o teste. Com base nessas considerações, foi decidido manter tanto a publicação do SON-'58 como do SSON.

Na terceira revisão dos testes SON, tanto o SON-'58 como o SSON foram substituídos pelo SON-R 5½-17 (Snijders, Tellegen & Laros, 1989), e o SON 2½-7 foi substituído pelo SON-R 2½-7 (Tellegen, Winkel, Wijnberg-Williams & Laros, 1998). O SON-R 5½-17 foi composto de sete subtestes: dois testes espaciais (Mosaicos e Padrões), dois testes de raciocínio concreto (Situações e Estórias), dois testes de raciocínio abstrato (Categorias e Analogias) e um teste perceptivo (Figuras Ocultas). Já o SON-R 2½-7, foi composto por seis subtestes: três subtestes de execução (Padrões, Mosaicos e Quebra-cabeças) e três subtestes de raciocínio (Situações, Categorias, e Analogias).

A distinção teórica em quatro tipos de tarefas no SON-R 5½-17- testes espaciais, testes de raciocínio concreto, testes de raciocínio abstrato e um teste de percepção - foi empiricamente apoiada por análise fatorial apenas no grupo mais jovem com idade de 6 anos e 6 meses. A divisão em quatro dimensões tornou-se menos aparente com o aumento da idade dos sujeitos. À medida que a idade dos sujeitos aumentou a influência do fator geral ficou cada vez mais forte. Assim, com base nestes resultados, foi decidido calcular apenas um escore de QI geral no SON-R 5½-17 e não escores para escalas específicas. Testes de memória não foram incluídos no SON-R 5½-17 levando em consideração de que um estudo da memória pode ser melhor realizado com uma bateria específica para medir memória, em vez com um único subteste.

O SON-R 5½-17 foi o primeiro teste SON onde foi utilizado um procedimento de testagem adaptativa. O uso deste procedimento diminuiu consideravelmente o número de itens aplicados que são fáceis demais ou difíceis demais para os respondentes. Em todos os sete subtestes do SON-R 5½-17, era oferecido *feedback*, que consistia em indicar se a solução para o item estava certa ou errada. Com o programa de computador, que desde a publicação do SON-R 5½-17 faz parte integral dos testes SON, os escores normatizados podem ser calculados e exibidos em uma forma simples. Com o programa de computador,

também é possível determinar a idade de referência dos respondentes, o cálculo do escore de QI para cada combinação de subtestes e o cálculo da correção do escore de QI pelo efeito Flynn. Desde 2013 esta correção é baseada em dois em vez de três pontos de QI por década.

Depois da publicação do SON-R 5½-17, a revisão do SON Pré-Escolar foi iniciada, o que resultou, em 1998, na publicação do SON-R 2½-7. O manual e relatório científico do SON-R 2½-7 (Tellegen, Winkel, Wijnberg-Williams & Laros, 1998) relata a grande quantidade de pesquisa, tanto nacional como internacionalmente, que foi realizada com o teste.

O SON-R 2½-7 é formado por seis subtestes diferentes: três subtestes de execução (Padrões, Mosaicos e Quebra-cabeças) e três subtestes de raciocínio (Situações, Categorias e Analogias). Com base na análise fatorial, foram identificados dois fatores: um fator de execução e um fator de raciocínio. Assim, levando em consideração estes resultados, foram calculados um QI geral (SON-QI) e também escores normatizados para as duas escalas: a Escala de Execução (SON-EE) e Escala de Raciocínio (SON-ER). O SON-R 2½-7 foi o primeiro teste SON no qual foram calculados escores normatizados para duas escalas e um escore de QI geral.

Para a avaliação da validade da distinção entre a Escala de Execução e a Escala de Raciocínio, foi pesquisado se existiam diferenças consistentes no grau em que essas escalas eram associadas com outros testes. Em comparação com a escala de raciocínio (SON-ER), foi encontrada para a escala de execução (SON-EE) uma correlação significativamente mais elevada com a escala de execução dos testes de Wechsler, com as partes de execução do LDT (Teste Diagnóstico de Leiden - Schroots & Alphen de Veer, 1976) e com o DTVP-2 (Teste Desenvolvidor de Percepção Visual - Hammill, Pearson & Voress, 1993). Em comparação com a SON-EE, a SON-ER mostrou uma

relação mais forte com a escala verbal dos testes Wechsler, com os subtestes verbais dos BAS (Escala Britânica de Habilidade - Elliott, Murray & Pearson, 1982), com os subtestes verbais e testes de memória do LDT e com o Teste Reynell para compreensão da linguagem (Escala Reynell de Desenvolvimento da Linguagem - Reynell, 1985).

Os resultados da coerência dos escores nas escalas do SON-R 2½-7 com outros testes apoia a distinção que é feita com base na análise da estrutura fatorial do teste. Os resultados indicam que, no SON-R 2½-7, dois aspectos da inteligência geral são representados: em primeiro lugar tarefas de execução relacionados à habilidade espacial e habilidade viso-motor, e tarefas que apelam mais para a capacidade de raciocínio abstrato e concreto. A capacidade de raciocínio mostrou uma relação mais forte com inteligência verbal em habilidades de linguagem. O fato que o SON-R 2½-7 medir dois aspectos de inteligência significa que o teste é mais versátil do que testes de inteligência não-verbal que são limitados a tarefas estritamente de execução.

Em 2007, uma versão abreviada do SON-R 2½-7 foi publicada, o SON-R 2½-7[a], onde a letra [a] significa abreviado. Esta versão é formada por quatro subtestes: Mosaicos, Categorias, Situações e Padrões. O SON-R 2½-7[a] é destinado para uso fora da Europa e também para facilitar a utilização do teste em países não-ocidentais. Para o subteste Categorias do SON-R 2½-7[a], foram realizados vários estudos para diminuir o viés cultural dos itens (Bijl, 2004; Boven & Rusthoven, 2005; Laros & Tellegen, 2004). Em 2015, a versão brasileira do SON-R 2½-7[a] foi publicada, voltada para a avaliação de crianças entre dois anos e meio e sete anos de idade (Laros, Tellegen, Jesus & Karino, 2015; Laros, Jesus & Karino, 2013).

O SON-R 2-8 e SON-R 6-40 fazem parte da quarta revisão do teste SON. O SON-R 6-40 foi publicado em 2011 (Tellegen & Laros, 2011) e é o primeiro teste SON com normas para adultos até quarenta anos. O teste consiste dos seguintes subtestes:

Analogias, Mosaicos, Categorias e Padrões. Em comparação com o SON-R 5½-17, o número de subtestes foi reduzido de seis a quatro, enquanto o número de itens de cada subteste foi aumentado. Deste modo, a duração de administração do teste foi diminuída para cerca de uma hora, enquanto a fidedignidade e a validade do teste foram mantidas. Especialmente para o subteste Categorias, foi realizado um grande número de pesquisas nacionais e internacionais para tornar o subteste menos dependente de culturas específicas (Tellegen & Laros, 2011; Stap, 2009).

A construção do SON-R 6-40 foi feita com base nos seguintes pontos de partida:

- (1) o interesse do cliente deve ter precedência;
- (2) a mensuração das capacidades cognitivas do cliente deve ocorrer de maneira mais exata possível;
- (3) e o resultado da mensuração deve ser visto em perspectiva.

O ponto de partida mais importante para os construtores do teste - assim como foi para Nan Snijders-Oomen - é que o teste é destinado a servir os interesses do cliente. A pessoa a ser testada deve ser capaz de confiar que a investigação diagnóstica é focada em seu bem-estar, como isso também é um ponto de partida óbvio numa investigação médica. A administração do teste pode ser vista como um processo de mensuração em que um resultado numérico é obtido de um modo padronizado. Se uma mensuração é realizada, é óbvio, ou pelo menos deveria ser óbvio, que se tenta realizar a mensuração mais precisa possível. Ou seja, isso também é do interesse do cliente. Esforçar-se para obter a precisão da mensuração não significa que o resultado deve ser considerado como um dado absoluto. É necessário perceber que o resultado do teste tem limitações em relação ao construto complexo de inteligência. A relatividade do resultado do teste é, portanto, também do interesse do cliente. Assim, o processo de validação do teste não teve somente

como objetivo mostrar o quão bom o teste é, mas também mostrar quais são as limitações do teste.

O SON-R 2-8, a última versão do teste SON para crianças pequenas, foi publicado em 2017 (Tellegen & Laros, 2017). Os resultados da investigação sobre a relação dos escores das duas escalas, SON-EE e SON-ER, com outros testes, confirmou a distinção que é feita entre os subtestes pertencentes à Escala de Execução e os subtestes associados a Escala de Raciocínio. Em comparação com a Escala de Raciocínio, a Escala de Execução correlaciona mais forte com: o escore total com o WNV-NL (Wechsler Non Verbal - Wechsler & Naglieri, 2008), o QI de Execução do WPPSI-III-NL (Wechsler, 2009) e o fator processamento visual do IDS-2-NL (Grob & Haggmann von Arx, 2017) enquanto que a Escala de Raciocínio apresenta uma correlação significativamente maior com o QI verbal do WPPSI-III-NL, com o Índice Geral de Linguagem do WPPSI-III, com o fator pensamento abstrato do IDS-2-NL e o fator de memória de longo prazo do IDS-2-NL (Tellegen & Laros, 2017).

O SON-R 5½-17, SON-R 2½-7 e SON-R 6-40 foram avaliados pela Comissão Nacional de Avaliação de Testes dos Países Baixos (COTAN), instituto responsável pela avaliação de testes psicológicos, e todos os aspectos avaliados (base teórica para construção de teste, manual do teste, material do teste, normas, fidedignidade e validade) têm recebido a melhor avaliação. As categorias de avaliação utilizadas pela COTAN são: insuficiente, suficiente e bom (Tellegen & Laros, 2017). O SON-R 2-8 está no processo de avaliação pela COTAN.

2.1 O teste não-verbal de inteligência SON-R 6-40

A quarta versão dos testes SON para crianças e para adultos, o SON-R 6-40 (Tellegen & Laros, 2011), foi publicada em 2011 e é o objeto sobre o qual versa o presente trabalho. O SON-R 6-40 é uma medida padronizada de inteligência voltado para avaliação

de pessoas com idade entre 6 e 40 anos. Nesta versão do teste SON, os escores em quatro subtestes são combinados para formar um escore de inteligência, que representa o desempenho dessa pessoa em relação ao seu grupo de idade.

O SON-R 6-40 foi desenvolvido para substituir a versão anterior do teste, o SON-R 5½-17. As principais razões para rever o SON-R 5½-17 foram as seguintes: a necessidade de atualizar as normas; modernizar os materiais do teste a fim de deixá-lo culturalmente menos específico; melhorar a capacidade de diferenciação, especialmente para pessoas com habilidades limitadas; adequar o teste para adultos e encurtar o tempo de administração. As razões citadas que fundamentaram a revisão do teste SON-R 5½-17 serão explicadas com mais detalhes a seguir.

Atualização das normas

O SON-R 5½-17 foi normatizado em 1985 e publicado em 1988. Depois de um período de mais de 25 anos, uma revisão de um teste de inteligência era, sem dúvida, desejável, pois normas de teste tendem a se tornar obsoletas ao longo do tempo. Estudos têm demonstrado (Lynn & Hampson, 1986; Flynn, 1987, 2009) que ao longo de um período de dez anos, o desempenho em testes de inteligência aumenta em cerca de dois ou três pontos de QI.

Materiais do teste

O SON-R 6-40 é composto pelos seguintes quatro subtestes do SON-R 5½-17: Analogias, Mosaicos, Categorias e Padrões. Como o teste também é utilizado fora da Holanda, as imagens tiveram de ser adaptadas para garantir que as tarefas pudessem ser facilmente compreensíveis em culturas diferentes. Alguns dos desenhos de Categorias não estavam muito claros ou pareciam antigos. A Figura 2.3 mostra exemplos de imagens que foram substituídas por serem consideradas inadequadas por apresentarem um forte viés cultural.

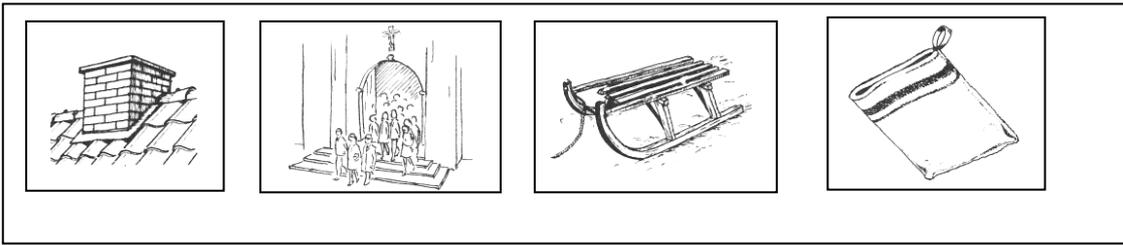


Figura 2.3 Figuras substituídas no subteste Categorias com viés cultural

Capacidade de diferenciação

A seleção dos quatro subtestes do SON-R 6-40 entre os sete subtestes do SON-R 5½-17 foi baseada em argumentos empíricos. Assim, nenhuma diminuição da validade do teste comparado com a validade do SON-R 5½-17 era esperada. Para compensar uma possível perda de fidedignidade do escore geral, o número total de itens que compõem os quatro subtestes foi aumentado de 98 para 124. Um esforço foi feito para melhorar a dificuldade progressiva dentro de cada um dos subtestes e para garantir que uma quantidade suficiente de itens mais fáceis seria incluída.

Inclusão de adultos

Os autores do teste afirmam que nenhuma grande mudança foi necessária para que o teste se tornasse adequado para adultos, pois o processo adaptativo limita o número de itens testados que são demasiadamente simples para este grupo. O estudo de normatização para o SON-R 5½-17 mostrou que os escores brutos ao redor de 18 anos quase não aumentam mais. No SON-R 5½-17, apenas no subteste Mosaicos um efeito de teto esteve presente. Assim, os autores decidiram aumentar o nível de dificuldade deste subteste. Entretanto, o estudo de normatização e validação do SON-R 6-40 na Holanda e Alemanha mostrou que um efeito de teto ocorre com os adultos no subteste Padrões.

Diminuição do tempo de administração do teste

Muitos benefícios práticos resultam da abreviação do teste e do encurtamento do tempo para sua administração. Considerando os dados da pesquisa de normatização do SON-R 5½-17, foi demonstrado que uma combinação de quatro subtestes mostrou uma validade igual que o conjunto completo de seis subtestes. Nem a versão abreviada do SON-R 2½-7 parecia resultar uma perda de validade em comparação com a versão maior com seis subtestes (Tellegen & Laros, 2011). Uma redução no número de subtestes, no entanto, limita as oportunidades de diferenciação em relação a diversos aspectos de desempenho do teste. Isso é compensado pelo fato de que é muito mais simples conseguir efetuar um teste curto quando a aplicação é realizada junto com outros testes. Estudos comparativos mostram que cada teste oferece informações sobre o examinando a partir de uma perspectiva específica. Se o resultado de uma aplicação de um teste não for claro, é recomendada a administração de um segundo teste para a tomada de decisões importantes. Como a aplicação do SON-R 6-40 agora dura em torno de 60 minutos em vez de 90 minutos, fica mais fácil de realizá-lo na prática.

2.2 Construção do teste SON-R 6-40

Baseados nas experiências com o estudo de normatização do teste SON-R 5½-17, os autores do teste realizaram uma série de estudos pilotos, que começaram em 2003 e finalizaram no início de 2009. Inicialmente, foram realizados estudos na China (Judistira, 1996; Zhang, Yanqing & Xiaoxun, 1997), no Peru (Le Clercq & Holvast, 1996) e no Brasil (Laros & Tellegen, 2001, 2004). Na China, todos os sete subtestes do SON-R 5½-17 foram respondidos por 302 crianças de cerca de dez anos de idade; no Peru, 160 crianças da mesma idade responderam o teste; no Brasil, 83 crianças com idades entre 7 e 14 anos responderam três subtestes (Categorias, Situações e Histórias). Os resultados indicaram que os escores dos subtestes que utilizavam imagens com significado

(Categorias, Situações e Histórias) foram significativamente menores do que dos outros quatro subtestes. Na China, uma diferença de nove pontos de QI foi observada entre estes dois tipos de subtestes. Os examinadores chineses avaliaram as imagens que, na sua opinião, eram menos adequados à sua cultura. No Peru e no Brasil, as crianças que deram uma resposta incorreta foram então questionadas se elas tinham entendido a imagem. Em Categorias, 14 dos 33 itens tinham problemas com uma ou mais fotos. No subteste Situações isso aconteceu com 6 dos 33 itens, enquanto que quase não houve dificuldades relatadas no subtestes Histórias.

Os resultados da pesquisa com o teste SON-R 5½-17 em outras culturas mostrou que era necessária uma revisão profunda dos subtestes Categorias e Situações para que eles se tornassem menos culturalmente tendenciosos e, portanto, adequados para inclusão na nova versão do teste. Assim, os autores resolveram criar diferentes versões dos subtestes para verificar o grau de entendimentos dos desenhos e figuras. Os estudos envolvendo os subtestes Analogias, Mosaicos e Padrões foram geralmente efetuados simultaneamente e são descritos aqui primeiro. A construção da pesquisa para o subteste Categorias foi realizada por meio de uma grande escala de testes realizados em muitos países diferentes e será discutida separadamente. É importante destacar que a autora desta tese não necessitou fazer estudos de adaptação ou estudos pilotos no Brasil devido ao diligente cuidado que os autores tiveram na construção do teste SON-R 6-40, contemplando amostras brasileiras em todas as fases do processo de construção do teste.

A construção dos subtestes Analogias, Mosaicos e Padrões

Inicialmente, vários cadernos dos subtestes Analogias, Mosaicos e Padrões foram construídos. Para Analogias, este caderno de teste consistia em dezoito itens simples, com sete deles sendo novos desenhos. No caso de Mosaicos, um caderno simples e um difícil

foram criados, contendo cada um, doze itens. Um caderno simples e um difícil, igualmente, foram criados para Padrões, contendo doze e catorze itens, respectivamente.

A pesquisa com estes cadernos de teste foi realizada em 2004 em mais de uma centena de alunos de escolas primárias na Holanda e com outras 21 crianças testadas também no Quênia. As crianças do Quênia, que tinham idade entre 6-12 anos, vieram de áreas urbanas e rurais. Os subtestes foram administrados individualmente (Boven & Rusthoven, 2005).

Com base nos resultados desse estudo inicial, novos cadernos de teste foram então criados para Analogias, Mosaicos e Padrões. Essa versão do teste recebeu o nome provisório de SON-I. Nesta versão, o subteste Analogias consistia de três séries, cada uma composta por doze itens, enquanto Mosaicos e Padrões consistia, cada um, de duas séries com onze itens. Alguns dos itens nesses novos cadernos de teste foram do SON-R 5½-17, mas os outros eram itens que tinham sido desenvolvidos durante o estudo anterior. Na compilação dos cadernos de teste, foi dada atenção especial à sequência de itens nas séries.

Os cadernos de teste do SON-I foram então utilizados em 2006 para efetuar testes individuais com um total de 337 alunos. Isso envolveu 90 alunos de uma escola primária e de uma escola secundária na Holanda (Hof, Eekhout & Mengerink, 2006), 65 crianças com ascendência imigrante de uma classe preparatória internacional e uma classe preparatória para crianças altamente qualificadas que falavam uma língua diferente (Letitre, van Calkar & Wemmenhove, 2007), 20 alunos de uma escola de chineses que viviam na Holanda (Wu, 2007), 72 crianças em idade escolar que viviam em Bangkok e 90 crianças do Brasil. A idade variou entre 6 e 22 anos.

Uma regra menos rigorosa de descontinuação foi aplicada na estrutura da presente pesquisa, com uma série não sendo parada depois de duas respostas incorretas, mas depois de três. Isto forneceu mais informações sobre a dificuldade de cada item.

Para obter uma melhor compreensão de como os itens mais simples funcionam no caso de jovens estudantes com baixas habilidades, um estudo separado foi realizado com 42 alunos de uma escola primária para crianças com necessidades especiais e/ou graves dificuldades de aprendizagem. A média de idade foi de oito anos. Outros testes de inteligência deram a essas crianças uma média de QI de 63 (intervalo de 50-85). Os cadernos de teste consistiam em duas ou três séries, cada uma contendo seis itens. No caso de Padrões, dois novos itens, relativamente fáceis, foram criados para este estudo (Boven, 2007).

Os resultados do SON-I foram então avaliados em 2008 (Tellegen & Laros, 2008). Uma análise de TRI com o programa *BILOG-MG* (Zimowski, Muraki, Mislevy & Bock, 2003) foi usada aqui para estimar os parâmetros de dificuldade dos itens com base nos dados de 337 pessoas que tinham sido testadas com o SON-I. Estes resultados e os resultados dos itens simples usados em 42 crianças com baixas habilidades levaram à conclusão de que um número pequeno de alterações era desejável para melhorar a composição do teste. Assim, três itens em Analogias foram então modificados para garantir uma transição mais tranquila entre os níveis de dificuldade dos itens. Em seguida, um caderno de teste contendo os itens alterados e os itens inalterados foi testado em 57 crianças de uma escola primária.

No caso de Mosaicos, o objetivo foi deixar o subtteste ser diferenciado mais claramente tanto no nível simples quanto no difícil. Para o nível simples, um caderno de teste com dez itens foi criado com dois itens novos e um item revisado (Aertsen, 2008). Esta versão foi utilizada com 41 crianças de uma escola primária (idade de 5 a 8 anos) e

15 alunos de uma escola para crianças com necessidades especiais (idade de 7-10 anos). Para fazer o teste mais difícil, novos itens complexos foram testados pela primeira vez em um estudo piloto. Um caderno de teste com dez itens foi criado, incluindo cinco novos itens difíceis. Esta versão foi utilizada com cinquenta pessoas em estabelecimentos de ensino profissional e de nível universitário (Arragon & Koster, 2008). A idade variou entre 18 e 26 anos.

Tanto um caderno de teste simples quanto um difícil também foram criados para Padrões. Na versão simples, o caderno consistia de doze itens, incluindo um novo item simples e dois itens que anteriormente haviam sido criados para aprendizes lentos. Esta versão foi testada em 57 alunos da escola primária regular ou da escola primária para crianças com necessidades especiais com idades compreendidas entre os 5 e os 10 anos (Hennemann, 2009).

No caso do caderno de teste difícil, nenhuma alteração foi feita nos próprios itens. Em vez disso, foi averiguado se o nível de dificuldade seria aumentado caso o tempo máximo para os itens difíceis fosse limitado a três minutos, ao invés dos quatro minutos permitidos com o SON-R 5½-17. A versão do teste consistiu nos dez itens mais difíceis do SON-I e foi administrada a 43 alunos no ensino superior (Kuipers, 2008). Para alguns dos indivíduos o tempo máximo permitido foi de quatro minutos, mas foi registrada qual pontuação eles teriam com um tempo máximo de três minutos. Para outros foi apenas permitido o tempo máximo de três minutos. Foi observado que apenas poucos indivíduos encontraram a solução correta após três minutos, ainda dentro do limite de tempo de quatro minutos. Em praticamente todos os casos os indivíduos encontraram a solução correta dentro de três minutos ou quatro minutos ainda não seriam suficientes para fazê-lo.

A construção de Categorias

Com base nos resultados do estudo piloto, todos os itens para as Categorias foram reelaborados em 2003, levando em consideração os resultados do estudo transcultural com o SON-R 5½-17 (Bijl, 2004). A Figura 2.4 mostra alguns exemplos de imagens que foram alteradas após o estudo de validação.

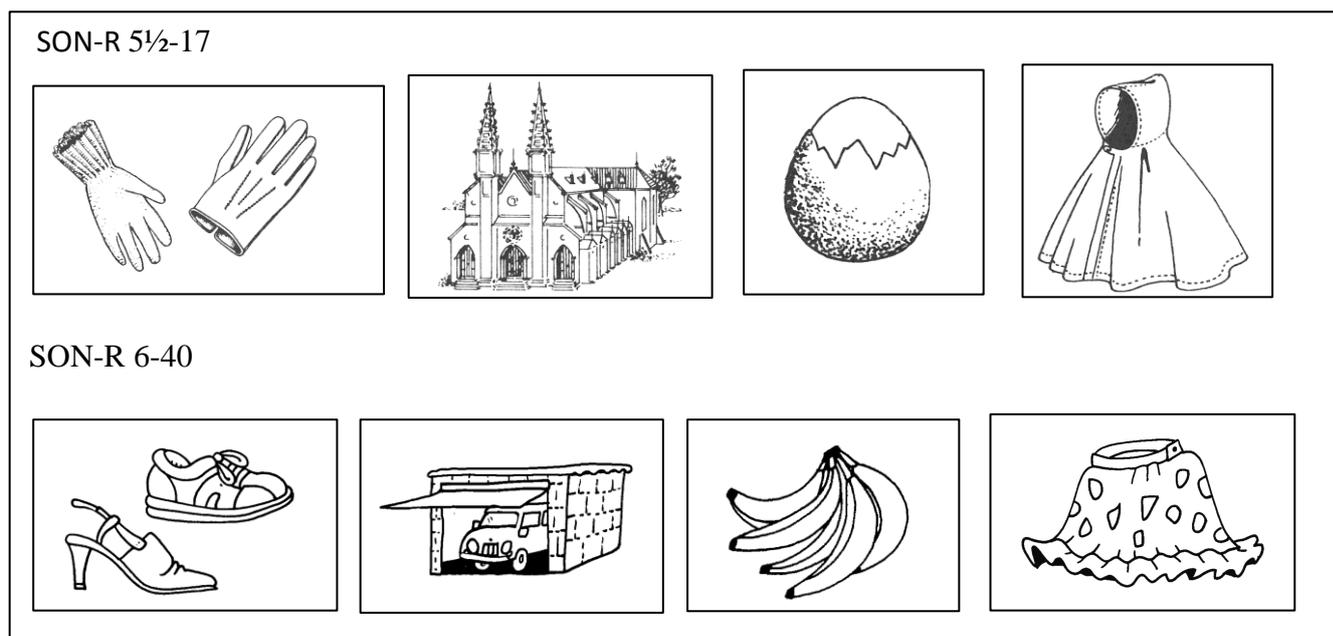


Figura 2.4 Figuras do SON-R 5½-17 e as alternativas do SON-R 6-40.

O motivo da alteração das imagens foi que elas eram muito culturalmente específicas ou não eram claras o suficiente. Para aumentar a capacidade de discriminação do subteste, tanto no nível simples quanto no difícil, quatro itens modificados do SON-R 2½-7 e nove itens difíceis recém criados foram incluídos em um novo caderno. Assim, um conjunto de 39 itens foi utilizado para realizar um estudo em grupo. Os itens foram mostrados em três cadernos de teste com diferentes níveis de dificuldade, que se sobrepunham até certa medida. Cada versão apresentava uma sequência de itens diferente. Isso se deu para evitar que os resultados fossem indevidamente influenciados pela posição do item no caderno de teste. O primeiro estudo foi realizado com 256 alunos de uma escola primária e de uma secundária na Holanda (Bijl, 2004).

Os resultados desse primeiro estudo realizado na Holanda indicaram que era necessário realizar alterações nas imagens de alguns itens. Foi também notado que o nível de dificuldade de alguns itens não era o que se esperava.

Assim, um estudo posterior foi realizado com os 39 itens alterados organizados em três cadernos de teste em uma nova sequência, com diferentes níveis de dificuldade que se sobrepunham em alguns casos. O subtteste Categorias foi administrado em 2004 desta maneira em 1.963 indivíduos nos seguintes países: Holanda ($N=446$), Quênia ($N=331$), Marrocos ($N=230$), Indonésia ($N=538$) e Brasil ($N=418$). O *feedback* dos alunos e examinadores indicou que algumas imagens não eram facilmente compreensíveis em certas culturas. As crianças no Quênia, por exemplo, não consideravam o tambor como um brinquedo, e não conheciam uma âncora (Boven & Rusthoven, 2005). Uma análise do Funcionamento Diferencial do Item (DIF) com o programa *BILOG-MG* (Zimowski et al., 2003) indicou que houve diferenças entre países no que diz respeito à dificuldade progressiva de Categorias.

Considerando esses resultados, muitas outras mudanças foram feitas nas imagens. A nova edição, assim como a anterior, consistiu de 39 itens distribuídos em três níveis de dificuldade. Em 2005 esta versão foi apresentada para 3.474 pessoas nos seguintes países: Holanda ($N=1.058$), Eslováquia ($N=338$), Suriname e Curaçao ($N=686$), Brasil ($N=478$), Tailândia ($N=536$) e Irã (378). Os resultados deste estudo foram também analisados com o *BILOG-MG*. Em geral, os níveis de dificuldade coincidiram bem entre os países. As alterações efetuadas em Categorias claramente tiveram o efeito desejado de reduzir o viés cultural desse subtteste.

Quando um estudo em grupo é realizado, a situação é substancialmente diferente do procedimento padrão para administração do SON-R. Não há interação com o examinador nesta situação. O subtteste não envolve um processo adaptativo e nenhum

feedback é dado. Em 2006, uma versão de Categorias foi criada na qual os itens foram testados individualmente. Nesta versão do teste, nove itens que resultaram em grandes diferenças na Holanda por um lado e no Brasil e na Tailândia por outro, foram omitidos. Esta versão do teste consistiu de três series paralelas de dez itens, que juntamente com Analogias, Mosaicos e Padrões constituíram o SON-I. Estas quatro partes do teste foram apresentadas para 337 pessoas, como descrito acima em relação à construção dos outros subtestes.

Uma versão de Categorias, que tinha sido modificada para torná-la mais simples, também foi aplicada em um grupo de 42 crianças com problemas de aprendizagem. Durante a avaliação do SON-I em 2008 foi decidido que mais mudanças para os desenhos de cinco itens em Categorias eram necessárias e estes foram testados em um estudo utilizando itens fáceis (no total de doze itens) e em um estudo com itens de dificuldade média (também doze itens). O caderno de teste fácil foi testado em 47 crianças de uma escola primária na Holanda e o caderno mais difícil em 153 alunos holandeses e 94 alunos brasileiros, também de escola primária (Stap, 2009).

Os resultados das modificações permitiram aos autores guiar a construção dos subtestes do teste SON-R 6-40 e melhorar o nível de dificuldade progressiva dos itens. A Figura 2.5 apresenta uma visão geral do grau de dificuldade de todos os itens do teste SON-R 5½-17 e a Figura 2.6 apresenta as informações deste mesmo parâmetro para o teste SON-R 6-40 com os dados da pesquisa de normatização realizada na Holanda/Alemanha.

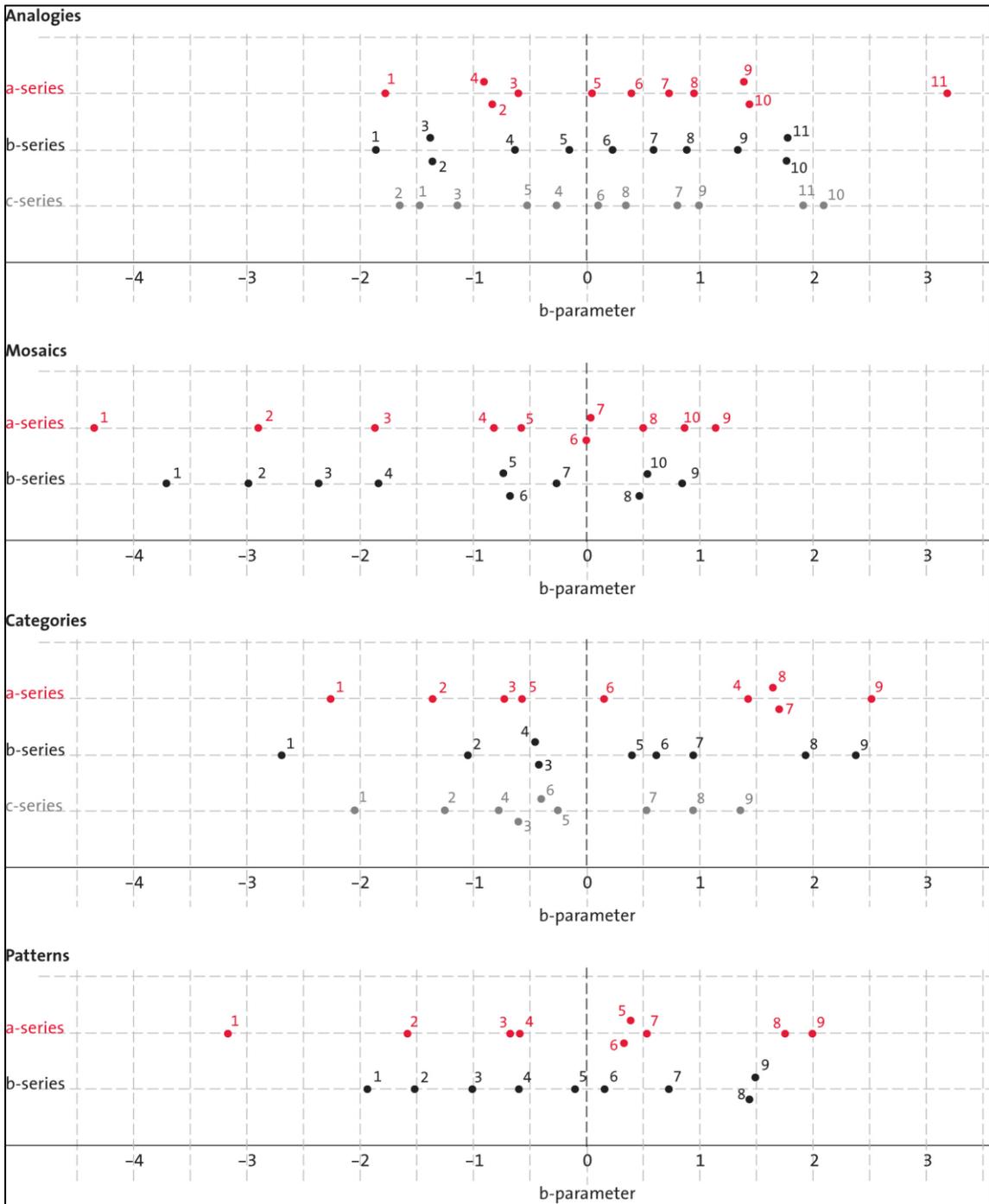


Figura 2.5 A dificuldade dos itens do teste SON-R 5½-17 ($N = 1.350$)

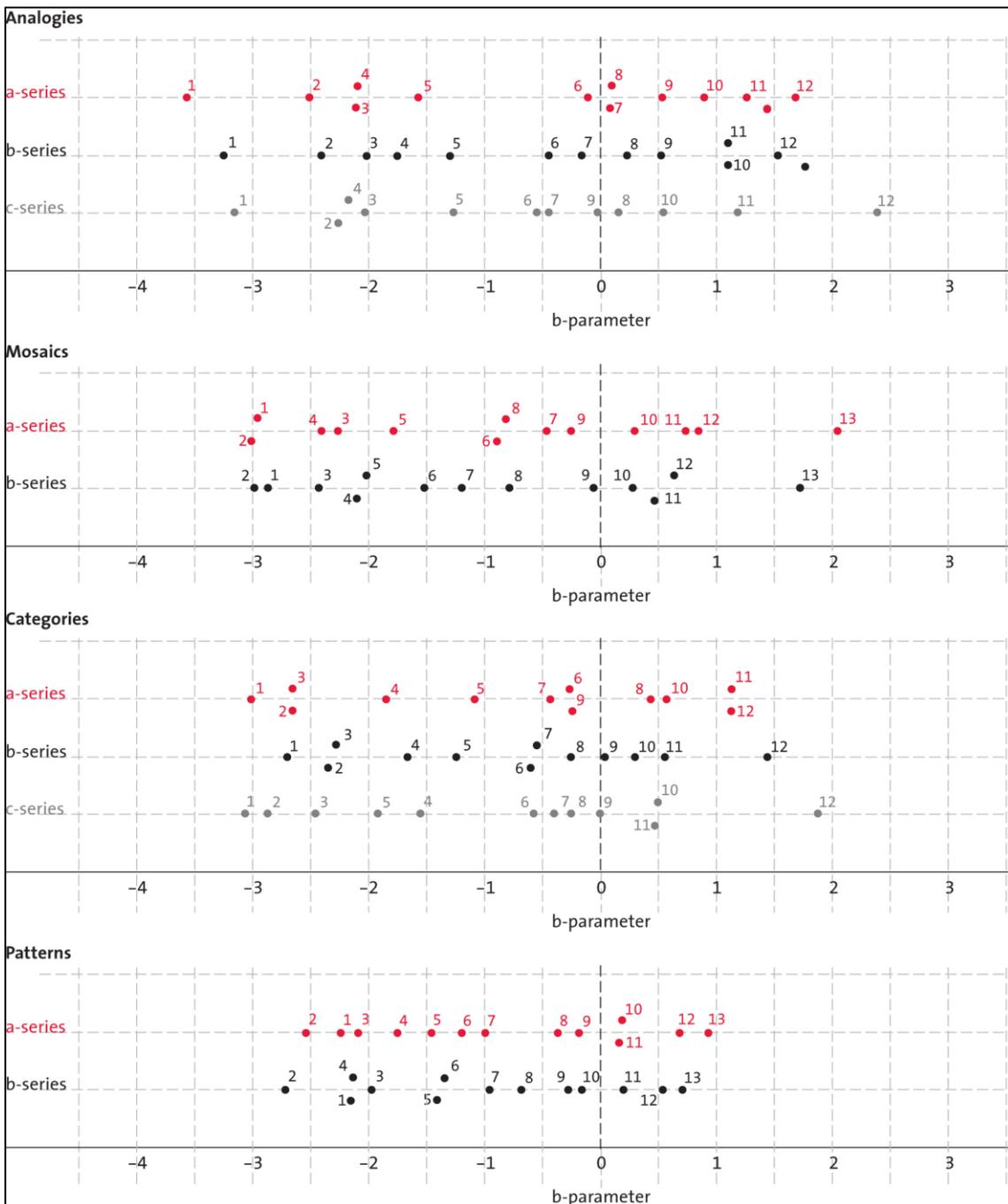


Figura 2.6 Dificuldade dos itens do teste SON-R 6-40 ($N = 1.933$).

A análise das Figuras 2.5 e 2.6 permite inferir que os autores do teste SON-R 6-40 conseguiram melhorar a dificuldade progressiva dos itens de cada subteste, um dos motivos observados quando decidiram fazer a atualização das normas do SON-R 5½-17. Além disso, itens mais fáceis foram inseridos em cada subteste visando melhorar a capacidade de diferenciação.

2.2 Características gerais do SON-R 6-40

O SON-R 6-40 (Tellegen & Laros, 2011) é um teste não-verbal de inteligência de aplicação individual. É composto por quatro subtestes, a saber: testes de raciocínio abstrato (Analogias e Categorias) e testes de desempenho com uma orientação espacial (Mosaicos e Padrões).

Embora instruções verbais possam ser dadas, é possível administrar o teste sem empregar a linguagem falada. Contudo, nada que é dito pelo examinador pode oferecer informações adicionais em comparação com a instrução não-verbal. A maneira pela qual o teste é apresentado é adaptada à habilidade de comunicação do indivíduo, permitindo que o teste corra o mais “naturalmente” possível.

O fato de o SON-R 6-40 ter a possibilidade de ser administrado sem o uso de linguagem escrita ou falada, torna o teste especialmente adequado para pessoas com deficiência em termos de linguagem ou comunicação assim como para imigrantes com pouco ou nenhum conhecimento da língua falada pelo examinador.

Uma característica distinta do SON-R 6-40 é que o *feedback* é dado durante o teste. Após um item ter sido completado pelo indivíduo, o examinador indica se a resposta está correta ou incorreta. Contudo, questões incorretas são corrigidas apenas nos exemplos. O *feedback* torna a situação de teste mais natural e permite que os indivíduos adaptem as suas estratégias de solução de problema (Assink et al, 1992; Tellegen & Laros, 1993).

Os itens dos subtestes são organizados em duas ou três séries paralelas em ordem crescente de dificuldade. Uma vez que o indivíduo tenha cometido alguns erros, o examinador avança para itens mais simples, de acordo com as regras estabelecidas para descontinuar e pular itens. Este procedimento tem um efeito positivo na motivação dos examinandos. Itens que são extremamente difíceis ou extremamente fáceis para uma determinada pessoa são pulados, limitando assim a duração do teste para cerca de uma

hora. Este procedimento também assegura que os indivíduos não percam sua motivação ou concentração.

O SON-R 6-40 está em concordância com outros testes de inteligência como o WISC-III e o WISC-IV (Wechsler, 2005a; Zhu & Weiss, 2005; Lima & Laros, 2017), o NIO (Dijk & Tellegen, 2004), o GIT2 (Luteijn & Barelds, 2004), o WAIS-IV e o WAIS-V (Wechsler, 2005b, 2007), a BPR-5 e o SON-R 2½-7[a] (Laros, Almeida, Valentini & Lima, 2015) e a MPC, a EMMC e o TONI-3, forma A (Alves, 2016) na medida em que a avaliação da inteligência é baseada nos escores obtidos em tarefas altamente diversificadas. Isto, é claro, sujeita-se à limitação de que o SON-R 6-40 não inclui itens de uma natureza especificamente verbal. Como itens verbais na maioria das vezes dependem significativamente de conhecimento e experiência, é esperado, em comparação a estes outros testes, que o SON-R 6-40 se concentre mais na medição da inteligência fluida e menos na medição de inteligência cristalizada (Cattell, 1971; Horn & Blankson, 2005; McGrew, 2005; Lima & Laros, 2017).

Existem duas diferenças importantes em relação aos subtestes não verbais incluídos em outros testes de inteligência. Em primeiro lugar, a parte não-verbal de outros testes é geralmente limitada aos subtestes de execução, enquanto o SON-R 6-40 também inclui tarefas de raciocínio. Em segundo lugar, apesar de que o material da parte de execução ser não-verbal, as instruções são geralmente dadas verbalmente (Tellegen & Laros, 2014).

Uma diferença importante em relação aos testes não-verbais de inteligência tais como Matrizes Progressivas Coloridas - MPC (Raven, 1962; Raven, Raven & Court, 2003) e o TONI-4 (Brown, Sherbenou & Johnsen, 2010) é que tais testes consistem em apenas um tipo de tarefa, deste modo dependendo grandemente da habilidade específica medida por aquele teste. Testes não-verbais de inteligência tais como o CTONI (Hammill, Pearson & Wiederholt, 1996), o UNIT (Bracken & McCallum, 1998) e o WNV (Wechsler

& Naglieri, 2008) se assemelham com o SON-R 6-40 pelo fato de ser construídos de vários subtestes. No entanto, uma diferença fundamental é que as instruções para esses testes são exclusivamente visuais e gesticuladas, enquanto as instruções que acompanham o SON-R 6-40 são adaptadas para a capacidade de comunicação do indivíduo, para tornar a situação de teste a mais natural possível. Outra diferença importante é que no SON-R 6-40 o examinando é informado se a solução foi correta ou incorreta. Os examinandos, assim, têm a oportunidade de adaptar a sua estratégia de resolução de problemas.

2.3 Descrição dos subtestes do SON-R 6-40

O SON-R 6-40 é formado por quatro subtestes:

1. Analogias
2. Mosaicos
3. Categorias
4. Padrões

Os subtestes são apresentados nesta sequência e podem ser divididos em dois tipos: testes de raciocínio abstrato (Analogias e Categorias) e testes de desempenho com uma orientação espacial (Mosaicos e Padrões). Os subtestes consistem de duas séries de treze itens (Mosaicos e Padrões) ou em três séries de doze itens (Categorias e Analogias), dispostos em ordem crescente de dificuldade. As tarefas de cada subteste são introduzidas com dois ou três exemplos e o examinando trabalha os itens de forma independente.

Analogias (ANA)

O subteste Analogias consiste de três séries de doze itens e é um teste de múltipla escolha. Cada item inclui um exemplo de uma analogia em que uma figura geométrica é alterada em uma ou mais formas para produzir uma outra forma geométrica. Isto é seguido por uma figura comparativa na qual o mesmo princípio de mudança deve ser aplicado. O examinando deve escolher a solução correta entre quatro alternativas. O nível de

dificuldade dos itens é relacionado ao número e complexidade das transformações. A

Figura 2.7 apresenta um exemplo de item desse subtteste.

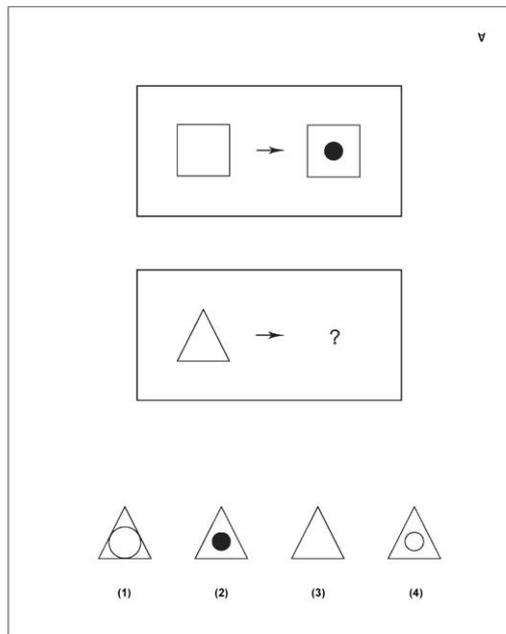


Figura 2.7 Exemplo de item do subtteste Analogias

Mosaicos (MOS)

O subtteste Mosaicos consiste de duas séries de treze itens. O objetivo aqui é copiar vários padrões de mosaicos em um quadro usando quadrados vermelhos, brancos e vermelhos/brancos. Para os itens mais fáceis são utilizados apenas quadrados vermelhos e brancos e o padrão de mosaico a ser copiado é mostrado em um quadro. Para os itens mais difíceis todos os quadrados são utilizados e o quadro é omitido. O maior tempo permitido aqui é de dois minutos para cada item. A Figura 2.8 apresenta um exemplo de item deste subtteste.

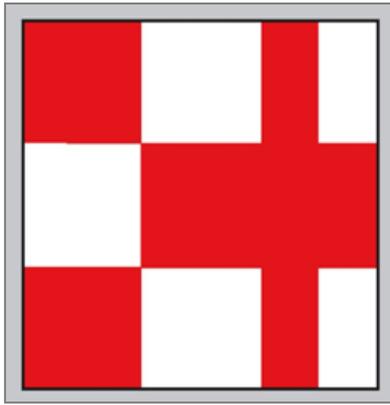


Figura 2.8 Exemplo de item do subtteste Mosaicos

Categorias (CAT)

O subtteste Categorias consiste de três séries de doze itens e é um teste de múltipla escolha. Neste subtteste, o examinando é apresentado a três figuras que possuem uma característica em comum. Duas novas figuras, com a mesma característica das três primeiras figuras, devem ser escolhidas de uma coluna de cinco figuras. O nível de dificuldade é determinado de acordo com o grau de abstração do conceito comum. A Figura 2.9 apresenta um exemplo de item desse subtteste.

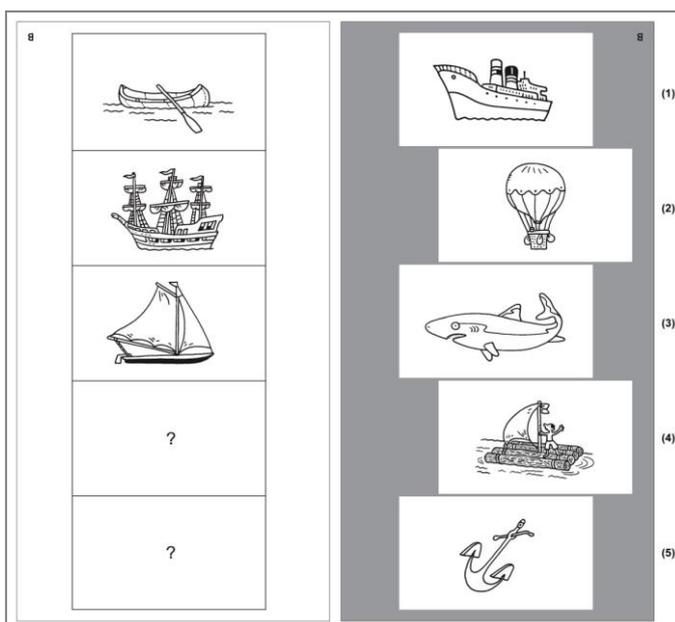


Figura 2.9 Exemplo de item do subtteste Categorias

Padrões (PAD)

O subtteste Padrões consiste de duas séries de treze itens. Cada item consiste de um padrão de linhas construído dentro de uma rede composta de quadrados. A parte do meio do padrão é omitida e precisa ser concluída utilizando-se um lápis. O nível de dificuldade é parcialmente determinado pelo número de linhas e o quanto do padrão foi omitido. O tempo máximo permitido aqui é de dois ou três minutos por item, dependendo do nível de dificuldade. A Figura 2.10 apresenta um exemplo de item desse subtteste.



Figura 2.10 Exemplo de item do subtteste Padrões.

2.4 Interpretação dos subttestes

O desempenho no SON-R 6-40 é direcionado para a resolução de problemas que requerem raciocínio abstrato e raciocínio espacial. O resultado no teste depende menos do conhecimento adquirido e mais da habilidade de descobrir métodos e regras e aplicá-los a novas situações. Dessa forma, o teste SON-R 6-40 busca mensurar a denominada “inteligência fluida”, compreendida como a habilidade de resolver problemas e de aprender, ao invés da “inteligência cristalizada” (Cattell, 1971; Carroll, 1993; Primi, 2002, McGrew, 2009). Entretanto, isso não quer dizer que a experiência obtida pelo respondente não influencia a sua habilidade de resolver problemas.

De acordo com os resultados da análise fatorial (ver capítulo 4), nos subttestes do SON-R 6-40 não existe uma distinção entre o raciocínio abstrato e espacial, como no teste SON-R 2½-7[a]. Apesar das diferenças entre os subttestes, observa-se que o raciocínio espacial e percepção visual desempenham um papel importante nos quatro subttestes.

Assim, o teste fornece somente um escore de QI total. A Tabela 2.1 oferece um sumário do conceito da escala do SON-R 6-40 e sugestão para a interpretação.

Tabela 2.1

A escala do SON-R 6-40 e sugestão para interpretação

Interpretação da escala do SON-R 6-40	
SON-QI	<ul style="list-style-type: none">• Avalia as habilidades de raciocínio para resolver problemas novos, que não podem ser executados automaticamente, e a capacidade de adaptação às situações novas, poucos estruturadas.• O SON-QI pretende medir a inteligência fluida, compreendida como a habilidade de aprender e resolver problemas sem usar conhecimento prévio.• Respondentes com altos escores geralmente apresentam facilidade em estabelecer relações abstratas, em desenvolver conceitos e solucionar problemas.

3. Pesquisa de normatização no Brasil

O primeiro passo para a pesquisa nacional foi construir o perfil dos respondentes do teste SON-R 6-40 nos diferentes municípios brasileiros. A população base para o processo de amostragem foram brasileiros recenseados de 6 a 40 anos de idade.

Durante a construção 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 Municipal - IDHM. Esta medida, IDHM, é uma mensuração derivada do Índice de Desenvolvimento Humano (IDH), ajustando o IDH para a realidade dos municípios e regiões metropolitanas.

O IDHM é uma medida composta de indicadores iguais ao IDH global – saúde, educação e renda –, assumindo valores entre 0 (valor mínimo) e 1 (valor máximo). Deve ser utilizada para comparar territórios nacionais entre si, como UFs, municípios, regiões metropolitanas e Unidades de Desenvolvimento Humano (UDH). O IDHM é classificado em cinco categorias:

- Muito baixo: entre 0 – 0,499
- Baixo: entre 0,500 – 0,599
- Médio: entre 0,600 – 0,699
- Alto: entre 0,700 – 0,799
- Muito alto: entre 0,800 - 1

O mapa a seguir apresenta a evolução do IDHM no Brasil no decorrer de quase 30 anos. Apesar de ter ocorrido uma grande mudança do IDHM dos municípios da região Norte e Nordeste, ainda é possível notar que grande parte dos municípios destas regiões possui IDHM menor que os municípios dos estados da região Sul e Sudeste.

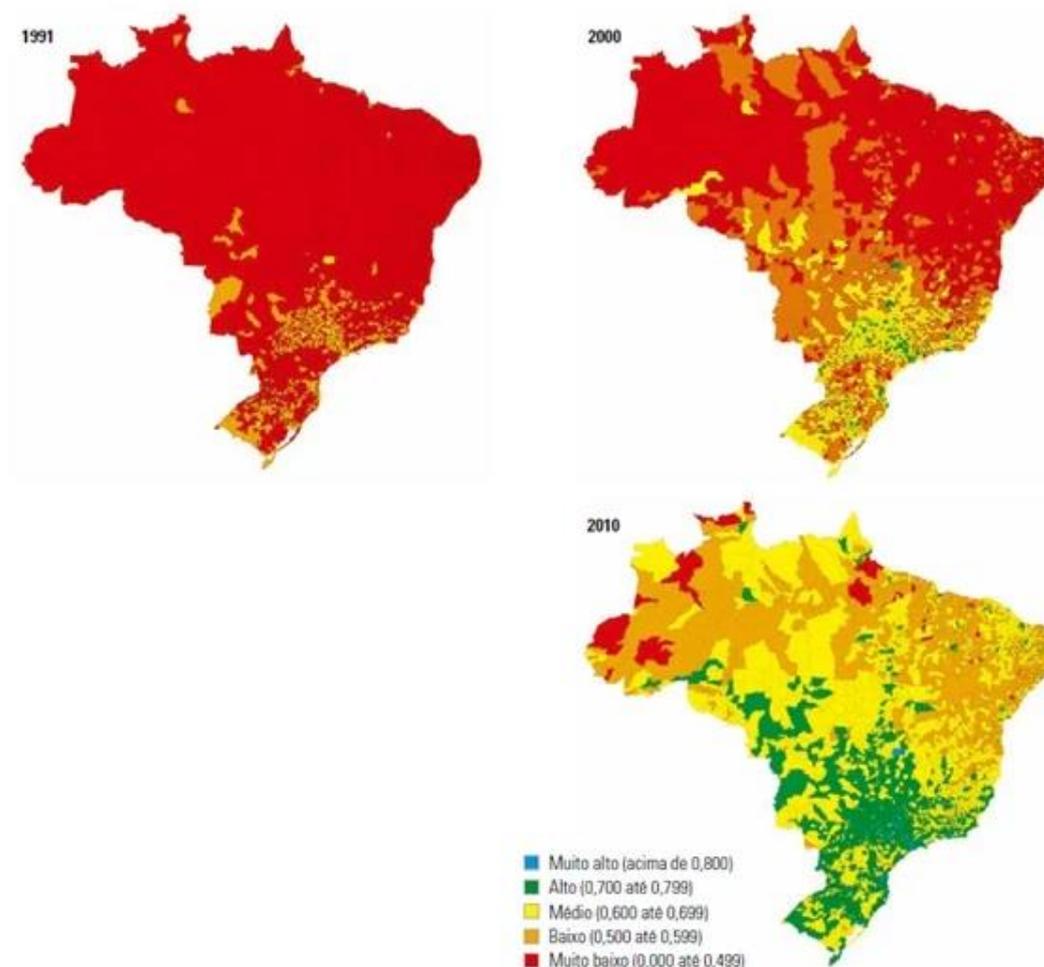


Figura 3.1 Evolução do IDHM no Brasil. Fonte: PNUD/2013.

Para a seleção dos municípios em cada região, além do IDHM, foram utilizados os seguintes critérios:

- (a) dentro de cada região deveria haver sempre uma cidade capital e uma cidade no interior;
- (b) as cidades selecionadas deveriam contemplar os maiores estados da região;
- (c) estados com maiores e menores IDHM do país deveriam ser contemplados na amostra;
- (d) disponibilidade de aplicativos no município ou próximo a ele;
- (e) facilidade de acesso.

A Tabela 3.1 mostra o plano que foi elaborado com base nesses critérios para a execução da pesquisa de normatização do SON-R 6-40. No total, 40 municípios foram selecionados, sendo São Caetano do Sul (SP) o município com o melhor IDHM do país e Melgaço (PA) o município com o pior IDHM (PNUD, 2013).

Tabela 3.1

Municípios selecionados em cada região do país

Região	Categorias IDHM	Municípios
Norte	$\geq 0,800$	
	entre 0,700 e 0,799	Belém (PA)
	entre 0,600 e 0,699	
	entre 0,500 e 0,599	
	entre 0 e 0,499	Melgaço (PA)
Nordeste	$\geq 0,800$	
	entre 0,700 e 0,799	Sobral (CE), Natal (RN), João Pessoa (PB), Salvador (BA), Santo Antônio de Jesus (BA)
	entre 0,600 e 0,699	Forquilha (CE), Monte Alegre (RN), Maragogipe (BA), Cruz das Almas (BA)
	entre 0,500 e 0,599	Cruz do Espírito Santo (PB), Lucena (PB),
	entre 0 e 0,499	
Centro-Oeste	$\geq 0,800$	Brasília (DF)
	entre 0,700 e 0,799	Anápolis (GO)
	entre 0,600 e 0,699	Abadiânia (GO)
	entre 0,500 e 0,599	
	entre 0 e 0,499	
Sudeste	$\geq 0,800$	São Paulo (SP), Santo André (SP), São Caetano do Sul (SP), Campinas (SP), Valinhos (SP), Belo Horizonte (MG), Vitória (ES)
	entre 0,700 e 0,799	Guarulhos (SP), Diadema (SP), Sumaré (SP), Bragança Paulista (SP), Uberlândia (MG), Santa Teresa (ES), Rio de Janeiro (RJ), São Gonçalo (RJ)
	entre 0,600 e 0,699	Indianópolis (MG), Belford Roxo (RJ), Itaboraí (RJ)
	entre 0,500 e 0,599	
	entre 0 e 0,499	

Tabela 3.1 (continuação)

Região	Categorias IDH-M	Municípios
Sul	≥ 0,800	Florianópolis (SC), Curitiba (PR)
	entre 0,700 e 0,799	Santa Maria (RS), São José dos Pinhais (PR), Almirante Tamandaré (PR), Águas Mornas (SC)
	entre 0,600 e 0,699	
	entre 0,500 e 0,599	
	entre 0 e 0,499	

A amostra brasileira para a normatização do SON-R 6-40 consiste de 1.360 pessoas divididas em 17 grupos de idade de 80 pessoas com percentuais iguais de homens e mulheres. A Tabela 3.2 apresenta os grupos de idade e como cada grupo é composto.

Tabela 3.2*Grupos de idade da amostra de normatização*

Grupo	Faixa etária	n de homens	n de mulheres	Total
1	6 anos 6 meses	40	40	80
2	7 anos 6 meses	40	40	80
3	8 anos 6 meses	40	40	80
4	9 anos 6 meses	40	40	80
5	10 anos 6 meses	40	40	80
6	11 anos 6 meses	40	40	80
7	12 anos 6 meses	40	40	80
8	13 anos 6 meses	40	40	80
9	14 anos 6 meses	40	40	80
10	15 anos 6 meses	40	40	80
11	16 anos 6 meses	40	40	80
12	18 anos 6 meses	40	40	80
13	20 anos 6 meses	40	40	80
14	22 anos 6 meses	40	40	80
15	27 anos 6 meses	40	40	80
16	32 anos 6 meses	40	40	80
17	37 anos 6 meses	40	40	80
Total		680	680	1.360

Para organizar a aplicação do teste nos diferentes municípios brasileiros, a amostra total de 1.360 pessoas foi dividida em 40 elementos (ou municípios) com 34 pessoas. Cada elemento possui 17 representantes do sexo masculino e 17 representantes do sexo feminino. Assim, cada elemento contém 34 pessoas ($17 \times 2 = 34$). A razão para a divisão em 40 elementos de 34 pessoas é evitar, por exemplo, o problema da desproporcionalidade, ou seja, avaliar jovens em cidades grandes e pessoas com mais idade em cidades pequenas, por exemplo; ou ainda, mais crianças no Sul e mais adultos no Norte. Então, para cada município da normatização, o número de pessoas avaliadas era para ter sido igual, isto é, 34 pessoas, sendo 17 do sexo masculino e 17 do sexo feminino, distribuídas nas 17 faixas etárias.

Além da quantidade de pessoas por município e a quantidade de participantes por sexo, outro critério na seleção dos sujeitos que participaram da pesquisa foi a idade. O critério de seleção por idade foi escolhido, principalmente, para evitar viés de seleção, aquele onde os professores selecionam as crianças a partir das suas preferências e impressões. Assim, o critério da idade ajuda na obtenção de uma amostra mais aleatória possível de crianças e adultos.

O critério da idade na seleção dos sujeitos também foi escolhido para garantir um grau de homogeneidade em relação às idades de cada faixa etária. Os participantes deveriam ter a idade mais próxima possível de X anos e 6 meses. Para as crianças do ensino fundamental (6 anos aos 14 anos), foi permitido um desvio da idade ideal de 3 meses. Assim, a criança poderia ter entre 6 anos e 3 meses a 6 anos e 9 meses. Para os adultos, o desvio permitido foi de 4 meses. Esse critério é especialmente importante para as faixas etárias de crianças mais novas, pois nesses grupos um período de 3 meses pode representar um grau significativo de desenvolvimento (Jesus, 2009). A Tabela 3.3

apresenta o exemplo de um elemento, isto é, como os dados deveriam ser coletados em cada município.

Tabela 3.3

Exemplo de um elemento da pesquisa de normatização

Faixa etária	Desvio da idade ideal	nº de homens	nº de mulheres
6 anos 6 meses	3 meses	1	1
7 anos 6 meses	3 meses	1	1
8 anos 6 meses	3 meses	1	1
9 anos 6 meses	3 meses	1	1
10 anos 6 meses	3 meses	1	1
11 anos 6 meses	3 meses	1	1
12 anos 6 meses	3 meses	1	1
13 anos 6 meses	3 meses	1	1
14 anos 6 meses	3 meses	1	1
15 anos 6 meses	4 meses	1	1
16 anos 6 meses	4 meses	1	1
18 anos 6 meses	4 meses	1	1
20 anos 6 meses	4 meses	1	1
22 anos 6 meses	4 meses	1	1
27 anos 6 meses	4 meses	1	1
32 anos 6 meses	4 meses	1	1
37 anos 6 meses	4 meses	1	1
		17	17

Por fim, foi estabelecido também o perfil da amostra normativa, isto é, seleção de variáveis relacionadas à situação educacional e econômica dos respondentes. Assim, em cada cidade, deveriam ser aplicados 34 testes observando o sexo, a faixa etária, a situação educacional, a situação econômica e a cidade previamente estabelecidos. Além disso, foram selecionadas escolas particulares tanto na capital quanto no interior.

Entretanto, após o início da coleta de dados, o plano amostral precisou ser ajustado para atender as dificuldades que foram surgindo durante a aplicação do teste. No final, o teste foi aplicado em 39 municípios das cinco regiões brasileiras. A Tabela 3.4 apresenta as cidades onde foi realizada a coleta dos dados.

Tabela 3.4
Plano amostral ajustado

Região	Categorias IDHM	Municípios
Norte	≥ 0,800	
	entre 0,700 e 0,799	Belém (PA)
	entre 0,600 e 0,699	Marituba (PA)
	entre 0,500 e 0,599	
	entre 0 e 0,499	
Nordeste	≥ 0,800	
	entre 0,700 e 0,799	Sobral (CE), Natal (RN), João Pessoa (PB), Salvador (BA), Santo Antônio de Jesus (BA)
	entre 0,600 e 0,699	Forquilha (CE), Crateús (CE) , Monte Alegre (RN), Maragogipe (BA), Cruz das Almas (BA)
	entre 0,500 e 0,599	Cruz do Espírito Santo (PB), Lucena (PB),
	entre 0 e 0,499	
Centro-Oeste	≥ 0,800	Brasília (DF)
	entre 0,700 e 0,799	Anápolis (GO)
	entre 0,600 e 0,699	Abadiânia (GO)
	entre 0,500 e 0,599	
	entre 0 e 0,499	
Sudeste	≥ 0,800	São Paulo (SP), Santo André (SP), São Caetano do Sul (SP), Campinas (SP), Valinhos (SP), Belo Horizonte (MG), Vitória (ES)
	entre 0,700 e 0,799	Guarulhos (SP), Diadema (SP), Sumaré (SP), Carapicuíba (SP) , Itatiba (SP) , Bragança Paulista (SP), Uberlândia (MG), Santa Teresa (ES), Rio de Janeiro (RJ), São Gonçalo (RJ)
	entre 0,600 e 0,699	Patrocínio (MG) , Belford Roxo (RJ)
	entre 0,500 e 0,599	
	entre 0 e 0,499	
Sul	≥ 0,800	Florianópolis (SC)
	entre 0,700 e 0,799	Santa Maria (RS), Canoas (RS)
	entre 0,600 e 0,699	
	entre 0,500 e 0,599	
	entre 0 e 0,499	

Nota. Em negrito, os municípios incluídos após o início da coleta.

Destaca-se que, mesmo após os ajustes, a maioria dos critérios estabelecidos previamente foi atendida. Dessa forma, foram mantidos a quantidade de respondentes

(N = 1.360), a quantidade de respondentes em cada grupo de idade (n=80) e o percentual igual de homens e mulheres (50%). Pequenos ajustes precisaram ser realizados no perfil dos respondentes, nos municípios onde ocorreram a aplicação do teste e a quantidade de respondentes por município. A Tabela 3.5 apresenta as alterações no perfil dos respondentes.

Tabela 3.5
Perfil dos respondentes da pesquisa de normatização

Grupo	Perfil	% previsto	% obtido
A 6 a 14 anos (N= 720)	Escolas Públicas	87,0%	82,5%
	Escolas Particulares	13,0%	17,5%
B 15 a 16 anos (N=160)	Escolas Públicas	87,0%	70,0%
	Escolas Particulares	11,8%	15,0%
	Escolas Federais	2,5%	2,5%
	Adolescentes fora da escola	12,5%	12,5%
C 18 a 40 anos (N=480)	Diretores/generais/presidentes/gestores	3,8%	2,5%
	Gerentes/funcionários públicos/tenentes/capitães	12%	10,0%
	Empregados com baixa renda /trabalhadores/soldados	71,7%	75,0%
	Beneficiários de programas sociais, por exemplo, bolsa família	12,5%	12,5%

Apesar da alteração no perfil dos respondentes da amostra normativa, observa-se que o percentual obtido no grupo A e B ainda se aproxima do percentual de crianças e adolescentes que estudam na rede pública (80%), particular (20%) e federal (0,4%) brasileira (Inep, 2016). No caso dos adultos, o perfil foi estabelecido a partir da renda/escolaridade. Também se optou 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. Quanto à alteração da quantidade de respondentes por município, a Tabela 3.6 apresenta o quantitativo exato por cidade.

Tabela 3.6*Quantidade de respondentes por municípios*

Município (UF)	N
Abadiânia (GO)	34
Anápolis (GO)	34
Belém (PA)	68
Belford Roxo (RJ)	34
Belo Horizonte (MG)	34
Bragança Paulista (SP)	34
Brasília (DF)	56
Campinas (SP)	34
Canoas (RS)	112
Carapicuíba (SP)	19
Cruz do Espírito Santo (PB)	34
Crateús (CE)	19
Cruz das Almas (BA)	34
Diadema (SP)	15
Florianópolis (SC)	41
Forquilha (CE)	34
Guarulhos (SP)	34
Itatiba (SP)	4
João Pessoa (PB)	34
Lucena (PB)	34
Maragogipe (BA)	34
Marituba (PA)	34
Monte Alegre (RN)	34
Natal (RN)	34
Patrocínio (MG)	34
Rio de Janeiro (RJ)	34
Salvador (BA)	34
Santa Maria (RS)	32
Santa Teresa (ES)	34
Santo André (SP)	29
Santo Antônio de Jesus (BA)	34
São Caetano do Sul (SP)	17
São Gonçalo (RJ)	34
São Paulo (SP)	34
Sobral (CE)	34
Sumaré (SP)	30
Uberlândia (MG)	34
Valinhos (SP)	34
Vitória (ES)	34
Total	1.360

Procedimento utilizado na pesquisa nacional de normatização

Os próprios aplicadores eram responsáveis por entrar em contato com as escolas e secretarias de educação, igrejas e ONGs. Durante o primeiro contato com as instituições, foram explicitados os objetivos, procedimentos, justificativa da pesquisa e foi solicitada a permissão da escola, por exemplo, para a realização da coleta dos dados nas dependências da instituição.

Para as escolas que concordaram em participar da pesquisa, foram entregues Termos de Consentimento Livre e Esclarecido e o Questionário Contextual para serem entregues aos pais. Os pais que aceitaram que seus filhos participassem da pesquisa, receberam um relatório breve com as seguintes informações: desempenho da criança no teste e sugestão de algumas atividades que pudessem contribuir para o desenvolvimento dos tipos de raciocínio avaliados.

Foram adotados os seguintes critérios de exclusão de participantes da amostra da pesquisa de normatização: sujeitos diagnosticados com algum problema físico ou condição que poderia afetar seu desempenho no teste; sujeitos com problema na visão; sujeitos com problema de audição. Outros estudos com pessoas que apresentam alguma deficiência específica foram realizados em outros momentos e os resultados serão apresentados no capítulo cinco deste relatório científico.

Quanto aos aplicadores, foram treinados em Brasília ou na região onde ocorreria a aplicação por um membro da equipe da coordenação da pesquisa. Em todos os treinamentos, os aplicadores receberam informações sobre o construto avaliado (inteligência), informações relacionadas aos testes SON, a relevância da pesquisa de normatização e validação do SON-R 6-40 e orientações sobre a amostragem. Além disso, o treinamento incluía uma parte prática, onde os aplicadores treinaram a aplicação do SON-R 6-40 em duplas.

Durante a pesquisa, diferentes psicólogos foram recrutados e agregados à equipe de aplicadores da pesquisa de normatização do SON-R 6-40 no Brasil. No total, 26 profissionais aplicaram o teste nas suas regiões. A Tabela 3.7 apresenta o nome dos aplicadores e as cidades onde cada um aplicou o teste.

Tabela 3.7

Nome dos examinadores e municípios.

Nome dos examinadores	Estado/Municípios	N
Ana Cláudia Fernandes	Rio de Janeiro: Rio de Janeiro, Belford Roxo e São Gonçalo	91
Andressa Tonini	Espírito Santo: Vitória e Santa Teresa	68
Annelise Carminatti	Santa Catarina: Florianópolis	41
Beatriz Marinho	Rio de Janeiro: Rio de Janeiro, Belford Roxo e São Gonçalo	11
Daniela Carcavila	São Paulo: São Paulo	12
Eliete Neves	Goiás: Anápolis e Abadiânia	68
Elisângela Alencar	Distrito Federal: Brasília	22
Elisete Cardoso	São Paulo: São Paulo e Guarulhos	34
Éverson Meireles	Bahia: Cruz das Almas e Santo Antônio de Jesus	68
Fábio Alves	Minas Gerais: Uberlândia e Patrocínio	68
Hunayara Tavares	Minas Gerais: Belo Horizonte	34
Janaína Nogueira	Pará: Belém e Marituba	102
Juliana Damásio	Rio Grande do Norte: Natal e Monte Alegre	68
Karoline Ribeiro	São Paulo: São Caetano do Sul, Santo André, Diadema e Carapicuíba	43
Larissa Lemos	Paraíba: João Pessoa, Cruz do Espírito Santo e Lucena	102
Lisandra Borges	São Paulo: Valinhos, Bragança Paulista e Itatiba	43
Luana Janssen Teixeira	Rio Grande do Sul: Canoas	112
Lucas Donato	São Paulo: São Paulo	7
Ludiana Moreno	Bahia: Salvador e Maragogipe	68
Márcia Santos	Rio Grande do Sul: Santa Maria	32
Marjorie Silva	São Paulo: Campinas	93
Paloma Vilhena	São Paulo: São Paulo	27
Renata Manuely	Distrito Federal: Brasília	34
Silvânia Onça	São Paulo: São Paulo	25
Suiany Ponte	Ceará: Sobral e Forquilha	68
Thelmara Bezerra	Ceará: Crateús	19
Total		1.360

Amostra da pesquisa de normatização

Todas as aplicações do teste SON-R 6-40 foram realizadas individualmente. A Tabela 3.8 sumariza algumas informações sobre o grupo amostral selecionado que respondeu o teste.

Tabela 3.8

Informações dos respondentes da pesquisa de normatização (N=1.360)

Região	Frequência	Percentual
Nordeste	393	28,9%
Norte	102	7,5%
Sudeste	556	40,9%
Centro-Oeste	124	9,1%
Sul	185	13,6%
Sexo	Frequência	Percentual
Feminino	680	50%
Masculino	680	50%
Localidade	Frequência	Percentual
Capital	403	29,6%
Interior	957	70,4%

Seguindo o delineamento do estudo de normatização realizado na Holanda, foram estabelecidos 17 grupos de idade formado por 80 pessoas, totalizando 1.360 participantes. Entre as idades de 6 a 16 anos, foram utilizados grupos de idade por ano. A partir da idade de 16 anos, foram aplicados intervalos entre os grupos de idade, pois os resultados de pesquisas anteriores indicam que não há uma alteração ou melhora significativa no desempenho cognitivo na idade adulta (Cliffordson & Gustafsson, 2008). Ackerman (1996) aponta que a inteligência fluida tem o seu pico de desenvolvimento no final da adolescência, mantendo-se estabilizada por alguns poucos anos e começando a declinar na fase adulta.

A Tabela 3.9 apresenta a estrutura geral dos grupos de idade da amostra normativa. Os grupos de idade foram classificados em três categorias: grupo A, formado por seis grupos de idade (6 a 11 anos); grupo B, formado por cinco grupos de idade (12 a 16 anos);

e o grupo C, composto por seis grupos de idade. Essa divisão foi útil para realizar algumas análises que serão apresentadas no capítulo sobre as características psicométricas.

A Tabela 3.9 mostra também a média da idade dos 17 grupos etários. Observa-se que a média ficou muito próxima da idade planejada de X anos e 6 meses. Este resultado traz uma forte evidência que os aplicadores realmente selecionaram os participantes com uma idade o mais próxima possível do que foi estabelecido.

Tabela 3.9
Estrutura geral dos grupos da amostra normativa (N=1.360)

Grupo A 6 a 11 anos			Grupo B 12 a 16 anos			Grupo C 18 a 40 anos		
Idade	N	Média	Idade	N	Média	Idade	N	Média
6 anos	80	6,55	12 anos	80	12,53	18 anos	80	18,45
7 anos	80	7,50	13 anos	80	13,50	20 anos	80	20,46
8 anos	80	8,50	14 anos	80	14,49	22 anos	80	22,46
9 anos	80	9,49	15 anos	80	15,50	27 anos	80	27,45
10 anos	80	10,53	16 anos	80	16,46	32 anos	80	32,47
11 anos	80	11,48				37 anos	80	37,45
Total	480		Total	400		Total	480	

A Tabela 3.10 apresenta informações quanto ao local de aplicação do teste, período do dia que o teste foi aplicado e o tempo médio, em minutos, que os participantes precisaram para responder o teste. Observa-se que a maioria das crianças e adolescentes responderam o teste nas dependências de uma unidade escolar, durante o dia. A média do tempo de aplicação do teste aumentou com a idade dos respondentes: de 52,5 minutos para o grupo A (6 a 11 anos) até 68,6 minutos para o grupo C (18 a 40 anos).

Tabela 3.10
Características da administração do SON-R 6-40

		Grupo A		Grupo B		Grupo C	
		N	%	N	%	N	%
Local de aplicação	Escola	363	75,6%	297	74,3%	98	20,4%
	Casa do participante	39	8,1%	41	10,3%	150	31,3%
	Outros	54	11,3%	44	11,0%	214	44,6%
	Sem informação	24	5,0%	18	4,5%	18	3,8%
	Total	480	100%	400	100%	480	100%
Momento de aplicação	Manhã	228	47,5%	217	54,3%	147	30,6%
	Tarde	248	51,7%	165	41,3%	209	43,5%
	Noite	4	0,8%	18	4,5%	123	25,6%
Duração da aplicação (em minutos)		Média	(DP)	Média	(DP)	Média	(DP)
		52,5	(15,1)	61,6	(17,5)	68,6	(17,2)

O modelo de normatização

As normas brasileiras do teste SON-R 6-40 foram estabelecidas em cooperação com o professor Peter Tellegen da Universidade de Groningen, utilizando um modelo para calcular as normas onde as distribuições dos escores brutos para todos os grupos de idade são ajustadas simultaneamente como uma função da idade usando regressão para cada subteste separadamente. A função contínua 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 distribuição populacional estimada, limitando assim o número de parâmetros da função.

Este método é chamado de normatização contínua (*continuous test norming*) ou normatização em base de regressão (*regression based test norming*), proposto por Zachary e Gorsuch (1985). Esse método também foi utilizado para a normatização do SON-R 5½-17 (Snijders, Tellegen & Laros, 1989).

Oosterhuis (2017) pontua que o grupo de normatização deve ser representativo da população alvo e suficientemente grande para assegurar que as normas apresentem pequenos erros padrão. Dessa forma, a composição da amostra deve refletir com precisão

a estrutura dos subgrupos da população que é relevante para a característica que é medida pelo teste. Por exemplo, se a idade é uma covariável relevante, os subgrupos que serão formados a partir desta variável devem ser grandes o suficiente para garantir a estimativa de normas precisas. Conseqüentemente, quanto maior o número de subgrupos, maior será a amostra total. Porém, coletar dados em amostras grandes e representativas para cada subgrupo é difícil, demorado e caro, levando os construtores de testes a limitar o número de subgrupos. Entretanto, essa redução de subgrupos pode levar ao problema da criação de categorias arbitrárias.

No caso da mensuração da inteligência, a variável idade é reconhecida como um fator que influencia diretamente o escore (Kittrell, 1999). Assim, a prática comum no procedimento tradicional de normatização é a construção de subgrupos, reunindo amplas faixas de idade. Oosterhuis (2017) e Kittrell (1999) explicam que uma diferença de idade de apenas alguns dias pode resultar em uma interpretação totalmente diferente dos resultados dos testes de indivíduos que estão próximos dos limites dessas categorias etárias. Então, para evitar esse problema, os construtores de testes podem criar um número maior de categorias de idade, mas isso resultará em uma amostra de normatização maior.

O método de normatização contínua apresenta vantagens que podem auxiliar os construtores dos testes quanto à questão do tamanho do grupo normativo. O tamanho necessário da amostra total e dos subgrupos é menor quando o método de normatização contínua é utilizado. Estudos de simulação apontam que esse método requer amostras 2,5 a 5,5 vezes menores que o processo de normatização tradicional (Oosterhuis, 2017; Voncken, Albers, & Timmerman, 2017; Kittrell, 1999).

Além disso, Oosterhuis (2017) e Kittrell (1999) alertam que a criação de subgrupos é adequada para variáveis nominais, por exemplo, estabelecimento de normas a partir do sexo. Entretanto, tal prática não é recomendada para variáveis contínuas, por exemplo,

idade, pois dividir uma variável contínua em categorias gera perda de informação e introduz erro (Kittrell, 1999). A título de exemplo, Zachary e Gorsuch (1985) mostram que o QI do WAIS-R de uma pessoa nos limites das categorias etárias pode ganhar até seis pontos com a transição de um mês de idade, fazendo que uma tabela diferente precise ser usada, gerando uma interpretação diferente. Assim, o objetivo da normatização contínua é fornecer estimativas de um dado parâmetro para a amostra normativa como uma função contínua da variável condicionante/relevante (Kittrell, 1999).

A Comissão Nacional de Avaliação de Testes dos Países Baixos (COTAN) tem estimulado os construtores de testes a utilizar o método de normatização contínua devido às vantagens que ele apresenta (Evers et al., 2009). As principais vantagens são: (1) os escores normatizados podem ser calculados com base na idade exata das pessoas; (2) o uso da informação de todos os grupos de idade simultaneamente torna a normatização mais precisa, (3) as amostras de normatização dos subgrupos e a amostra total não precisam ser tão grande como na normatização tradicional para obter resultados estáveis, (4) o processo de ajustamento usando a regressão evita as discontinuidades que frequentemente acontecem na normatização tradicional, e (5) o modelo permite fazer extrapolações fora da faixa de idade na qual a pesquisa de normatização foi feita (Laros & Tellegen, 1991; Kittrell, 1999; Oosterhuis, 2017; Lenhard, Lenhard, Suggate & Segerer, 2018).

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

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

Foram adotados dez passos no processo de normatização. A Figura 3.2 apresenta a sequência das etapas efetuadas.

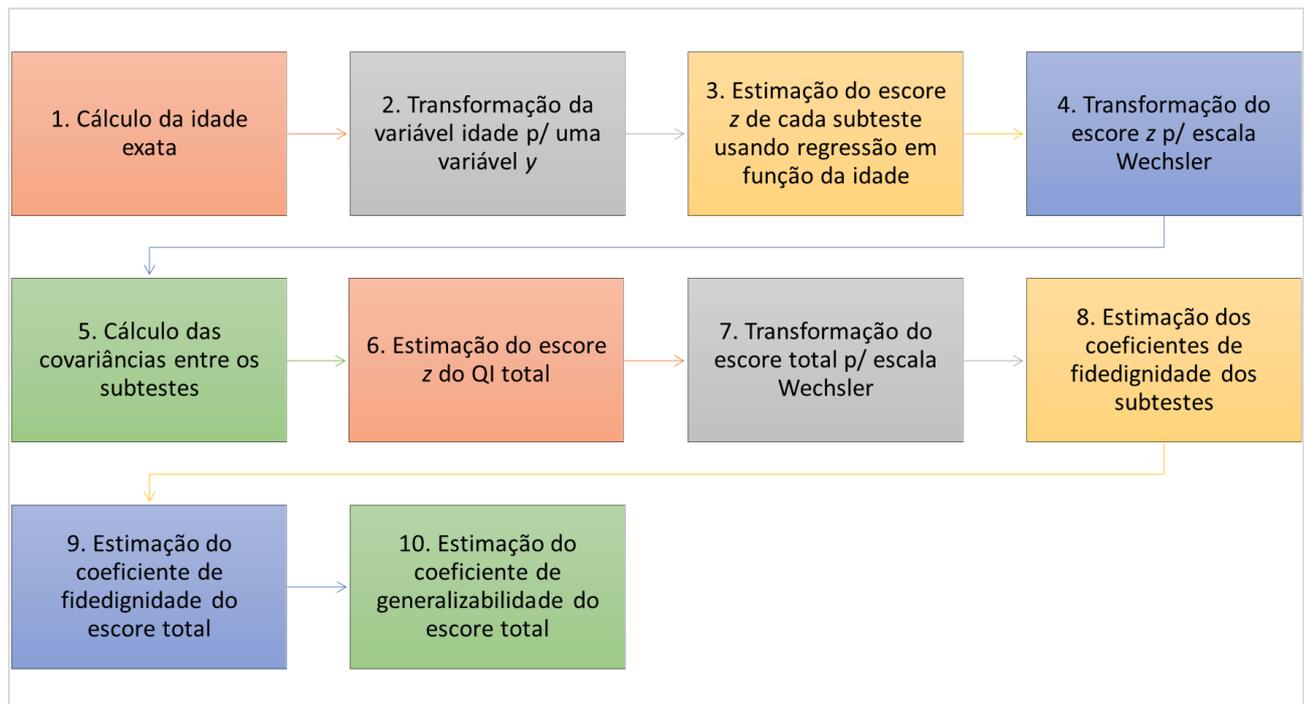


Figura 3.2 Passos adotados no processo de normalização dos escores brutos.

O primeiro passo no processo de normalização foi calcular a idade exata de cada participante da pesquisa de normalização. Para tanto, foi utilizada a fórmula descrita abaixo:

$$\textit{idade exata} = (\textit{ano de aplicação} - \textit{ano do nascimento}) + (\textit{mês de aplicação} - \textit{mês do nascimento})/12 + (\textit{dia de aplicação} - \textit{dia do nascimento})/365,25$$

Após o cálculo da idade exata, a variável foi transformada em uma variável y , com valores entre -1 e +1, utilizando a seguinte fórmula:

$$y = (\textit{idade exata} - 23)/17$$

A transformação da idade exata para a variável y com valores entre -1 e +1 foi feita para evitar valores muito grandes da idade exata depois de elevar a variável y à segunda e terceira potência ($40^2 = 1.600$; $40^3 = 64.000$). A entrada de y^2 e y^3 nas análises de regressão para calcular os escores normalizados foi necessária uma vez que a relação entre a idade e os escores brutos é uma relação curvilínea. Quando se espera uma relação

curvilínea, é necessário que potências superiores (y^2 e y^3) sejam testadas juntamente com a própria variável independente (y) (Kittrell, 1999). Assim, entraram na análise de regressão, além de y , y^2 e y^3 .

$$y^2 = y \times y$$

$$y^3 = y \times y \times y$$

O passo seguinte foi calcular o escore z de cada subtteste. Para obter o escore z , o escore normatizado foi dividido pela raiz quadrada da variância. Este cálculo foi feito da seguinte forma:

$$\text{Escore normatizado } z \text{ Analogias} = (\text{escore bruto Analogias} - (\text{constante} + \text{coeficiente de regressão} \cdot y - \text{coeficiente de regressão} \cdot y^2 + \text{coeficiente de regressão} \cdot y^3)) / (\sqrt{(37.483 + 17.222 \cdot y)}).$$

$$\text{Escore normatizado } z \text{ Mosaicos} = (\text{escore bruto Mosaicos} - (\text{constante} - \text{coeficiente de regressão} \cdot y^2 + \text{coeficiente de regressão} \cdot y^3)) / (\sqrt{(19.179 + 2.565 \cdot y - 6.271 \cdot y^2)})$$

$$\text{Escore normatizado } z \text{ Categorias} = (\text{escore bruto Categorias} - (\text{constante} + \text{coeficiente de regressão} \cdot y - \text{coeficiente de regressão} \cdot y^2 + \text{coeficiente de regressão} \cdot y^3)) / (\sqrt{(39.466 + 27.563 \cdot y^3)}).$$

$$\text{Escore normatizado } z \text{ Padrões} = (\text{escore bruto Padrões} - (\text{constante} - \text{coeficiente de regressão} \cdot y^2 + \text{coeficiente de regressão} \cdot y^3)) / (\sqrt{(22.825 + 3.307 \cdot y - 7.530 \cdot y^2 - 2.792 \cdot y^3)}).$$

Após o cálculo do escore z , estes escores foram transformados para os escores da escala Wechsler, com média 10 e desvio padrão 3. Este cálculo foi feito da seguinte forma (os escores Wechsler foram arredondados para números inteiros):

$$\text{Escore normatizado Wechsler Analogias} = 10 + (3 \times \text{escore normatizado } z \text{ Analogias})$$

$$\text{Escore normatizado Wechsler Mosaicos} = 10 + (3 \times \text{escore normatizado } z \text{ Mosaicos})$$

Escore normalizado Wechsler Categorias = 10 + (3 x escore normalizado z Categorias)

Escore normalizado Wechsler Padrões = 10 + (3 x escore normalizado z Padrões).

O próximo passo foi o cálculo das covariâncias entre os subtestes – as covariâncias entre os subtestes são iguais às correlações entre os subtestes, uma vez que foram calculados em base dos escores z. O resultado deste cálculo foi:

Covariância entre Analogias e Mosaicos = 0,651

Covariância entre Analogias e Categorias = 0,655

Covariância entre Analogias e Padrões = 0,646

Covariância entre Mosaicos e Categorias = 0,576

Covariância entre Mosaicos e Padrões = 0,722

Covariância entre Categorias e Padrões = 0,560

Com o cálculo das covariâncias entre os subtestes foi possível calcular o escore normalizado do escore total. Para obter o escore z, o escore normalizado foi dividido pela raiz quadrada da variância:

Escore total normalizado z = (Escore normalizado z Analogias + Escore normalizado z Mosaicos + Escore normalizado z Categorias + Escore normalizado z Padrões) / √ (4 + 2 x (Covariância Analogias-Mosaicos + Covariância Analogias-Categorias + Covariância Analogias-Padrões + Covariância Mosaicos-Categorias + Covariância Mosaicos-Padrões + Covariância Categorias-Padrões)).

O escore total normalizado z foi transformado para a escala de QI usando a seguinte fórmula:

Escore total normalizado Escala QI = 100 + (15 x Escore normalizado z).

Depois de obter os escores normatizados dos subtestes e do escore total, foram estimados os coeficientes de fidedignidade dos subtestes como uma função da idade exata usando a seguinte fórmula:

$$\alpha \text{ Analogias} = 0,909 + 0,038y$$

$$\alpha \text{ Mosaicos} = 0,893 - 0,034y^2$$

$$\alpha \text{ Categorias} = 0,912 + 0,064y^3$$

$$\alpha \text{ Padrões} = 0,893.$$

O coeficiente de fidedignidade do escore total foi calculado usando a fórmula abaixo:

$$\text{Alfa escore total} = 1 - (\sum \text{variância do erro} / \text{Variância total}).$$

$$\text{Alfa} = 1 - [(1 - \alpha \text{ Analogias}) + (1 - \alpha \text{ Mosaicos}) + (1 - \alpha \text{ Categorias}) + (1 - \alpha \text{ Padrões})] / [4 + 2x (\text{Covariância Analogias-Mosaicos} + \text{Covariância Analogias-Categorias} + \text{Covariância Analogias-Padrões} + \text{Covariância Mosaicos-Categorias} + \text{Covariância Mosaicos-Padrões} + \text{Covariância Categorias-Padrões})].$$

Como último passo, foi calculado o coeficiente de generalizabilidade do escore total usando a fórmula de Alfa de Cronbach (Hogan, 2006) tomando os escores dos subtestes como a unidade de análise em vez de os escores nos itens:

$$\text{Generalizabilidade escore total} = ((4 / 4 - 1)(1 - (\sum \text{variância dos subtestes} / \text{Variância total})).$$

4. Características Psicométricas

Este capítulo trata das características psicométricas do SON-R 6-40 com base na amostra normativa. Os dados foram analisados por meio de procedimentos estatísticos paramétricos de modelagem e análise de itens, usando os seguintes pacotes estatísticos: *SPSS 21.0*, *AMOS 21.0*, *FACTOR 10.8* e *BILOG-MG*. Foram realizadas análises de dificuldade e discriminação dos itens por meio da Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI), análise de fidedignidade e análise da dimensionalidade usando quatro métodos: Método *Hull*, Análise Paralela *Optimal Implementation*, Análise Paralela *Classical Implementation* e Schwarz's *Bayesian Information Criterion Dimensionality Test*. A estrutura fatorial do SON-R 6-40 foi examinada por meio da *Minimum Rank Factor Analysis* (MRFA) e Análise Fatorial Confirmatória (CFA).

Análise dos itens segundo a TCT

4.1 Dificuldade dos itens

A análise de itens pela TCT, geralmente, inclui a análise do índice de dificuldade e do poder discriminativo dos itens. Pela TCT, estima-se a dificuldade a partir do valor p (proporção de pessoas que responderam o item corretamente) (Primi, 2012). A Tabela 4.1 apresenta a dificuldade dos itens nas séries *a*, *b* e *c* para os subtestes Analogias e Categorias e nas séries *a* e *b* para os subtestes Mosaicos e Padrões. Observa-se que, nos subtestes Analogias e Mosaicos, todos os itens dentro da mesma série estão em ordem crescente de dificuldade. Já no subteste Categorias, um dos itens, o item 4 da série *c*, apresentou dificuldade maior que o item seguinte. Por fim, o item 1 das duas séries do subteste Padrões apresentou maior dificuldade que o item posterior. Apesar de três itens apresentarem dificuldade maior que o item seguinte, observa-se que há poucos desvios e

a diferença entre o valor p dos itens não ultrapassou 0,03. Resultados semelhantes foram encontrados na normatização do SON-R 6-40 na Holanda/Alemanha.

Tabela 4.1
Dificuldade dos itens dos subtestes do SON-R 6-40 (N = 1.360)

Analogias					Categorias				
Item	série a	série b	série c	média	Item	série a	série b	série c	média
1	1,00	1,00	0,98	0,99	1	0,97	0,97	0,97	0,97
2	0,96	0,96	0,96	0,96	2	0,96	0,92	0,97	0,95
3	0,93	0,91	0,92	0,92	3	0,93	0,88	0,91	0,91
4	0,84	0,86	0,87	0,85	4	0,84	0,82	0,77*	0,83
5	0,75	0,73	0,75	0,74	5	0,64	0,71	0,80	0,72
6	0,40	0,51	0,58	0,49	6	0,37	0,45	0,54	0,45
7	0,38	0,43	0,49	0,43	7	0,33	0,35	0,42	0,37
8	0,28	0,30	0,36	0,31	8	0,18	0,34	0,35	0,29
9	0,18	0,19	0,34	0,23	9	0,14	0,19	0,23	0,18
10	0,10	0,10	0,18	0,12	10	0,09	0,13	0,12	0,11
11	0,07	0,07	0,10	0,08	11	0,05	0,10	0,08	0,08
12	0,03	0,04	0,05	0,04	12	0,03	0,04	0,04	0,03
Média	0,49	0,51	0,55	0,52	Média	0,46	0,49	0,49	0,49
Mosaicos				Padrões					
Item	série a	série b	média	Item	série a	série b	média		
1	0,99	0,99	0,99	1	0,95*	0,92*	0,93		
2	0,98	0,98	0,98	2	0,96	0,95	0,95		
3	0,87	0,90	0,88	3	0,89	0,90	0,89		
4	0,87	0,88	0,87	4	0,84	0,88	0,86		
5	0,74	0,79	0,76	5	0,71	0,75	0,73		
6	0,54	0,71	0,62	6	0,67	0,72	0,69		
7	0,40	0,61	0,50	7	0,55	0,62	0,58		
8	0,37	0,47	0,42	8	0,30	0,43	0,36		
9	0,22	0,25	0,23	9	0,25	0,31	0,28		
10	0,13	0,20	0,16	10	0,13	0,21	0,17		
11	0,05	0,11	0,08	11	0,10	0,14	0,12		
12	0,04	0,07	0,05	12	0,04	0,07	0,05		
13	0,01	0,03	0,02	13	0,03	0,04	0,03		
Média	0,47	0,53	0,51	Média	0,49	0,53	0,51		

Nota. * item é mais difícil que o item seguinte.

No teste SON-R 6-40 é importante garantir a dificuldade progressiva dos itens dentro de cada série porque o procedimento adaptativo usado na aplicação do SON-R 6-40 é baseado no pressuposto de uma dificuldade progressiva dos itens (Laros & Tellegen, 1991). Para conseguir construir itens em ordem progressiva foi desenvolvida uma teoria

de dificuldade para cada um dos subtestes (Snijders, Tellegen & Laros, 1989; Geerlings, Laros, Tellegen & Glas, 2011).

4.2 Discriminação dos itens

Na TCT, a discriminação refere-se ao potencial que os itens têm de diferenciar estatisticamente os sujeitos com diferentes escores totais no subteste (Hogan, 2006). Para verificar o parâmetro da discriminação dos itens segundo a TCT, foi calculada a correlação item-total corrigida, que é uma medida de associação entre o desempenho no item e o desempenho no subteste onde o próprio item é retirado do escore do subteste. A Tabela 4.2 apresenta os valores da correlação item-total corrigida dos itens de todos os subtestes. Observa-se que valores relativamente baixos ($< 0,30$) aparecem apenas nos itens iniciais e finais de cada série. A variância destes itens com proporção de acerto muito alta ou muito baixa é reduzida, o que leva à correlações baixas. A inspeção da Tabela 4.2 revela que os itens dos subtestes de execução Padrões e Mosaicos têm maior poder discriminativo: as médias da correlação item-total corrigida são 0,56 e 0,50, respectivamente.

Tabela 4.2Correlação item-total dos itens dos subtestes do SON-R 6-40 ($N = 1.360$)

Analogias					Categorias				
Item	série a	série b	série c	média	Item	série a	série b	série c	média
1	0,11	0,11	0,24	0,15	1	0,16	0,34	0,32	0,27
2	0,28	0,36	0,35	0,33	2	0,36	0,38	0,29	0,34
3	0,37	0,43	0,45	0,42	3	0,37	0,45	0,45	0,42
4	0,48	0,49	0,52	0,50	4	0,46	0,49	0,51	0,49
5	0,53	0,57	0,59	0,56	5	0,47	0,55	0,53	0,52
6	0,58	0,65	0,62	0,62	6	0,54	0,68	0,63	0,62
7	0,61	0,62	0,66	0,63	7	0,57	0,60	0,65	0,61
8	0,63	0,65	0,62	0,63	8	0,59	0,70	0,65	0,65
9	0,60	0,58	0,63	0,60	9	0,58	0,63	0,59	0,60
10	0,51	0,50	0,54	0,52	10	0,50	0,57	0,49	0,52
11	0,46	0,45	0,48	0,46	11	0,41	0,54	0,45	0,47
12	0,29	0,37	0,37	0,34	12	0,34	0,39	0,34	0,36
Média	0,45	0,48	0,51	0,48	Média	0,45	0,53	0,49	0,49

Mosaicos				Padrões			
Item	série a	série b	média	Item	série a	série b	média
1	0,21	0,22	0,22	1	0,38	0,51	0,45
2	0,25	0,29	0,27	2	0,41	0,45	0,43
3	0,50	0,51	0,51	3	0,56	0,56	0,56
4	0,53	0,55	0,54	4	0,61	0,60	0,61
5	0,62	0,61	0,62	5	0,67	0,68	0,68
6	0,64	0,66	0,65	6	0,72	0,71	0,72
7	0,66	0,70	0,68	7	0,72	0,73	0,73
8	0,69	0,70	0,70	8	0,59	0,70	0,65
9	0,61	0,64	0,63	9	0,62	0,66	0,64
10	0,54	0,58	0,56	10	0,53	0,60	0,57
11	0,38	0,53	0,46	11	0,51	0,55	0,53
12	0,38	0,46	0,42	12	0,40	0,46	0,43
13	0,21	0,34	0,28	13	0,32	0,38	0,35
Média	0,48	0,52	0,50	Média	0,54	0,58	0,56

Análise dos itens segundo a TRI

Antes de apresentar as estimativas dos parâmetros da TRI, é necessário avaliar qual modelo se adequou melhor aos dados empíricos. Embreston e Reise (2000) afirmam que existem vários testes estatísticos que podem ser utilizados para avaliar o ajuste aos dados. Neste estudo, foi seguida a orientação de De Ayala (2009), que afirma que a decisão pelo modelo precisa ser baseada no cálculo da diferença do ajuste ($-2 \log \text{likelihood}$) dos dois modelos, dividido pela diferença dos graus de liberdade dos dois modelos. Para ser

significativo, esse valor, que é chamado de razão crítica, deve ser maior que 1,96. Outra estimativa apontada por De Ayala (2009) é o *Akaike Information Criterion* (AIC). No caso desse índice, o melhor modelo é aquele que apresenta o menor AIC.

As análises com a TRI foram realizadas usando o programa *BILOG-MG 3.0*. A Tabela 4.3 apresenta os valores de ajuste para a escala total do SON-R 6-40 de 124 itens. A análise com a TRI foi executada com todos os 124 itens juntos para garantir a comparabilidade dos valores dos parâmetros estimados entre os diferentes subtestes. O modelo de 2 parâmetros foi escolhido por apresentar melhor ajuste.

Tabela 4.3

Estatísticas de ajuste dos modelos de TRI do conjunto de 124 itens dos quatro subtestes do SON-R 6-40

Modelo	Escala total (124 itens)				
	-2 lnL	Δ ajuste	Δ df	r.c.	AIC
1 PL	88.673,73	---	---	---	88.921,73
2 PL	87.056,67	1.617,06	124	13,04	87.552,67
3 PL	87.559,55	-502,88	124	-4,54	88.303,55

Notas: -2 lnL = -2 log *likelihood* (ajuste); Δ ajuste = diferença de -2 log *likelihood* entre modelos; Δ df = diferença de graus de liberdade; r.c. = razão crítica (Δ ajuste / Δ df); AIC = *Akaike Information Criterion*.

4.3 Parâmetro *b*

Para avaliar a dificuldade do item ou parâmetro *b* segundo a TRI, utilizou-se o programa *BILOG-MG 3.0*. Esse parâmetro é medido na mesma escala da habilidade e indica que quanto maior for o valor do parâmetro *b*, maior será a habilidade requerida (Andrade, Laros & Gouveia, 2010). A Tabela 4.4 apresenta os valores do parâmetro *b*, que foi estimado utilizando escala *z*, isto é, média zero e desvio padrão um. A avaliação da Tabela 4.4 mostra que o subteste Analogias é o subteste mais fácil, com média de parâmetro *b* de -0,26 e o subteste Padrões é o subteste mais difícil, com a média de parâmetro *b* de -0,08. Os resultados também mostram que, com exceção do subteste Categorias, a série *a* é a mais difícil que a série *b* e série *c*. Na construção dos subtestes do SON-R 6-40 a intenção foi distribuir os itens de um subteste de tal maneira que a série

a fosse mais difícil. Isso foi feito em relação ao procedimento adaptativo usado na aplicação do SON-R 6-40 (Tellegen & Laros, 2014).

Tabela 4.4

Parâmetro *b* dos itens do SON-R 6-40 segundo a TRI (*N* = 1.360)

Analogias					Categorias				
Item	série a	série b	série c	média	Item	série a	série b	série c	média
1	-3,37	-3,92	-2,73	-3,34	1	-4,82	-2,43	-2,61	-3,29
2	-2,59	-2,29	-2,22	-2,37	2	-2,17	-2,51*	-2,54	-2,41
3	-2,20	-1,87	-1,80	-1,96	3	-2,11	-1,84	-1,80	-1,92
4	-1,44	-1,44	-1,34	-1,41	4	-1,53	-1,33	-1,19	-1,35
5	-0,96	-0,81	-0,83	-0,87	5	-0,67	-0,85	-1,11	-0,88
6	0,38	0,02	-0,25	0,05	6	0,53	0,22	-0,11	0,21
7	0,45	0,27	0,08	0,27	7	0,62	0,59*	0,30	0,50
8	0,77	0,69	0,49	0,65	8	1,17	0,54	0,55	0,75
9	1,12	1,09	0,55	0,92	9	1,33	1,06	0,99	1,13
10	1,47	1,45	1,15	1,36	10	1,53	1,28	1,48	1,43
11	1,63	1,63	1,50	1,59	11	1,86	1,45	1,67	1,66
12	2,23	1,82	1,88	1,98	12	2,02	1,86	2,01	1,96
Média	-0,21	-0,28	-0,29	-0,26	Média	-0,19	-0,16	-0,20	-0,18

Mosaicos				Padrões			
Item	série a	série b	média	Item	série a	série b	média
1	-2,77	-2,61	-2,69	1	-2,19	-1,67	-1,93
2	-2,72	-2,47	-2,60	2	-2,06	-1,92*	-1,99
3	-1,46	-1,51	-1,49	3	-1,49	-1,49	-1,49
4	-1,40	-1,38	-1,39	4	-1,16	-1,37	-1,27
5	-0,77	-0,97	-0,87	5	-0,65	-0,80	-0,73
6	-0,11	-0,64	-0,38	6	-0,47	-0,63	-0,55
7	0,35	-0,27	0,04	7	-0,09	-0,29	-0,19
8	0,43	0,14	0,29	8	0,69	0,26	0,48
9	0,95	0,81	0,88	9	0,79	0,61	0,70
10	1,30	1,06	1,18	10	1,26	0,94	1,10
11	1,81	1,41	1,61	11	1,38	1,23	1,31
12	1,80	1,63	1,72	12	1,72	1,59	1,66
13	2,17	2,14	2,16	13	1,90	1,82	1,86
Média	-0,03	-0,20	-0,12	Média	-0,03	-0,13	-0,08

Nota. * item é mais difícil que o item seguinte.

4.4 Parâmetro a

Para estimar o parâmetro a , segundo a TRI, também foi utilizado o programa *BILOG-MG 3.0*. A discriminação do item na TRI, também é chamado de parâmetro a , é definida como o poder do item para diferenciar testandos que têm habilidades abaixo da localização do item (dificuldade do item) e testandos que têm habilidades acima da localização do item (Andrade, Laros, & Gouveia, 2010; Baker, 2001). O parâmetro a é representado pela inclinação da curva característica do item no ponto de inflexão, onde a probabilidade de resposta correta é de 50% (Andrade, Laros, & Gouveia, 2010; De Ayala, 2009). Itens com curvas características dos itens mais inclinadas são mais úteis para diferenciar alunos com habilidades diferentes do que itens com curvas mais achatadas.

Geralmente, os valores do parâmetro a variam entre 0,0 e 2,0, sendo que valores negativos não são esperados, pois indicariam que a probabilidade de responder o item corretamente diminui com o aumento da habilidade. A Tabela 4.5 apresenta os valores do parâmetro a de todos os itens do teste.

Tabela 4.5*Parâmetro a dos itens do SON-R 6-40 segundo a TRI (N = 1.360)*

Analogias					Categorias				
Item	série a	série b	série c	média	Item	série a	série b	série c	média
1	2,03	1,14	1,55	1,57	1	0,44	1,46	1,21	1,04
2	0,99	1,31	1,41	1,24	2	1,66	0,73	1,31	1,23
3	0,93	1,16	1,46	1,18	3	1,12	0,86	1,24	1,07
4	0,98	1,27	1,72	1,32	4	0,90	0,99	0,79	0,89
5	0,99	1,10	1,41	1,17	5	0,63	0,83	1,19	0,88
6	1,06	1,25	1,03	1,11	6	0,83	1,28	1,10	1,07
7	1,11	1,04	1,38	1,18	7	1,07	0,94	1,24	1,08
8	1,31	1,45	1,21	1,32	8	1,33	1,42	1,25	1,33
9	1,66	1,47	1,40	1,51	9	1,51	1,60	1,25	1,45
10	1,73	1,73	1,43	1,63	10	1,74	1,79	1,40	1,64
11	1,95	1,99	1,67	1,87	11	1,75	1,73	1,55	1,68
12	1,54	2,29	1,57	1,80	12	1,94	1,84	1,59	1,79
Média	1,36	1,43	1,44	1,41	Média	1,24	1,29	1,26	1,26
Mosaicos				Padrões					
Item	série a	série b	Média	Item	série a	série b	média		
1	2,17	3,14	2,66	1	1,24	2,12	1,68		
2	1,57	1,74	1,66	2	1,96	2,00	1,98		
3	1,27	2,09	1,68	3	1,78	1,94	1,86		
4	1,59	1,76	1,68	4	1,77	2,22	2,00		
5	1,52	1,57	1,55	5	1,40	1,65	1,53		
6	1,18	1,64	1,41	6	1,94	1,99	1,97		
7	1,53	1,87	1,70	7	1,91	2,01	1,96		
8	1,70	1,85	1,78	8	1,29	1,85	1,57		
9	1,73	1,74	1,74	9	2,06	1,96	2,01		
10	1,91	1,43	1,67	10	1,95	1,90	1,93		
11	1,92	1,79	1,86	11	2,21	2,17	2,19		
12	2,14	1,83	1,99	12	2,59	2,02	2,31		
13	2,72	1,49	2,11	13	2,71	2,05	2,38		
Média	1,77	1,84	1,80	Média	1,91	1,99	1,95		

A Tabela 4.5 indica que os dois subtestes de execução, Padrões e Mosaicos, têm os itens que discriminam melhor: o valor médio do parâmetro a é 1,95 e 1,80, respectivamente. A média do parâmetro a para os subtestes Analogias e Categorias é de 1,41 e 1,26. Usando os critérios de Baker (2001), podemos concluir que, em geral, os itens de Padrões e Mosaicos têm um poder discriminativo muito alto ($> 1,70$), enquanto que o poder discriminativo dos itens de Analogias e Categorias, em geral, fica entre moderado e alto (entre 0,65 e 1,69).

4.5 Escores brutos

Foi calculada a média do escore bruto em cada faixa etária de cada subteste para examinar a diferença nos escores brutos com o aumento da idade. Como na faixa de idade entre 6 e 40 anos ocorrem mudanças desenvolvimentais, há a expectativa que o escore aumente (Hogan, 2006). O autor afirma que a curva de crescimento do escore bruto em testes de inteligência é bastante inclinada durante aproximadamente os primeiros 12 anos de vida. Entre 12 e 20 anos, a inclinação da curva de crescimento é muito menos inclinada. Entre 20 e 60 anos, as médias do escore bruto nos testes de inteligência permanecem praticamente as mesmas, contudo com um ligeiro declínio. A partir de 60 a 65 anos, o declínio torna-se mais acentuado. A Tabela 4.6 apresenta os valores nas 17 faixas etárias e nos quatro subtestes.

Tabela 4.6*Estatísticas descritivas do escore bruto por subteste em cada faixa etária (N = 1.360)*

Analogias					Mosaicos				
Idade	Média	(DP)	Assimetria	Curtose	Idade	Média	(DP)	Assimetria	Curtose
6 anos	10,1	(4,3)	0,10	-0,13	6 anos	6,0	(3,1)	0,56	-0,36
7 anos	12,4	(4,8)	-0,02	-0,56	7 anos	8,0	(3,3)	-0,38	-0,50
8 anos	13,1	(5,0)	0,04	-0,38	8 anos	8,7	(3,5)	-0,41	0,05
9 anos	15,3	(4,5)	-0,19	0,28	9 anos	10,2	(3,8)	0,49	0,98
10 anos	15,7	(4,9)	-0,05	-0,41	10 anos	10,9	(4,2)	-0,17	-0,54
11 anos	16,7	(4,6)	-0,13	-0,10	11 anos	12,1	(3,4)	0,01	-0,39
12 anos	16,7	(4,3)	0,19	0,08	12 anos	12,5	(4,0)	0,11	-0,08
13 anos	19,4	(5,6)	0,35	0,71	13 anos	14,3	(3,7)	0,69	0,50
14 anos	19,0	(6,2)	0,01	-0,19	14 anos	14,6	(4,4)	-0,06	-0,19
15 anos	19,8	(4,9)	-0,29	0,15	15 anos	14,7	(3,9)	0,09	-0,27
16 anos	20,8	(6,3)	-0,03	-0,41	16 anos	15,3	(4,1)	0,08	0,49
18 anos	22,3	(6,4)	-0,39	-0,24	18 anos	16,1	(4,5)	0,16	-0,59
20 anos	22,2	(6,1)	-0,15	-0,08	20 anos	16,6	(4,4)	0,30	-0,37
22 anos	23,3	(5,6)	-0,05	-0,37	22 anos	16,5	(4,1)	-0,46	0,14
27 anos	23,9	(6,5)	-0,15	-0,72	27 anos	16,8	(4,3)	-0,45	-0,34
32 anos	23,4	(7,1)	-0,25	-0,72	32 anos	15,8	(4,5)	-0,26	0,00
37 anos	21,1	(6,9)	0,02	-0,84	37 anos	15,3	(4,1)	0,13	-0,15

Categorias					Padrões				
Idade	Média	(DP)	Assimetria	Curtose	Idade	Média	(DP)	Assimetria	Curtose
6 anos	9,7	(4,0)	-0,22	-0,51	6 anos	4,1	(3,4)	0,67	-0,21
7 anos	12,3	(4,6)	-0,29	0,96	7 anos	7,0	(4,0)	0,19	-0,11
8 anos	12,8	(4,9)	-0,35	0,14	8 anos	8,8	(4,1)	-0,21	0,00
9 anos	14,5	(4,4)	-0,21	0,21	9 anos	10,2	(4,3)	0,37	1,56
10 anos	15,4	(5,0)	0,23	0,58	10 anos	11,0	(4,8)	-0,36	-0,53
11 anos	15,8	(5,5)	0,11	0,54	11 anos	12,3	(2,9)	-0,32	-0,58
12 anos	16,1	(5,5)	0,21	1,82	12 anos	12,9	(4,3)	-0,23	0,95
13 anos	17,7	(5,2)	-0,02	1,90	13 anos	14,9	(5,0)	-0,11	0,40
14 anos	18,9	(6,5)	0,21	-0,38	14 anos	15,2	(4,9)	-0,10	-0,01
15 anos	19,0	(6,0)	0,35	-0,54	15 anos	15,2	(4,0)	-0,07	-0,09
16 anos	18,8	(6,1)	-0,04	0,87	16 anos	15,8	(4,7)	-0,29	0,59
18 anos	21,1	(6,6)	-0,20	-0,39	18 anos	16,5	(4,5)	0,12	-0,21
20 anos	20,7	(6,0)	0,17	-0,78	20 anos	16,8	(4,7)	0,11	-0,41
22 anos	22,0	(6,6)	-0,36	-0,56	22 anos	17,3	(4,3)	0,05	0,21
27 anos	22,4	(6,2)	-0,07	-1,03	27 anos	16,7	(4,7)	-0,51	0,92
32 anos	21,8	(7,5)	-0,05	-0,89	32 anos	16,6	(5,2)	-0,61	1,04
37 anos	20,9	(7,3)	0,09	-0,62	37 anos	15,6	(4,1)	-0,03	0,28

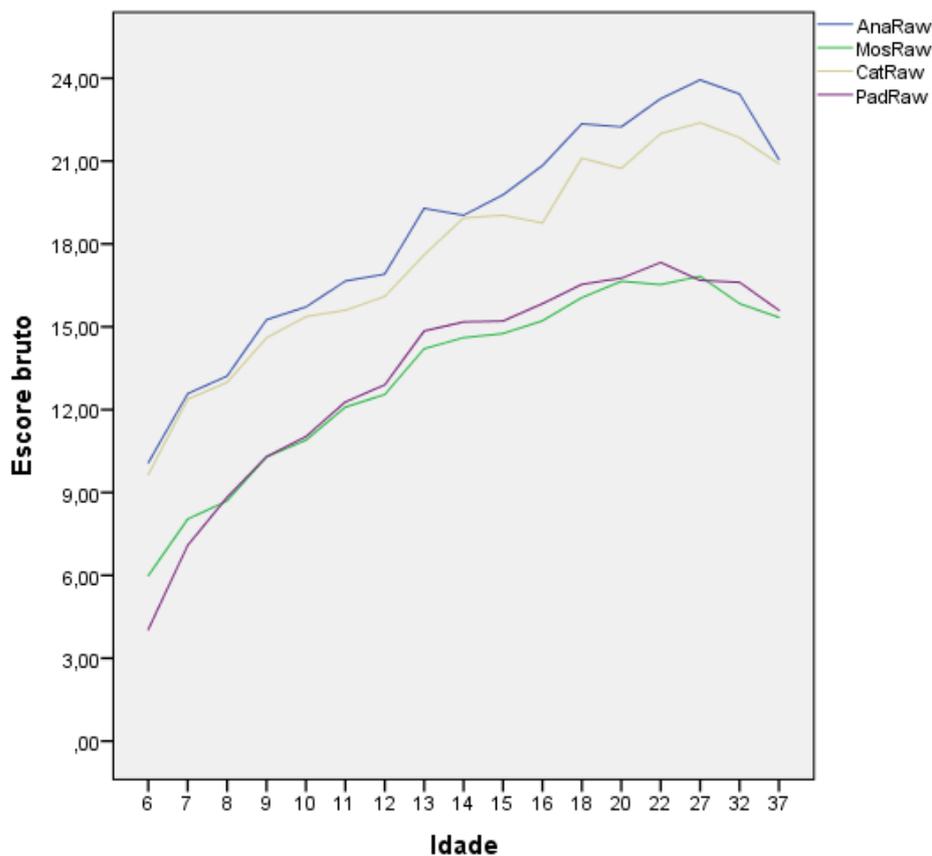


Figura 4.1 Escore médio bruto por subteste em cada faixa etária

A Tabela 4.6 mostra que as médias dos escores brutos dos subtestes aumentam com a idade até uns 22 anos para Mosaicos e Padrões e até 27 anos para os subtestes Analogias e Categorias. É importante ressaltar que não somente o escore bruto dos subtestes aumenta com a idade, mas também a variabilidades destes escores.

Ainda em relação aos escores brutos, a Tabela 4.7 apresenta os valores da média do escore total em cada faixa etária. O valor apresentado nesta tabela é calculado a partir da soma dos escores brutos dos quatro subtestes. Para tornar a contribuição de cada um dos subtestes de forma equivalente no escore bruto total, o escore bruto de cada subteste foi dividido pelos respectivos desvios padrões no grupo total (N = 1.360).

Assim como nas Tabelas 4.6 e 4.7, as Figuras 4.1 e 4.2 mostram que há um aumento progressivo com o aumento da idade e que o crescimento é curvilíneo, como indicado por pesquisas prévias que afirmam que o desenvolvimento cognitivo não se faz de

maneira contínua, podendo haver picos ou estagnações (Wechsler & Schelini, 2006; Schrank & Flanagan, 2003). O mesmo padrão curvilíneo foi encontrado no estudo de normatização do teste SON-R 6-40 na Alemanha e Holanda (Tellegen & Laros, 2011). Observa-se na Tabela 4.7 que entre as faixas de 6 a 13 anos, o aumento no escore total foi de 5,97 pontos, enquanto que entre as faixas etárias de 14 a 22 anos o aumento caiu para 1,80 pontos. Na faixa de 27 a 37 anos, parece que o escore total está diminuindo. A fim de verificar se o escore total está realmente diminuindo depois de 27 anos, foi realizada uma comparação com o teste *t* da diferença do escore total entre o grupo de 14 a 22 anos e o grupo de 27 a 37 anos. O teste *t* mostrou que não existe uma diferença significativa no nível de 1% entre os grupos. Assim, podemos concluir que a aparente queda no escore total do grupo acima de 27 anos não é estatisticamente significativa.

Tabela 4.7

Características do escore total bruto em cada faixa etária (N = 1.360)

Idade	Média da idade	Média do escore	(DP)	Assimetria	Curtose
6	6,55	4,8	(1,9)	0,35	-0,36
7	7,50	6,4	(2,3)	-0,48	-0,17
8	8,50	7,0	(2,4)	-0,45	0,38
9	9,49	8,1	(2,3)	0,02	1,13
10	10,53	8,5	(2,6)	-0,35	-0,71
11	11,48	9,2	(2,2)	-0,50	0,14
12	12,53	9,4	(2,5)	-0,03	0,68
13	13,50	10,7	(2,7)	0,18	1,19
14	14,49	11,0	(3,2)	-0,17	-0,15
15	15,50	11,1	(2,5)	-0,01	-0,31
16	16,46	11,5	(2,9)	-0,38	0,40
18	18,45	12,3	(3,1)	-0,04	-0,79
20	20,46	12,4	(2,8)	0,11	-0,41
22	22,46	12,8	(3,0)	-0,23	0,01
27	27,45	12,9	(3,0)	-0,34	-0,41
32	32,47	12,5	(3,5)	-0,33	-0,09
37	37,45	11,8	(3,0)	-0,04	-0,67

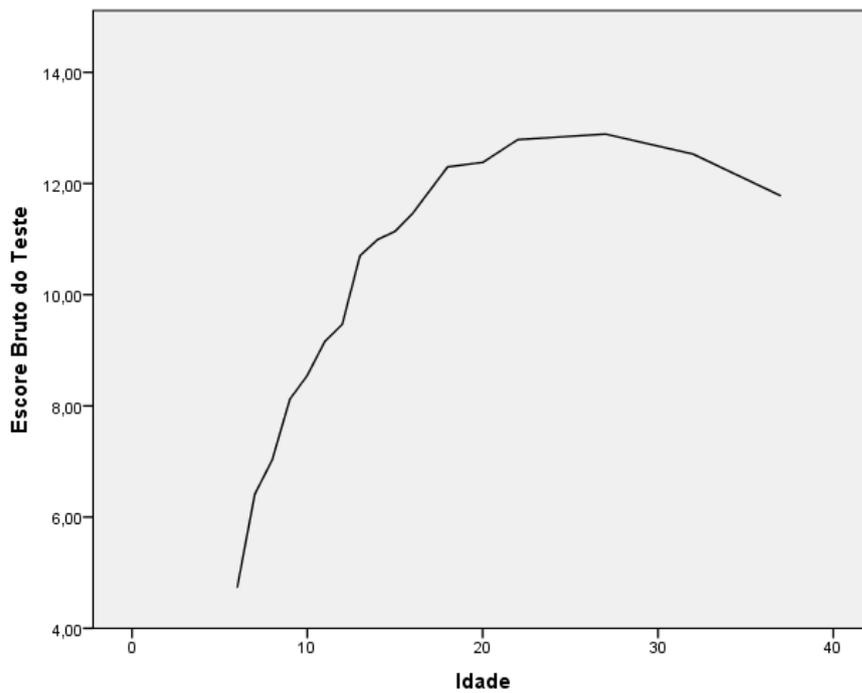


Figura 4.2 Escore médio bruto por faixa etária

A Tabela 4.8 apresenta a correlação múltipla entre os escores brutos com a idade e a correlação linear para todo o grupo normativo. A correlação múltipla entre a soma dos escores brutos dos quatro subtestes e a idade é de 0,65. Isso significa que 43% da variância do escore bruto total está sendo explicada pela variável idade. A maior parte da variância do escore bruto total está sendo explicada na faixa etária de 6 a 11 anos. Na faixa etária seguinte, de 12 a 18 anos, a relação com a variável idade já não é mais tão pronunciada. Por fim, no grupo dos adultos, de 20 a 40 anos, a relação entre idade e escore bruto não tem mais significância prática.

Tabela 4.8*Correlações dos escores brutos com a idade*

	6-40 anos	6-11 anos	12-18 anos	20-40 anos
N	1.360	480	480	400
Média de idade	16,78	9,01	15,2	28,1
(DP idade)	(8,58)	(1,70)	(1,95)	(6,29)
	<i>R</i>	<i>r</i>	<i>r</i>	<i>r</i>
Analogias	0,58	0,41	0,28	-0,06
Mosaicos	0,62	0,48	0,24	-0,11
Categorias	0,53	0,38	0,23	-0,01
Padrões	0,63	0,55	0,22	-0,10
Soma 4 subtestes	0,65	0,53	0,28	-0,08

R = correlação múltipla com idade e idade elevada ao quadrado; *r* = correlação linear com idade

4.6 Escores normatizados

As normas nacionais do teste SON-R 6-40 foram estabelecidas em cooperação com o prof. Peter Tellegen da Universidade de Groningen, utilizando o modelo de normatização contínua. Nesse modelo, os escores normatizados são calculados com base na idade exata e as distribuições dos escores para todos os grupos de idade são ajustadas simultaneamente.

Após a finalização do processo de normatização, descrito capítulo 3, foram analisadas as estatísticas descritivas dos escores normatizados dos subtestes e do SON-QI. Os escores dos subtestes são apresentados em uma escala Wechsler com média 10 e desvio padrão 3 e o SON-QI em uma escala com média 100 e desvio padrão 15. A Tabela 4.9 apresenta as características dos escores normatizados dos subtestes e do SON-QI para o grupo total e para os três grupos de idade: de 6 a 11 anos, de 12 a 16 anos e de 18 a 40 anos. A Tabela 4.9 mostra que, tanto no grupo total como nos três grupos de idade especificados, a média e o desvio-padrão ficam muito perto do valor das escalas escolhidas.

Tabela 4.9

Características dos escores normalizados dos subtestes e escore total (SON-QI) na amostra total e nos três grupos de idade.

Grupo Total						
6 a 37 anos (N=1.360)						
Subteste	Média	(DP)	Mínimo	Máximo	Assimetria	Curtose
Analogias	10,0	(3,0)	1	19	-0,05	-0,23
Mosaicos	10,0	(3,0)	2	19	-0,01	-0,15
Categorias	10,0	(3,0)	1	19	-0,05	-0,07
Padrões	10,0	(3,0)	0	22	-0,07	0,30
SON-QI	100,0	(15,0)	53	147	-0,17	-0,08
Grupo A						
6 a 11 anos (N=480)						
Subteste	Média	(DP)	Mínimo	Máximo	Assimetria	Curtose
Analogias	10,0	(2,9)	3	19	-0,02	-0,25
Mosaicos	10,0	(3,0)	2	19	-0,08	-0,21
Categorias	10,0	(3,0)	2	19	-0,19	0,24
Padrões	10,0	(2,9)	2	22	0,09	0,17
SON-QI	100,1	(14,6)	59	147	-0,25	-0,09
Grupo B						
12 a 16 anos (N=400)						
Subteste	Média	(DP)	Mínimo	Máximo	Assimetria	Curtose
Analogias	9,9	(3,1)	1	19	0,06	0,12
Mosaicos	9,9	(3,0)	2	18	0,12	0,07
Categorias	9,9	(3,0)	1	19	0,16	0,50
Padrões	10,1	(3,1)	0	18	-0,16	0,44
SON-QI	99,8	(15,3)	53	145	-0,09	0,31
Grupo C						
18 a 37 anos (N=480)						
Subteste	Média	(DP)	Mínimo	Máximo	Assimetria	Curtose
Analogias	10,0	(3,0)	1	16	-0,17	-0,51
Mosaicos	10,1	(3,0)	2	18	-0,05	-0,24
Categorias	10,0	(3,1)	1	16	-0,07	-0,71
Padrões	9,9	(3,0)	0	17	-0,17	0,30
SON-QI	100,0	(15,3)	56	134	-0,15	-0,37

A Tabela 4.10, a seguir, apresenta de forma mais específica e detalhada as características do SON-QI nos 17 grupos etários. A média do escore varia entre 97,3 e 102,2 e o desvio padrão do escore de QI varia entre 12,7 e 17,4. Levando em consideração o resultado da análise de variância, as diferenças entre os escores não são significativas

($p=0,878$). Observa-se que o valor mínimo do SON-QI foi 53, na faixa etária dos 14 e 16 anos e o valor máximo foi 147, na faixa etária dos 9 anos. No estudo de normatização na Holanda/Alemanha, o escore normatizado variou entre 55 e 145.

Tabela 4.10

Características do SON-QI nas 17 faixas etárias

Idade	SON-QI					
	Média	(DP)	Min	Máx	Assimetria	Curtose
6	97,3	(13,5)	74	134	0,28	-0,54
7	101,9	(15,4)	63	131	-0,44	-0,20
8	99,9	(15,3)	59	137	-0,43	0,43
9	101,8	(14,3)	64	147	0,00	0,93
10	99,8	(15,9)	65	130	-0,38	-0,71
11	100,1	(12,7)	66	128	-0,49	0,04
12	97,7	(14,5)	61	136	-0,05	0,70
13	102,2	(14,9)	59	145	0,18	1,08
14	100,8	(17,4)	53	138	-0,18	-0,17
15	99,2	(13,7)	64	129	-0,03	-0,32
16	99,3	(15,7)	53	139	-0,38	0,32
18	100,5	(16,1)	68	134	-0,05	-0,80
20	98,9	(14,5)	65	132	0,12	-0,42
22	99,8	(15,1)	56	131	-0,24	0,06
27	100,2	(15,1)	63	126	-0,35	-0,35
32	100,7	(16,8)	58	133	-0,35	0,01
37	99,7	(14,5)	66	128	-0,04	-0,65

A Tabela 4.11 apresenta os escores normatizados mínimos e máximos alcançados em cada subteste e no SON-QI por faixa etária. Percebe-se que houve grande variação nos escores normatizados dos subtestes e do SON-QI.

Tabela 4.11

Valor mínimo e máximo dos escores normalizados dos subtestes e do SON-QI por faixa de idade

Idade	Mínimo escore normalizado					Máximo escore normalizado				
	Ana	Mos	Cat	Pad	SON-QI	Ana	Mos	Cat	Pad	SON-QI
6	4	4	3	6	74	17	17	16	17	134
7	5	3	2	5	63	19	15	19	19	131
8	3	2	2	4	59	17	16	17	18	137
9	4	3	4	3	64	17	19	17	22	147
10	4	2	4	2	65	16	18	19	17	130
11	3	4	3	5	66	17	16	17	14	128
12	4	2	2	1	61	16	17	19	18	136
13	2	5	1	0	59	19	18	18	18	145
14	1	2	2	1	53	18	17	17	17	138
15	2	3	5	4	64	16	17	17	16	129
16	2	2	1	0	53	17	17	17	17	139
18	1	3	1	3	68	16	16	16	16	134
20	1	4	5	2	65	16	17	16	16	132
22	3	2	1	2	56	16	15	16	15	131
27	3	3	4	1	63	15	15	15	16	126
32	3	2	3	0	58	16	17	16	16	133
37	4	2	4	2	66	16	18	16	17	128

4.7 Fidedignidade

Os coeficientes de fidedignidade dos escores brutos dos quatro subtestes e da escala geral foram estimados para cada faixa etária utilizando o coeficiente Alfa de Cronbach. Foi escolhido este coeficiente em vez do Lambda 2 de Guttman porque alguns estudos na Holanda com o SON-R 5½-17 e SON-R 6-40 mostraram que o Lambda 2 superestimou em 0,015 a estimativa da fidedignidade. Os autores afirmam que essa superestimação é uma consequência do *feedback* oferecido ao examinando durante a aplicação dos subtestes que contem duas ou três séries para possibilitar uma aplicação adaptativa (Tellegen & Laros, 2014).

A Tabela 4.12 mostra, além dos coeficientes de fidedignidade, o desvio padrão dos escores brutos. Essa última informação é apresentada para demonstrar que existem

diferenças substanciais na variância dos escores brutos dos participantes nas 17 faixas etárias (observe os valores do desvio-padrão), o que influencia diretamente no tamanho do coeficiente de fidedignidade (Nunnally & Bernstein, 1994).

Tabela 4.12

Fidedignidade dos escores brutos dos quatro subtestes e SON-QI (N = 1.360)

Idade	N	Fidedignidade (Alfa de Cronbach)									
		Ana	DP	Mos	DP	Cat	DP	Pad	DP	SON-QI	DP
6 anos	80	0,85	(4,3)	0,86	(3,1)	0,84	(4,0)	0,87	(3,4)	0,94	(1,9)
7 anos	80	0,89	(4,8)	0,85	(3,3)	0,87	(4,6)	0,89	(4,0)	0,96	(2,3)
8 anos	80	0,90	(5,0)	0,87	(3,5)	0,89	(4,9)	0,89	(4,1)	0,96	(2,4)
9 anos	80	0,86	(4,5)	0,87	(3,8)	0,86	(4,4)	0,89	(4,3)	0,95	(2,3)
10 anos	80	0,88	(4,9)	0,90	(4,2)	0,88	(5,0)	0,91	(4,8)	0,96	(2,6)
11 anos	80	0,87	(4,6)	0,85	(3,4)	0,90	(5,5)	0,81	(2,9)	0,95	(2,2)
12 anos	80	0,86	(4,3)	0,88	(4,0)	0,90	(5,5)	0,89	(4,3)	0,96	(2,5)
13 anos	80	0,90	(5,6)	0,87	(3,7)	0,88	(5,2)	0,91	(5,0)	0,96	(2,7)
14 anos	80	0,92	(6,2)	0,90	(4,4)	0,92	(6,5)	0,91	(4,9)	0,97	(3,2)
15 anos	80	0,87	(4,9)	0,88	(3,9)	0,91	(6,0)	0,88	(4,0)	0,96	(2,5)
16 anos	80	0,92	(6,3)	0,88	(4,1)	0,91	(6,1)	0,91	(4,7)	0,97	(2,9)
18 anos	80	0,92	(6,4)	0,90	(4,5)	0,92	(6,6)	0,91	(4,5)	0,97	(3,1)
20 anos	80	0,91	(6,1)	0,90	(4,4)	0,91	(6,0)	0,91	(4,7)	0,96	(2,8)
22 anos	80	0,90	(5,6)	0,88	(4,1)	0,92	(6,6)	0,89	(4,3)	0,97	(3,0)
27 anos	80	0,92	(6,5)	0,90	(4,3)	0,91	(6,2)	0,91	(4,7)	0,97	(3,0)
32 anos	80	0,93	(7,1)	0,90	(4,5)	0,94	(7,5)	0,92	(5,2)	0,97	(3,5)
37 anos	80	0,93	(7,0)	0,88	(4,1)	0,94	(7,3)	0,89	(4,1)	0,97	(3,0)
Média	-	0,90		0,88		0,90		0,89		0,96	
Mín	-	0,85	(4,3)	0,85	(3,1)	0,84	(4,0)	0,81	(2,9)	0,94	(1,9)
Máx	-	0,93	(7,1)	0,90	(4,5)	0,94	(7,5)	0,92	(5,2)	0,97	(3,5)

Nota. Ana = Analogias; Mos = Mosaicos; Cat = Categorias; Pad = Padrões; Mín = Mínimo; Máx = Máximo; DP = Desvio Padrão.

A inspeção da Tabela 4.12 revela que, principalmente na faixa etária de 15 anos, a variância dos escores brutos é menor em comparação com as outras faixas etárias. É mais provável que a irregularidade dos desvios-padrão dos escores brutos dos subtestes nas 17

faixas etárias seja uma consequência da instabilidade nos casos constituintes das amostras do que um fenômeno real na população. Para obter estimativas mais estáveis, os coeficientes de fidedignidade foram ajustados para cada subteste usando regressão múltipla como uma função da idade exata dos testandos. Assim, foram obtidas estimativas dos coeficientes de fidedignidade que têm menos instabilidade de uma faixa etária para outra. A Tabela 4.13 apresenta os coeficientes de fidedignidade dos escores normatizados dos subtestes e da escala geral (SON-QI). A Tabela 4.13 também mostra o coeficiente de generalizabilidade do SON-QI.

Tabela 4.13

Fidedignidade dos escores normatizados por subteste e SON-QI (N = 1.360)

Idade	Fidedignidade (Alfa de Cronbach)					
	Analogias	Mosaicos	Categorias	Padrões	SON-QI	Gen. SON-QI
6 anos	0,87	0,86	0,85	0,89	0,955	0,874
7 anos	0,87	0,86	0,86	0,89	0,957	0,874
8 anos	0,88	0,87	0,87	0,89	0,958	0,874
9 anos	0,88	0,87	0,88	0,89	0,959	0,874
10 anos	0,88	0,87	0,89	0,89	0,960	0,874
11 anos	0,88	0,88	0,89	0,89	0,961	0,874
12 anos	0,89	0,88	0,90	0,89	0,962	0,874
13 anos	0,89	0,88	0,90	0,89	0,962	0,874
14 anos	0,89	0,88	0,90	0,89	0,963	0,874
15 anos	0,89	0,87	0,91	0,89	0,964	0,874
16 anos	0,89	0,89	0,91	0,89	0,964	0,874
18 anos	0,90	0,89	0,91	0,89	0,965	0,874
20 anos	0,90	0,89	0,91	0,89	0,966	0,874
22 anos	0,91	0,89	0,91	0,89	0,966	0,874
27 anos	0,92	0,89	0,91	0,89	0,967	0,874
32 anos	0,93	0,88	0,92	0,89	0,968	0,874
37 anos	0,94	0,87	0,95	0,89	0,970	0,874
Média	0,90	0,88	0,90	0,89	0,963	-
Mínimo	0,87	0,86	0,85	0,89	0,955	-
Máximo	0,94	0,89	0,95	0,89	0,970	-

No caso do subteste Padrões, a estimativa da fidedignidade é fixa em todos os grupos de idade porque foi observado que a variável idade não influenciava na estimativa da fidedignidade nas diferentes faixas etárias, ou seja, não houve relação clara entre as correlações entre a fidedignidade de Padrões e a idade exata.

Outra informação que também pode ser observada na Tabela 4.13 é o coeficiente de generalizabilidade do SON-QI. O coeficiente de generalizabilidade do SON-QI pode ser interpretado como a correlação esperada entre o SON-QI e o escore de qualquer outro teste não-verbal de inteligência com o mesmo número e tipo de subtestes. O coeficiente de generalizabilidade também indica o grau em que, como base nos subtestes administrados, é possível generalizar para o domínio total de subtestes comparáveis. Para calcular o coeficiente de generalizabilidade, os quatro subtestes do SON-R 6-40 são considerados como uma amostra do domínio de subtestes não-verbais similares. A generalizabilidade foi calculada utilizando o coeficiente Alfa de Cronbach, tomando os escores dos subtestes como a unidade de análise em vez dos escores nos itens. Quando no cálculo do coeficiente de Alfa de Cronbach o escore total nos itens são tomados como unidade de análise, obtemos uma estimativa de fidedignidade do escore total.

4.7.1 Estudo de fidedignidade teste-reteste

Foi realizado um estudo de fidedignidade teste-reteste para verificar o grau de estabilidade do escore. Neste estudo, participaram 77 pessoas com idade variando entre 9 e 21 anos, a média de idade foi de 15,36 anos ($DP = 2,75$), 67,9 % eram do sexo feminino e 64,1 % do total de participantes estava matriculada em escolas públicas do Distrito Federal. A coleta dos dados ocorreu em escolas públicas, creches, abrigos e residências, após assinatura do Termo de Consentimento Livre e Esclarecido pelos responsáveis. As aplicações do teste ocorreram em duas sessões e foram administrados

de forma individual. O intervalo entre as duas sessões variou entre 27 a 50 dias ($M=33,8$; $DP = 4,8$).

Quanto às análises dos dados, primeiramente, foram calculados os escores normatizados dos quatro subtestes e da escala total. Em seguida, foram realizadas comparações das médias em cada aplicação por meio do teste t de Student, considerando a significância de $p \leq 0,05$. Para calcular o tamanho do efeito das diferenças de médias, foi utilizada a fórmula de d de Cohen = $média_{II} - média_{I} / \sqrt{(dp_{II}^2 + dp_I^2)/2}$ (Vachhaase & Thompson, 2004). Os resultados das análises estão descritos na Tabela 4.14.

Tabela 4.14

Médias, desvios-padrão e diferença entre as médias dos escores normatizados no estudo de fidedignidade teste-reteste (N = 77)

Subteste	Aplicação I		Aplicação II		Diferença entre as médias	t	Valor- p	d
	média	(DP)	média	(DP)				
Analogias	10,4	(2,7)	11,3	(2,8)	0,87	-4,04	$\leq 0,01$	0,32
Mosaicos	10,2	(2,8)	10,6	(2,9)	0,40	-1,96	$\leq 0,05$	0,14
Categorias	10,1	(2,7)	11,2	(2,8)	1,10	-5,41	$\leq 0,01$	0,40
Padrões	10,2	(2,9)	10,7	(2,9)	0,54	-3,19	$\leq 0,01$	0,19
SON-QI	101,1	(12,5)	105,6	(13,7)	4,52	-6,59	$\leq 0,01$	0,34

Como pode ser observado na Tabela 4.14, houve uma tendência dos participantes apresentarem maiores médias na segunda aplicação, sendo as diferenças entre as médias significantes em todos os subtestes e escala total. Além disso, os valores do tamanho do efeito para os subtestes Analogias, Categorias e SON-QI indicam um efeito fraco. Já para os demais subtestes, o efeito foi muito fraco. Segundo Cohen (1988), os valores para interpretação dos tamanhos de efeito devem ser interpretados da seguinte forma:

- entre $0,0 < d \leq 0,20$ = efeito muito fraco;

- entre $0,20 < d \leq 0,50$ = efeito fraco;

- entre $0,50 < d \leq 0,80$ = efeito moderado;

- $d \geq 0,80$ = efeito forte.

Quanto aos valores das correlações entre as médias, os resultados obtidos indicaram relações fortes variando de 0,76 a 0,87 entre os escores dos quatro subtestes. A correlação entre os escores totais do SON-R 6-40 nas duas aplicações é de 0,90, o que indica uma fidedignidade teste-reteste muito satisfatória.

4.8 Dimensionalidade

Para avaliar a estrutura fatorial do SON-R 6-40, foram efetuadas análises fatoriais com os escores normatizados dos subtestes (M=10; DP=3). Como primeiro passo, foi avaliada a dimensionalidade do SON-R 6-40, ou com outras palavras, foi avaliada a questão da quantidade de fatores que precisam ser extraídos em uma análise fatorial. Para avaliar a dimensionalidade, foi utilizado o programa *Factor* 10.8.02 (Lorenzo-Seva, Timmerman & Kiers, 2006). Os seguintes métodos foram utilizados para avaliar a dimensionalidade do SON-R 6-40: (1) método Hull (Lorenzo-Seva, Timmerman & Kiers, 2011); (2) Análise Paralela *Optimal Implementation* (Timmerman & Lorenzo-Seva, 2011); (3) Análise Paralela *Classical Implementation* (Horn, 1965) e (4) Schwarz's *Bayesian Information Criterion Dimensionality Test – BIC* (Schwarz, 1978). Foram observados os seguintes índices de unidimensionalidade de cada método: *Unidimensional Congruence* (UniCo), *Explained Common Variance* (ECV) e *Mean of Item Residual Absolute Loadings* (MIREAL) (Lorenzo-Seva, Timmerman & Kiers, 2011, Damásio, 2012; Rodriguez, Reise & Haviland, 2015). Quanto aos índices de unidimensionalidade, um valor maior que 0,95 do coeficiente UniCo, maior que 0,85 do ECV e menor que 0,30 do MIREAL sugerem unidimensionalidade dos dados.

A Tabela 4.15 apresenta a quantidade de fatores indicados em cada análise e os valores dos índices de unidimensionalidade. Observa-se que todos os métodos indicaram a presença de um único fator e todos os índices também sugeriram a unidimensionalidade.

Tabela 4.15
Resultado das análises da dimensionalidade do SON-R 6-40

Método	Quantidade de fatores indicados
Hull	1
Índices de unidimensionalidade	
Unico	0,983
ECV	0,871
MIREAL	0,265
Método	Quantidade de fatores indicados
Análise paralela (<i>Optimal Implementation</i>)	1
Índices de unidimensionalidade	
Unico	0,993
ECV	0,905
MIREAL	0,225
Método	Quantidade de fatores indicados
Análise paralela (<i>Classical Implementation</i>)	1
Índices de unidimensionalidade	
UniCo	0,990
ECV	0,905
MIREAL	0,223
Método	Quantidade de fatores indicados
BIC	1
Índices de unidimensionalidade	
UniCo	0,983
ECV	0,872
MIREAL	0,263

4.9 Análise Fatorial Exploratória (AFE)

Fundamentando-se nas informações obtidas na análise da dimensionalidade do teste SON-R 6-40, foi realizada uma análise fatorial exploratória com o método *Minimum Rank Factor Analysis* (MRFA) (Ten Berge & Kiers, 1991; Lorenzo-Seva & Ferrando, 2006), utilizando o *software* Factor 10.8.02. A análise fatorial foi realizada com os escores normatizados dos subtestes (M=10 e DP=3). Os resultados da análise fatorial com MRFA podem ser observados na Tabela 4.16.

Tabela 4.16
Resultado da análise fatorial usando MRFA

<i>Eigenvalue e percentual de variância explicada por componente e grupo de idade</i>								
	Total (N=1.360)		Grupo A (N=480)		Grupo B (N=400)		Grupo C (N=480)	
Componente 1	2,91	72,8%	2,79	69,9%	2,95	73,8%	2,99	74,8%
Componente 2	0,49	12,3%	0,55	13,7%	0,48	12,1%	0,47	11,8%
Componente 3	0,32	8,0%	0,37	9,3%	0,31	7,9%	0,27	6,8%
Componente 4	0,27	6,8%	0,29	7,2%	0,25	6,2%	0,26	6,5%
<i>Carga fatorial média por subteste e grupo de idade</i>								
	Total (N=1.360)		Grupo A (N=480)		Grupo B (N=400)		Grupo C (N=480)	
Analogias	0,82		0,76		0,82		0,81	
Mosaicos	0,82		0,82		0,87		0,83	
Categorias	0,74		0,74		0,72		0,79	
Padrões	0,83		0,81		0,83		0,83	

Os resultados da análise fatorial no grupo total mostram que com apenas um fator já 72,8% da variância comum é explicada. Nos três grupos de idade, A (de 6 a 11 anos), B (12 a 16 anos) e C (18 a 37 anos), a percentagem da variância explicada pelo primeiro fator é, respectivamente, 69,9%, 73,8% e 74,8%. Os outros fatores não acrescentam uma quantidade de variância substancial. Segundo Baglin (2014), a percentagem de variância comum explicada pode ser considerada uma medida de ajuste do modelo aos dados. A estrutura fatorial com um fator dominante foi similar nos três grupos de idade.

Em relação às análises da estrutura interna do SON-R 6-40, os resultados evidenciam a unidimensionalidade da escala, indicando que as tarefas avaliadas pelo teste podem ser explicadas por uma dimensão geral. Esse resultado permite a predição de um escore geral (SON-QI) e é coerente com o estudo realizado na Holanda, no qual também se encontrou evidência de uma dimensão única (Tellegen & Laros, 2014).

4.10 Análise Fatorial Confirmatória (CFA)

A estrutura fatorial encontrada na análise fatorial exploratória foi submetida à CFA. Os parâmetros foram estimados a partir da matriz de covariância, utilizando-se o método da *Maximum Likelihood* (MLE), que permite estimativas mais precisas por ser sensível às violações da normalidade (Ullman, 2007; Schumacker & Beyerlein, 2000). Para avaliar a adequação do modelo, foram utilizados os seguintes índices de ajuste: qui-quadrado (X^2); *Normed Fit Index* (NFI $\geq 0,95$); *Incremental Fit Index* (IFI $\geq 0,95$); *Comparative Fit Index* (CFI $\geq 0,95$); *Standardized Root Mean Square Residual* (SRMR $\leq 0,08$) (Schreiber et al., 2006; Marsh, Hau, & Wen, 2004; Jackson, Gillaspay, & Purc-Stephenson, 2009). Os índices comparativos, NFI, IFI e CFI contrastam a matriz de dados amostrais com um modelo nulo ou de independência. O índice SRMR é categorizado como um índice de ajuste baseado em resíduos, que avalia a distância entre os elementos preditos por meio da estimação dos parâmetros do modelo e os dados observados na matriz de covariância (Ullman, 2007; Pilati & Laros, 2007). O pacote estatístico *AMOS* 21.0 foi utilizado para efetuar a CFA.

A Tabela 4.17 apresenta os índices de ajuste e as informações de cargas fatoriais padronizadas do modelo investigado. O modelo unifatorial testado se mostrou ajustado, apresentando valores nos índices de ajuste superiores a 0,95, segundo os critérios de pontos de corte adotado. Além disso, todas as cargas fatoriais são altas. Os valores das cargas fatoriais estão em consonância com os resultados da *Minimum Rank Factor Analysis*, indicando cargas levemente maiores para os subtestes de execução Mosaicos e Padrões (0,84 e 0,83, respectivamente) em comparação com os valores das cargas fatoriais para os subtestes Analogias e Categorias (0,81 vs. 0,72).

Segundo Schreiber et al. (2006), os índices de ajuste do modelo testado devem ser avaliados em conjunto e se a grande maioria dos índices indicarem um bom ajuste, então

provavelmente o modelo é adequado. Quanto à correlação entre os subtestes, o menor valor foi entre Mosaicos e Categorias ($r = 0,58$) e a maior correlação foi entre os subtestes de execução Mosaicos e Padrões ($r = 0,72$).

Tabela 4.17

Índices de ajuste pelo método de estimação da máxima verossimilhança e cargas fatoriais padronizadas (N=1.360)

Índices de ajuste	
Índices	Valor obtido
χ^2	83.739
Número de parâmetros	8
Graus de liberdade	2
<i>p-value</i>	< 0,01
NFI	0,97
IFI	0,97
CFI	0,97
SRMR	0,03
Subtestes	Carga fatorial padronizada
Analogias	0,81
Mosaicos	0,84
Categorias	0,72
Padrões	0,83

A Figura 4.3 apresenta o modelo estrutural com as respectivas estimativas padronizadas. Assim como na AFE, a CFA também sugere uma estrutura unidimensional.

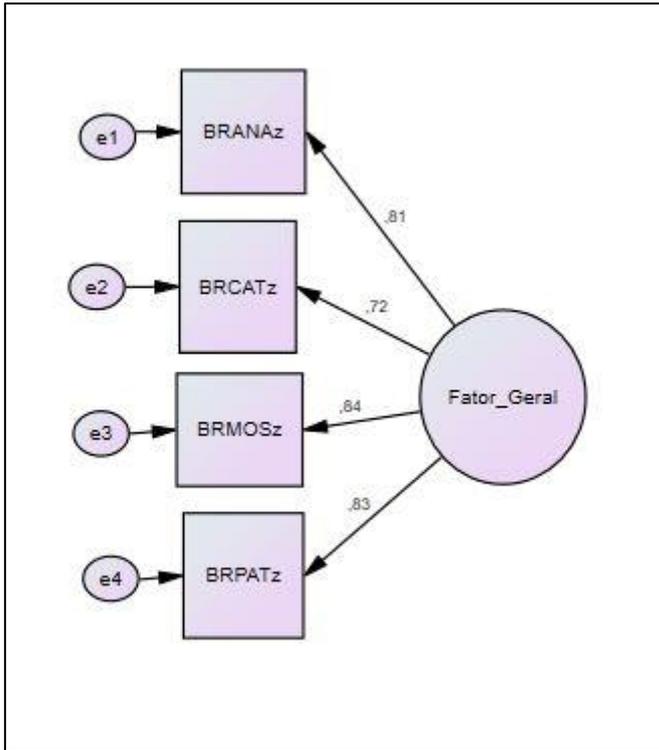


Figura 4.3 Modelo estrutural com as estimativas padronizadas ($N=1.360$)

5. Evidências de validade do SON-R 6-40

No contexto internacional e nacional, existem padrões muito bem estabelecidos para guiar o processo de construção ou adaptação de testes psicológicos. Uma publicação conjunta da *American Educational Research Association* (AERA), *American Psychological Association* (APA) e do *National Council on Measurement in Education* (NCME) são os *Standards for Educational and Psychological Testing*, daqui para frente tratados apenas como *Standards*, é amplamente reconhecida como uma declaração autorizativa proveniente de um consenso profissional em relação aos padrões para a testagem (AERA; APA; NMCE, 2014). Os *Standards* são “uma força global para a testagem” e desempenham um papel pedagógico importante na comunidade internacional, pois abordam conceitos tão fundamentais (validade, fidedignidade, normas, equalização, etc.) para a construção de testes e avaliação, que podem ser facilmente empregados em diferentes contextos (Zumbo, 2014). Os *Standards* gozam desse *status* tanto pela forma como foram desenvolvidos e aprovados, contando com a chancela das associações mais importantes da área de psicologia e educação americanas, quanto pela ampla história que possuem (Linn, 2006).

Atualmente, validade pode ser definida como o grau com que as evidências empíricas e teóricas apoiam as interpretações dos escores dos testes para um propósito específico (AERA, APA, NCME, 2014). O modelo dos *Standards* indica cinco fontes de evidências de validade que devem orientar os estudos sobre essa questão. São evidências baseadas: no conteúdo do teste, no processo de resposta, na estrutura interna, na relação com outras variáveis e evidências baseadas nas consequências da testagem (AERA, APA, & NCME, 2014).

O objetivo deste capítulo é relatar a relação dos escores normatizados do teste SON-R 6-40 com outras variáveis obtidas na pesquisa de normatização brasileira, por exemplo,

grau de escolaridade dos pais, duração da testagem, diferenças regionais, e apresentar os estudos de validação realizados no Brasil. Todas essas informações são importantes para os usuários dos testes julgarem se o teste é adequado para uma situação específica.

É importante destacar que a literatura da área aponta para a importância de mensurar outras variáveis, além do QI, quando se investiga capacidade intelectual (Nisbett et al., 2012; Flores-Mendoza & Nascimento, 2007). O conhecimento dessas outras variáveis são importantes para reconhecimento de déficits e, principalmente, para criação ou promoção de ambientes onde ocorra a estimulação da inteligência.

5.1 Relação com variáveis demográficas

Diferenças regionais

Participantes de diferentes municípios de todas as regiões brasileiras foram incluídos na amostra normativa, totalizando 39 municípios de 13 estados brasileiros. A fim de analisar a influência da variável região e localidade (capital e interior) sobre o desempenho no teste, foi observado o valor do QI médio dos respondentes destes grupos.

A Tabela 5.1 apresenta estes resultados.

Tabela 5.1

Estatísticas descritivas do SON-QI médio dos respondentes por região e localidade

Região	N	M	DP	EP	IC 95%	
					Inf.	Sup.
Norte	102	92,9	(13,9)	1,4	90,2	95,7
Nordeste	393	95,5	(15,9)	0,8	93,9	97,1
Centro-Oeste	124	102,2	(12,4)	1,1	99,9	104,4
Sul	185	102,9	(14,4)	1,1	100,8	105,0
Sudeste	556	103,0	(14,2)	0,6	101,8	104,2
Localidade	N	M	DP	EP	IC 95 %	
Interior	957	99,9	(14,9)	0,5	98,9	100,8
Capital	403	100,3	(15,5)	0,8	98,8	101,8

Notas. M = Média; DP = desvio padrão; EP= erro padrão da média; Inf = inferior; Sup = superior

De forma geral, uma comparação entre as médias das regiões revelou que as diferenças entre as médias foi significativa ($p < 0,05$). Mas ao observar os valores dos intervalos de confiança, nota-se que os valores de algumas regiões se sobrepõem e que a diferença entre as médias das regiões Norte e Nordeste não foi significativa. Além disso, a diferença entre as médias das regiões Centro-Oeste, Sul e Sudeste também não apresentou diferenças. Entretanto, há diferenças significativas quando as médias das regiões Norte e Nordeste são comparadas com as demais regiões brasileiras. É necessário destacar que essas diferenças entre as médias pode ser resultado de outros fatores, como NSE, escolaridade e erro amostral. Quanto à variável localidade, a diferença da média não é significativa ($p > 0,05$) e os intervalos de confiança apresentam sobreposição.

Sexo

A fim de verificar a influência do gênero no desempenho do teste, foi observado o valor do QI normatizado e a média obtida em cada subteste. Foi realizado também uma comparação entre as médias, teste t , com o objetivo de verificar se as diferenças apresentavam significância estatística entre homens e mulheres. Para calcular o tamanho de efeito das diferenças, foi utilizado a seguinte fórmula: $d \text{ de Cohen} = \text{média II} - \text{média I} / \sqrt{(dp_{II}^2 + dp_I^2)/2}$ (Vacha-Haase & Thompson, 2004). A Tabela 5.2 apresenta estes resultados.

Tabela 5.2
Médias nos subtestes e no SON-QI por sexo

Escore	Masculino		Feminino		Diferença entre as médias
	Média	Desvio Padrão	Média	Desvio Padrão	
Analogias	10,2	(3,0)	9,8	(3,0)	-0,4
Mosaicos	10,4	(3,1)	9,6	(2,9)	-0,8
Categorias	10,2	(3,1)	9,8	(3,0)	-0,3
Padrões	10,3	(3,0)	9,8	(3,0)	-0,5
QI-SON	101,5	(15,2)	98,5	(14,7)	-2,9

Os dados da Tabela 5.2 mostram que o desempenho do grupo masculino foi superior ao grupo feminino em todos os subtestes e no escore total do teste SON-R 6-40, porém a diferença entre as médias é muito pequena. Vale ressaltar que o teste *t* indicou que todas as diferenças entre as médias do grupo não são significativas e os tamanhos de efeito para tais diferença é considerado muito fraco, *d* entre 0,04 e 0,09.

Nível de escolarização

O questionário contextual apresentava itens acerca do grau de escolaridade dos pais ou responsáveis. No caso das crianças, foi solicitado que o questionário fosse respondido por um dos responsáveis. A Tabela 5.3 apresenta dados sobre o nível de escolarização dos responsáveis pelos participantes da amostra de normatização.

Tabela 5.3

Grau de escolaridade dos pais ou responsáveis

Escolaridade do pai	Frequência	%
Não estudou em escolas formais	38	2,8%
Ensino fundamental incompleto	251	18,5%
Ensino fundamental completo	143	10,5%
Ensino médio incompleto	61	4,5%
Ensino médio completo	297	21,8%
Ensino superior incompleto	34	2,5%
Ensino superior completo	123	9,0%
Pós-graduação incompleta	2	0,1%
Pós-graduação completa	9	0,7%
Em branco	402	29,5%
Escolaridade da mãe	Frequência	%
Não estudou em escolas formais	32	2,4%
Ensino fundamental incompleto	244	17,9%
Ensino fundamental completo	157	11,5%
Ensino médio incompleto	63	4,6%
Ensino médio completo	313	23,0%
Ensino superior incompleto	59	4,3%
Ensino superior completo	146	10,7%
Pós-graduação completa	15	1,1%
Em branco	331	24,3%

Foi calculada a correlação entre o grau de escolaridade dos pais e o desempenho no SON-R 6-40. A correlação entre a escolaridade da mãe e o SON-QI foi de 0,32 ($N=1.029$) e a correlação entre a escolaridade do pai e o SON-QI foi de 0,19 ($N=979$).

Nível Socioeconômico (NSE)

O questionário contextual também apresentou itens acerca da presença de bens de consumo no domicílio. Quanto aos bens de consumo, os itens investigados foram: televisão, rádio, DVD, máquina de lavar, geladeira, computador, automóvel e banheiro.

Os itens sobre a quantidade de bens presentes na residência foram submetidos a uma análise dos componentes principais. Os resultados indicaram um KMO superior a 0,87 e 44,4% de variância explicada para uma estrutura com um fator. As cargas fatoriais podem ser observadas na Tabela 5.4

Tabela 5.4
Cargas fatoriais do questionário de nível socioeconômico (NSE)

Itens na residência	Carga fatorial
TV	0,78
Computador	0,76
Banheiro	0,73
Automóvel	0,71
DVD	0,64
Máquina de lavar	0,59
Geladeira	0,56
Rádio	0,51

Os valores das cargas fatoriais da Tabela 5.4 revelam que o número de TVs tem a maior carga fatorial de NSE, enquanto o número de rádios tem a menor carga. Aparentemente, a quantidade de TVs é um indicador mais forte do NSE dos pais ou responsáveis do que o número de rádios na residência.

Após a análise dos componentes principais foi calculado o escore no fator nível socioeconômico de cada participante, obtido por meio da média dos itens respondidos.

Esse procedimento foi efetuado a fim de possibilitar a verificação do perfil socioeconômico da amostra. A Figura 5.1 mostra a distribuição dos escores fatoriais do NSE da amostra de normatização. Uma vez que 184 participantes tinham dados faltosos nas questões sobre os bens na residência, a distribuição é baseada em 1.176 em vez de 1.360 respondentes. O escore fatorial do NSE variou de 0 a 4,0 e a média foi 1,13 (DP=0,51).

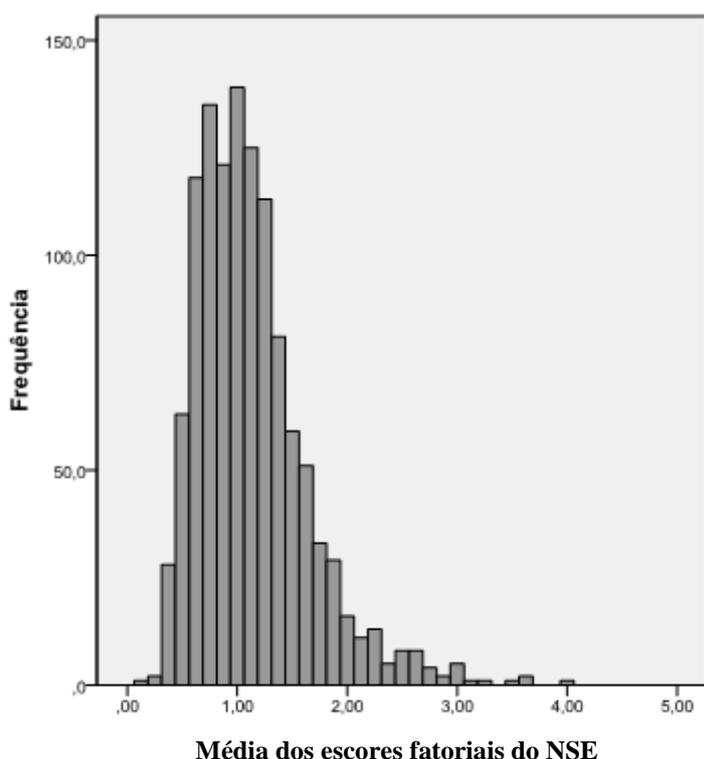


Figura 5.1 Distribuição da amostra segundo o nível socioeconômico

A correlação entre o escore fatorial do NSE e o escore total no SON-R 6-40 foi de 0,31. Esse valor significa que 9,6% da variância do SON-QI é explicada pelo escore fatorial do NSE, indicando que o desempenho no teste é relativamente pouco influenciado pelo nível socioeconômico da família. É importante ressaltar que a correlação entre o escore do NSE e o SON-QI do SON-R 2½-7[a] foi mais elevada, a saber, de 0,47 (Laros, Tellegen, Jesus & Karino, 2015). Foi observada uma correlação relativamente baixa de

0,24 entre o grau de escolaridade da mãe e escore fatorial do NSE. Quanto à escolaridade do pai, a correlação com o NSE foi de 0,22.

Avaliação dos aplicadores

A maioria dos respondentes (91%) da amostra normativa foi avaliada por 19 examinadores, que testaram de 32 a 112 participantes. A Tabela 5.5 apresenta a quantidade de testes aplicados por cada administrador, o QI médio dos respondentes que eles avaliaram e a diferença da média do QI. É importante observar que outras variáveis, como sexo, NSE e idade, podem ter influenciado o desempenho no teste.

Tabela 5.5
Estatísticas descritivas do SON-QI por examinador (N=1.237)*

Examinador	N	M	(DP)	Desvio
1	68	92,3	(16,3)	-7,7
2	102	92,9	(13,9)	-7,1
3	68	94,9	(19,4)	-5,1
4	68	95,8	(14,1)	-4,2
5	102	96,3	(15,9)	-3,7
6	91	96,9	(13,3)	-3,1
7	34	98,5	(12,6)	-1,5
8	41	98,6	(18,7)	-1,4
9	68	98,8	(13,3)	-1,2
10	43	99,7	(12,7)	-0,3
11	43	101,7	(12,3)	1,7
12	32	103,7	(11,9)	3,7
13	68	103,9	(13,8)	3,9
14	34	104,3	(12,0)	4,3
15	112	104,3	(13,0)	4,3
16	68	104,6	(10,7)	4,6
17	93	106,3	(12,6)	6,3
18	68	107,2	(15,3)	7,2
19	34	110,5	(14,6)	10,5
Total	1.237	100,6	(14,0)	

Nota. * Os examinadores que aplicaram o teste em menos de 30 pessoas não foram incluídos na tabela, resultando em um total de 1.237 casos. M = média; DP = desvio padrão; Desvio = desvio em relação à média geral de 100.

Durante a aplicação do teste, os examinadores avaliaram o comportamento dos respondentes levando em consideração quatro aspectos: motivação, concentração, cooperação e compreensão das instruções. Foi utilizada a seguinte escala de mensuração: 1=muito bom; 2=bom; 3=baixa; 4=muito baixa. A Tabela 5.6 apresenta os resultados da relação entre a avaliação feita pelo examinador e o desempenho no teste em cada condição.

Tabela 5.6
Avaliação do examinador e desempenho no teste

Aspecto avaliado	Avaliação	Grupo A 6 a 11 anos			Grupo B 12 a 16 anos			Grupo C 18 a 37 anos		
		%	M	(DP)	%	M	(DP)	%	M	(DP)
Motivação	Muito baixa	1,0	85,8	(18,5)	0,8	93,7	(7,6)	0,5	98,5	(7,8)
	Baixa	8,5	87,9	(14,9)	5,5	88,9	(15,2)	4,1	88,2	(17,0)
	Bom	43,5	98,4	(13,9)	42,3	96,6	(14,5)	36,5	96,7	(13,6)
	Muito bom	41,9	103,5	(13,4)	45,0	103,9	(15,1)	51,5	104,2	(15,3)
Concentração	Muito baixa	1,9	77,3	(12,3)	0,8	85,0	(20,8)	0,4	89,5	(20,5)
	Baixa	16,3	89,2	(14,2)	8,8	86,4	(13,6)	5,8	84,0	(14,8)
	Bom	41,3	100,2	(12,0)	35,8	94,2	(12,3)	40,4	97,6	(13,9)
	Muito bom	35,6	104,6	(14,1)	48,3	106,3	(14,6)	45,4	105,3	(14,4)
Cooperação	Muito baixa	0,8	88,8	(19,9)	0,3	97,0	--	0,2	104,0	--
	Baixa	3,1	84,5	(16,7)	2,0	90,0	(13,7)	1,9	87,4	(17,8)
	Bom	42,1	96,0	(14,4)	31,3	93,3	(14,3)	30,4	95,1	(14,3)
	Muito bom	49,0	103,7	(13,0)	60,0	103,3	(14,9)	59,6	103,7	(14,8)
Compreensão das instruções	Muito baixa	4,4	76,1	(12,8)	1,8	72,9	(14,1)	0,6	78,3	(24,2)
	Baixa	10,0	86,9	(8,7)	4,0	88,6	(12,7)	3,5	78,8	(12,3)
	Bom	40,8	99,7	(12,4)	34,8	93,0	(12,0)	35,2	95,0	(13,3)
	Muito bom	39,8	105,1	(13,5)	52,8	105,7	(14,5)	52,7	105,9	(13,8)

Nota. M = média; DP = desvio padrão.

O número de participantes avaliados em cada situação variou de 1 a 286. No aspecto cooperação, apenas uma pessoa no grupo B (12 a 16 anos) recebeu avaliação “muito baixa”, impossibilitando o cálculo do desvio padrão.

No primeiro aspecto avaliado, é possível observar que a motivação cresce com o aumento da idade. De forma geral, observou-se que quanto pior a avaliação nos aspectos, menor o escore no teste, principalmente no aspecto compreensão das instruções, onde em todos os grupos houve uma diminuição do escore quando o examinador atribuiu uma nota menor.

Tempo de administração do teste

Na pesquisa de normatização, o teste foi administrado em apenas uma sessão, sendo oferecidas pequenas pausas entre os subtestes quando solicitado pelo respondente. A Tabela 5.7 apresenta de forma mais detalhada o tempo necessário que os grupos precisaram para responder todo o teste e o tempo médio para responder cada subteste.

Tabela 5.7
Tempo de administração do teste completo e por subteste

Minutos	Tempo médio de administração do teste completo			Tempo de administração em minutos		
	Grupo A	Grupo B	Grupo C	Subteste	Média	DP
< 40 minutos	23,5%	10,5%	2,5%	Analogias	8,2	4,8
41-50 minutos	22,1%	17,0%	10,2%	Mosaicos	20,9	6,3
51-60 minutos	26,3%	23,5%	21,3%	Categorias	8,7	4,9
61-70 minutos	15,6%	18,8%	27,1%	Padrões	17,0	8,4
> 70 minutos	10,4%	27,5%	37,1%	Total	54,8	17,5
Em branco	2,1%	2,8%	1,9%			

Nota. DP = desvio padrão

Considerando o tempo total de testagem, verificou-se uma média geral de 60 minutos ($DP=17,9$). Para as crianças do grupo A (6 a 11 anos), a média foi de 52,5 minutos ($DP=15,1$), as do grupo B (12 a 16 anos) precisaram de 64,6 minutos ($DP=17,5$) e os adultos do grupo C (18 a 37 anos) tiveram tempo médio de 68,6 minutos ($DP=17,2$).

5.2 Correlação com outros testes de inteligência

Estudo com teste não-verbal de inteligência Leiter-R em amostra com TEA

Entre os sintomas presentes nos Transtornos do Espectro do Autismo (TEA) estão os prejuízos de linguagem e comunicação. Observa-se um atraso ou ausência de linguagem falada e quando presente, há dificuldades em iniciar e manter uma conversa, uso estereotipado e restrito como as ecolalias e idiossincrasias que prejudicam a compreensão e expressão adequada da linguagem (*American Psychiatric Association*, 2014). Testes verbais podem apresentar limitações para avaliar indivíduos com TEA por demandarem uso funcional da linguagem, atribuição de estados mentais, compreensão de sequências lógicas com intencionalidade de personagens e conhecimento cristalizado que depende de situações formais de aprendizagem. Neste sentido, a literatura na área reporta as vantagens do uso de instrumentos não-verbais (Antonio, Mecca & Macedo, 2012; Dawson, Soulières, Gernsbacher & Mottron, 2007).

Estudos apontam para o perfil de “picos e vales” em diferentes tarefas não-verbais, sendo as habilidades visuo-espaciais e de execução mais preservadas em relação às habilidades de raciocínio fluido (Macedo, Mecca, Valentini, Laros & Lima, 2013; Mecca, Orsati, Macedo, 2014). Apesar das diferenças significativas em tarefas com demandas cognitivas distintas, o desempenho nestas podem estar relacionados (Macedo, Mecca, Valentini, Laros & Lima, 2013).

O objetivo do presente estudo foi investigar se há correlação entre o desempenho de indivíduos com Transtornos do Espectro do Autismo (TEA) em dois testes não-verbais: SON-R 6-40 e Leiter-R. A Leiter-R é amplamente utilizado no contexto internacional para avaliar as habilidades cognitivas de pessoas com distúrbios no desenvolvimento (Myrelid et al., 2010; Caudle, Katzenstein, Oghalai, Lin & Caudle,

2014; Lane, 2011; Kushner, Bennetto & Yost, 2007) e está em fase de normatização para o Brasil.

Participaram deste estudo 21 indivíduos com TEA, entre 6 e 19 anos ($M=10,3$; $DP=4,3$), sendo 4 meninas (19,1%) e 17 meninos (80,9%). O diagnóstico foi realizado por uma equipe especializada, no ambulatório de TEA em um hospital na cidade de São Paulo.

A Leiter-R avalia inteligência de forma não-verbal. A partir dos 6 anos, são utilizadas as seguintes provas: Figura-Fundo, Analogias, Formas Completas, Dobra de Papel (avaliam processamento visual – Gv), Sequências, Padrões Repetidos (avaliam raciocínio fluido – Gf). As aplicações dos instrumentos foram realizadas em dois encontros, individualmente, junto com uma terapeuta já conhecida pelo participante. As correlações entre os escores do SON-R 6-40 e Leiter-R são apresentadas na Tabela 5.8.

Tabela 5.8

Coefficientes de correlação entre os subtestes do SON-R 6-40 e da Leiter-R e entre o QI de ambos os instrumentos.

Subtestes Leiter-R	Subtestes SON-R 6-40				
	Analogias	Mosaicos	Categorias	Padrões	SON-QI (IC 95%)
Figura fundo	0,69	0,76	0,84	0,82	0,86 (0,68-0,94)
Analogias	0,82	0,81	0,83	0,81	0,90 (0,77-0,96)
Formas completas	0,56	0,67	0,66	0,75	0,72 (0,42-0,88)
Sequências	0,71	0,89	0,78	0,83	0,87 (0,70-0,95)
Padrões	0,49	0,66	0,62	0,61	0,65 (0,30-0,85)
Dobra de papel	0,61	0,71	0,68	0,66	0,73 (0,44-0,88)
QI-Total Leiter-R	0,76	0,90	0,85	0,89	0,93 (0,83-0,97)

Nota. As correlações mais altas dos subtestes do SON-R 6-40 com os subtestes do Leiter-R estão destacadas em negrito.

Entre os subtestes de ambos os teste de inteligência foram observadas correlações variando entre moderadas e altas. Dos seis subtestes da Leiter-R, o subteste Analogias mostrou as correlações mais alta com o SON-QI do SON-R 6-40. Os subtestes Sequências e Figura-Fundo da Leiter-R também mostraram uma alta correlação com o SON-QI (0,87

e 0,86, respectivamente). Dos subtestes do SON-R 6-40, os subtestes de execução, Mosaicos e Padrões, apresentaram correlações mais altas com o QI-Total da Leiter-R (0,90 e 0,89, respectivamente). O subteste Analogias do SON-R 6-40 apresentou a correlação mais alta com o subteste Analogias da Leiter-R, e o subteste Categorias do SON-R 6-40 apresentou correlação mais alta com o subteste Figura Fundo da Leiter-R. Todos os subtestes do SON-R 6-40 apresentaram altas correlações com os subtestes que avaliam processamento visual (*Gv*) (Figura-Fundo, Analogias, Formas Completas e Dobra de Papel) e com os subtestes que avaliam inteligência fluida (*Gf*) (Sequências e Padrões Repetidos) da Leiter-R, corroborando achados prévios que mostram que apesar das tarefas demandarem algumas habilidades específicas, há uma parte significativa da variância no desempenho que é compartilhada. A correlação muito alta de 0,93 entre o escore total do SON-R 6-40 e o escore total da Leiter-R indica que ambos avaliam o mesmo construto. No mesmo tempo, a correlação alta entre os dois testes é uma evidência forte da validade convergente do SON-R 6-40.

Estudos anteriores de validação do teste SON-R 6-40

O teste SON-R 6-40 já possui diferentes estudos realizados no contexto brasileiro com testes normatizados para o Brasil (Lima & Laros, 2017; Alves, 2016; Laros, Almeida, Valentini & Lima, 2015). Nestas pesquisas, os testes administrados juntamente com o SON-R 6-40 foram: a BPR-5 ($N=68$); o WISC-IV ($N=120$); o TONI-3, a Escala de Maturidade Mental Colúmbia (EMMC), as Matrizes Progressivas Coloridas (MPC) de Raven ($N=150$) e o SON-R 2½-7[a] ($N=44$). A Tabela 5.9 apresenta um resumo dos resultados obtidos desses estudos.

Os outros testes de inteligência que foram aplicados juntamente com o SON-R 6-40 (WISC-IV, BPR-5, EMMC, MPC, SON-R 2½-7[a]) apresentaram uma correlação com o escore total do SON-R 6-40 entre 0,67 e 0,74 com um valor médio de 0,72. Uma

correlação 0,72 significa que o SON-R 6-40 tem mais do que 70% da variância em comum com estes outros cinco testes pesquisados. O valor médio de 0,72 fica abaixo do valor médio de 0,80 encontrado na pesquisa sobre a validade convergente realizada na Holanda e Alemanha (Tellegen & Laros, 2014). Uma possível explicação da diferença é que as correlações na pesquisa da Holanda e Alemanha foram corrigidas para a falta de variância o que não ocorreu na pesquisa no Brasil.

Tabela 5.9

Coefficientes de correlação entre o escore de QI do teste SON-R 6-40 e os escores de outros testes de inteligência normalizados no Brasil

BPR-5 (N = 68)					
	RA	RV	RE	RN	QI BPR-5
SON-QI 6-40	0,71 (0,68)	0,66 (0,63)	0,68 (0,64)	0,57 (0,54)	0,74 (0,71)
WISC-IV (N = 120)					
	ICV	IOP	IMO	IVP	QI-WISC
SON-QI 6-40	0,48 (0,45)	0,84 (0,71)	0,44 (0,39)	0,32 (0,26)	0,73 (0,63)
MPC, EMMC e TONI-3 (N = 150)					
	QI-TONI-3	QI-EMMC	QI-MPC		
SON-QI 6-40	0,63 (0,57)	0,67 (0,61)	0,72 (0,64)		
SON-R 2½-7[a] (N = 44)					
	SON-R EE	SON-R ER	SON-QI 2½-7		
SON-QI 6-40	0,62 (0,54)	0,72 (0,60)	0,75 (0,67)		

RA=Raciocínio Abstrato; RV=Raciocínio Verbal; RE=Raciocínio Espacial; RN=Raciocínio Numérico; ICV=Índice de Compreensão Verbal; IOP=Índice de Organização Perceptual; IMO=Índice de Memória Operacional; IVP=Índice de Velocidade de Processamento; SON-R EE=Escala de Execução do SON-R 2½-7[a]; SON-R ER=Escala de Raciocínio do SON-R 2½-7[a]. Os primeiros valores na tabela são os coeficientes de correlação corrigidos para atenuação e os valores entre parênteses são os coeficientes brutos.

As correlações relativamente baixas de 0,32 e 0,44 entre o QI-SON e os índices IVP (Índice de Velocidade de Processamento) e IMO (Índice de Memória Operacional) do WISC-IV são evidências da validade divergente do SON-R 6-40, uma vez que os testes SON não pretendem medir a velocidade de processamento e a memória operacional.

Os cinco estudos realizados entre o SON-R 6-40 e outros testes de inteligência fornecem evidências satisfatórias da validade convergente e da validade divergente do teste. A correlação entre os escores totais mais elevada encontrada foi com o Leiter-R, um teste não-verbal de inteligência que tem muito em comum com o SON-R 6-40.

5.3 Grupos especiais

Como o estudo de normatização não contemplou indivíduos com problemas auditivos, foi realizado um estudo piloto para verificar a adequação das instruções não-verbais de aplicação. Além disso, pessoas com transtornos no desenvolvimento responderam o teste SON-R 6-40 para verificar o grau de adequação do teste para avaliar grupos difíceis.

Estudo com pessoas com problemas auditivos

Participaram desse estudo 21 indivíduos com deficiência auditiva diagnosticada, com idade variando entre 7 e 22 anos ($M=14,7$; $DP=3,7$), sendo 14 do sexo masculino. Quanto ao tipo de deficiência auditiva e ao tipo de comunicação que os participantes usam rotineiramente, a Tabela 5.10 apresenta uma síntese dessas informações.

Tabela 5.10

Tipo de deficiência auditiva e tipo de comunicação utilizada pelos participantes (N=21)

Tipo de deficiência auditiva	N	Tipo de comunicação	N
Surdez profunda	12	Oral	8
Surdez severa	3	Libras	11
Surdez moderada	3	Leitura labial	2
Surdez leve	2		
Surdez profunda + implante coclear	1		
Total	21	Total	21

Todas as aplicações foram realizadas em sessões individuais, com tempo médio de aplicação de 40 minutos ($DP=9,8$), em uma instituição situada no interior de São Paulo

que oferece atendimento especial para pessoas surdas. Todos os participantes tiveram os termos de consentimento livre e esclarecido assinados pelos responsáveis.

Utilizando as normas brasileiras, foi calculado o QI médio dos participantes e realizada uma comparação entre os escores desse grupo com os escores da amostra de normatização. Além disso, foi utilizada a fórmula de *d de Cohen* = $\frac{\text{média II} - \text{média I}}{\sqrt{(\text{dp}_{II}^2 + \text{dp}_I^2)/2}}$ para calcular o tamanho do efeito das diferenças de médias (Vacha-Haase & Thompson, 2004). Os resultados são apresentados na Tabela 5.11.

Tabela 5.11

Diferenças entre os escores normatizados do grupo com deficiência auditiva e da amostra de normatização brasileira do SON-R 6-40 (N=21)

Subtestes	Média	(DP)	Média N	(DP) N	Dif	Sign.	Tamanho de efeito (<i>d</i> de Cohen)
Analogias	6,5	(3,1)	10,1	(3,0)	-3,6	< 0,01	1,14
Mosaicos	7,2	(3,5)	9,99	(3,0)	-2,8	< 0,05	0,91
Categorias	7,7	(3,8)	10,01	(3,0)	-2,3	< 0,05	0,67
Padrões	7,2	(3,5)	10,02	(3,0)	-2,8	< 0,05	0,86
SON-QI	83,1	(18,2)	100,0	(15,0)	-16,9	< 0,01	1,01

Notas. Média N= média dos escores da amostra de normatização; (DP) N = desvio-padrão dos escores da amostra de normatização; Dif. = diferença entre a média dos escores do grupo com deficiência auditiva e da amostra de normatização; Sign.= Significância estatística da diferença.

Os resultados obtidos indicam que o desempenho do grupo com deficiência auditiva é inferior ao desempenho considerado como médio para a população brasileira, conforme os dados apresentados na Tabela 5.11. Todas as diferenças entre os grupos com surdez e da amostra de normatização são significativas e o tamanho de efeito para tais diferenças é considerado um efeito moderado a forte, *d* entre 0,67 e 1,14. Os resultados da Tabela 5.11 também indicam uma variabilidade maior dos escores em comparação com a variabilidade da amostra de normatização.

Levando em consideração os valores baixos obtidos nos escores no SON-R 6-40 do grupo com deficiência auditiva, é provável que pessoas com múltiplas deficiências fizeram parte do grupo. A literatura da área indica que a deficiência auditiva vem

acompanhada frequentemente com outros tipos de deficiência (Caudle, Katzenstein, Oghalai, Lin & Caudle, 2014). Por exemplo, caso uma gestante seja infectada com o vírus citomegalovírus (CMV) ou o vírus da rubéola, é possível que o bebê sofra perda auditiva. Essas infecções durante a gravidez podem causar outros tipos de deficiência. Infelizmente, informações sobre o tempo da perda da audição, escores em outros testes de inteligência, a frequência e o tipo de intervenções terapêuticas não estão disponíveis para esta amostra. É conhecido que estes fatores podem desempenhar um papel importante no desenvolvimento das habilidades cognitivas.

Quanto a adequação das instruções não-verbais contidas no manual de aplicação, não houve relato de nenhuma dificuldade ou problema para entender a atividade proposta de cada subteste. Embora seja um grupo reduzido de respondentes, obtido a partir de uma amostra selecionada por conveniência, é possível concluir que todas as aplicações foram realizadas normalmente, sem irregularidades ou intercorrências. Desta forma, o teste SON-R 6-40 mostrou-se apropriado para uma caracterização cognitiva do grupo, bem como uma ferramenta adequada para compor uma avaliação mais ampla para elaboração de intervenção ou prognósticos.

Estudo com pessoas com Transtornos do Espectro do Autismo

Participaram do estudo 25 meninos e 12 meninas com idades entre 6 e 24 anos ($M=12,0$; $DP=4,5$), diagnosticados com TEA por uma equipe interdisciplinar, com base nos critérios do DSM-5 (*American Psychiatric Association*, 2014) (Mecca, Lima, Laros, Macedo & Lowenthal, no prelo). Foram excluídos indivíduos que apresentavam outras condições genéticas associadas aos TEA ou que não conseguiram responder ao SON-R 6-40 por limitações físicas, sensoriais ou alterações comportamentais que dificultaram uma avaliação padronizada. Foram incluídos tanto participantes com fala preservada ($N=30$) quanto indivíduos não-verbais ($N=7$). As avaliações com o SON-R 6-40 foram

realizadas de forma individualizada, em um encontro com o participante com duração em torno de 40 minutos.

Em relação à análise dos dados, inicialmente foram realizadas as transformações dos escores brutos em escores normalizados levando em consideração as normas do teste SON-R 6-40 para o contexto brasileiro. Em seguida, foram realizadas análises descritivas (média, desvio-padrão e amplitude) dos escores dos grupos com TEA e do grupo de normalização do teste SON-R 6-40. A Tabela 5.12 apresenta os resultados.

Tabela 5.12

Diferenças entre os escores normalizados do grupo com TEA e da amostra de normalização brasileira do SON-R 6-40

Subtestes	Grupo com TEA				Grupo de normalização		Diferença*
	Média	(Desvio-Padrão)	Mínimo	Máximo	Média	(Desvio-Padrão)	
Analogias	2,5	(1,6)	1	6	10,0	(3,0)	-7,5
Mosaicos	3,7	(2,4)	1	9	10,0	(3,0)	-6,3
Categorias	2,7	(1,7)	1	7	10,0	(3,0)	-7,3
Padrões	4,0	(2,3)	1	8	10,0	(3,0)	-6,1
SON-QI	58,5	(12,1)	36	81	100,0	(15,0)	-41,5

Notas. *Nesta coluna são apresentadas as diferenças entre as médias dos escores normalizados dos participantes com TEA e os escores normalizados do SON-R 6-40 para a população brasileira.

Quanto ao desempenho do grupo com TEA, os resultados indicam que os participantes apresentaram desempenho inferior em todos os subtestes, indicando que de modo geral, o grupo TEA apresenta habilidades cognitivas inferiores ao grupo com desenvolvimento típico. A literatura na área aponta que 45 a 60% dos casos de TEA ocorrem com deficiência intelectual (Carlsson et al., 2013; Joseph, 2011; Lai et al., 2013; Levy, Mandell & Schultz, 2009). A inspeção da Tabela 5.12 também permite inferir que há uma certa variabilidade de nível intelectual na presente amostra, com indivíduos que apresentaram inteligência na média.

Quanto a adequação do teste, os aplicadores (psicólogos treinados e habituados a trabalharem com esse grupo), consideraram algumas características do teste SON-R 6-40

como um instrumento atrativo para o uso com pessoas que são difíceis de avaliar. Tais características incluem, materiais atraentes, variedade de tarefas, oportunidade de desempenhar um papel ativo, exemplos detalhados, *feedback* do examinador, nível de entrada fácil e um procedimento adaptativo, que repercute tanto no tempo de aplicação dos subtestes quanto no engajamento do respondente para resolver as tarefas.

Considerações Finais

O presente trabalho teve como objetivo geral estabelecer normas para o teste não-verbal de inteligência SON-R 6-40. Além disso, paralelamente ao estudo de normatização, foram realizados vários estudos de validação dos escores do teste. Quanto à contribuição da pesquisa, destaca-se, principalmente, duas contribuições: o desenvolvimento de critérios para a construção da amostra de normatização e o método utilizado para obtenção das normas.

Quanto à amostra de normatização, buscou-se representar a diversidade da realidade socioeconômica brasileira ao aplicar o teste em 39 municípios de 13 estados das cinco regiões brasileiras. Destaca-se que houve um percentual igual de homens e mulheres na amostra, assim como a quantidade de respondentes em cada grupo de idade se manteve ao que foi estabelecido inicialmente ($N=80$). A média de idade dos participantes em cada grupo de idade também traz uma forte evidência que os aplicadores selecionaram os participantes com uma idade o mais próxima possível do que foi planejado; o desvio da média de idade não foi acima de 0,10 anos. Ademais, a amostra normativa apresenta mais respondentes residindo no interior ($N=957$) do que residindo nas capitais ($N=403$), se contrapondo aos estudos que em geral têm os centros urbanos ou cidades específicas como locais para a coleta dos dados.

Quanto ao processo de normatização, foi utilizado o método de normatização contínua, onde as distribuições dos escores brutos para todos os grupos de idade são ajustadas simultaneamente como uma função da idade usando regressão para cada subteste separadamente. Esse método apresenta algumas vantagens, tais como: (1) os escores normatizados podem ser calculados com base na idade exata das pessoas; (2) o uso da informação de todos os grupos de idade simultaneamente torna a normatização mais precisa, (3) as amostras de normatização dos subgrupos e a amostra total não

precisam ser tão grande como na normatização tradicional para obter resultados estáveis, (4) o processo de ajustamento usando a regressão evita as discontinuidades que frequentemente acontecem na normatização tradicional, e (5) o modelo permite fazer extrapolações fora da faixa de idade na qual a pesquisa de normatização foi feita. Será disponibilizado para os usuários do teste tanto um *software*, que permite o cálculo com base na idade exata do respondente, como também as tabelas de conversão de escores brutos dos subtestes em escores normatizados com intervalos de um mês para crianças de até oito anos de idade, intervalos de dois meses para crianças até 11 anos e com intervalos de três meses para as pessoas com mais de 11 anos de idade. A conversão da soma dos escores normatizados dos subtestes para o SON-QI é realizado com intervalos de três meses para crianças de até os oito anos de idade. Dos oito anos até 11 anos, a conversão do SON-QI é feita com intervalos de seis meses; para as demais idades, o intervalo é de 12 meses.

Outro avanço deste trabalho é que também os coeficientes de fidedignidade foram ajustados em função da idade dos respondentes. Assim, foi evitado que a fidedignidade em alguns grupos apresentasse grandes flutuações. Os valores dos coeficientes de fidedignidade dos subtestes variaram entre 0,85 e 0,95, e para o SON-QI, os coeficientes variaram entre 0,95 e 0,97.

O teste também foi aplicado em amostras clínicas específicas, pessoas com deficiência auditiva e com transtorno do espectro do autismo, para verificar a adequação das instruções não-verbais de aplicação contidas no manual e o grau de adequação do teste para avaliar grupos que são difíceis para testar. Em ambos os estudos, todas as aplicações foram realizadas em sessões individuais por profissionais altamente qualificados para trabalharem com as amostras selecionadas. No caso dos participantes com deficiência auditiva, não houve relato de nenhuma dificuldade ou problema para

entender a atividade proposta de cada subteste. Na avaliação das pessoas com TEA, os psicólogos treinados consideraram o teste SON-R 6-40 como um instrumento atrativo para o uso com essas pessoas que são difíceis de avaliar. Nas duas situações, o teste SON-R 6-40 mostrou-se apropriado para uma caracterização cognitiva do grupo, bem como uma ferramenta adequada para compor uma avaliação mais ampla para elaboração de intervenção ou prognósticos.

Além disso, o teste SON-R 6-40 foi aplicado juntamente com o Leiter-R, em 21 pessoas com TEA. O Leiter-R é um teste que avalia a inteligência de forma não-verbal, amplamente utilizado no contexto internacional para avaliar as habilidades cognitivas de pessoas com distúrbios no desenvolvimento. Entre os escores totais nos dois testes de inteligência foi observada uma correlação muito elevada ($r=0,93$). Todos os subtestes do SON-R 6-40 mostraram uma alta correlação com os subtestes que avaliam processamento visual (Gv) e com os subtestes que avaliam inteligência fluida (Gf) da Leiter-R. A correlação muito alta de 0,93 entre o escore total do SON-R 6-40 e o escore total da Leiter-R indica que ambos avaliam o mesmo construto. No mesmo tempo, a correlação alta é uma evidência forte da validade convergente do SON-R 6-40.

Frente aos resultados encontrados, a presente pesquisa disponibiliza um teste não-verbal de inteligência que poderá ser utilizado no contexto profissional após a avaliação do Conselho Federal de Psicologia, seguindo as recomendações da Resolução 009/2018. Espera-se que outros estudos com o teste SON-R 6-40 sejam realizados no contexto brasileiro para acumular evidências dos escores do teste em diferentes situações ou com diferentes públicos.

Como limitações e agenda de futuras pesquisas, destaca-se a necessidade de ampliar as amostras dos estudos com grupos especiais e, especialmente no caso dos participantes com deficiência auditiva, adotar critérios claros para que apenas as pessoas com

deficiência auditiva componham a amostra deste grupo. Estudos anteriores com os testes SON com pessoas que têm somente perda auditiva demonstram que o desempenho deste grupo é semelhante ao grupo de ouvintes (Snijders, Tellegen, & Laros, 1989).

Referências

- Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests and knowledge. *Intelligence*, 22, 227-257.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: APA.
- Aertsen, M. C. E. (2008). *De SON-I op het laagste niveau te moeilijk? [O SON-I no nível mais baixo é difícil demais?]*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Almeida, L. S. (1994). *Inteligência: Definição e medida*. Aveiro, PT: CIDInE.
- Alves, T. A. (2016). *Propriedades psicométricas dos escores do SON-R 6-40, Teste não-verbal de Inteligência*. Dissertação de Mestrado, Universidade de Brasília, Brasília.
- American Psychiatric Association (2014). *Manual diagnóstico e estatístico de transtornos mentais – DSM-5*. Artmed Editora.
- Andrade, J. M., Laros, J. A., & Gouveia, V. V. (2010). O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Avaliação Psicológica*, 9(3), 421-435.
- Antonio, A. G. M., Mecca, T. P., & Macedo, E. C. (2012). O uso do teste não-verbal Leiter-R na avaliação de inteligência em distúrbios no desenvolvimento. *Cadernos de Pós-Graduação em Distúrbios do Desenvolvimento*, 12(2), 9-15.
- Arragon, R. C., & Koster, K. J. M. (2008). *Puzzels te gemakkelijk? Onderzoek naar het discriminerend vermogen van de subtest Mozaïeken van de SON-I [Quebra-cabeças fácil demais? Estudo o sobre a capacidade discriminatória do subteste Mosaicos do SON-I]*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Assink, M., Boersma, N., Brenninkmeijer, J., Judistra, M., Renckens, L., Rooks, G., Verweij, P., Wennink, E., & Werkman, W. (1992). *Effect van de moeilijkheidsvolgorde en feedback op testprestaties [Efeito da sequência de dificuldade e feedback no desempenho do teste]*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation*, 19(5), 1-14.
- Baker, F. B. (2001). *The basics of item response theory*. United States of America: ERIC Clearinghouse on Assessment and Evaluation.
- Bijl, L. (2004). *Multi-cultureel testen: de herziening van de subtest Categorieën van de*

- SON-R 5½-17* [Pesquisa em várias culturas: A revisão do subteste Categorias do SON-R 5½-17]. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061-1071.
- Boven, G. (2007). *Onderzoek SON-I bij zwakke leerlingen* [Pesquisa com o SON-I com alunos fracos]. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Boven, G., & Rusthoven, H. (2005). *De invloed van cultuur op een niet-verbale redeneertest: Onderzoek met de subtest Categorieën van de SON-E in Nederland en Kenia* [A influência da cultura em um teste de raciocínio não verbal: Pesquisa com o subteste categorias do SON-E na Holanda e no Quênia]. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Bracken, B. A., & McCallum, R. S. (1998). *UNIT Universal Non-verbal Intelligence Test*. Itaska, IL: Riverside Publishing.
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Non-Verbal Intelligence, Fourth Edition (TONI-4)*. Austin, TX: Pro-Ed.
- Carlsson, L. H., Norrelgen, F., Kjellmer, L., Westerlund, J., Gillberg, C., & Fernell, E. (2013). Coexisting disorders and problems in preschool children with Autism Spectrum Disorders. *The Scientific World Journal*, 213979. doi: 10.1155/2013/213979.
- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports *g* and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5-21). New York: Pergamon Press.
- Caudle, S. E., Katzenstein, J. M., Oghalai, J. S., Lin, J., & Caudle, D. D. (2014). Nonverbal cognitive development in children with cochlear implants: Relationship between the Mullen Scales of Early Learning and Later Performance on the Leiter International Performance Scales-Revised. *Assessment*, *21*(1), 119-128.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, *38*, 592.
- Cattell, R. B. (1971). *Abilities; their structure, growth, and action*. Boston: Houghton Mifflin.
- Cliffordson, C., & Gustafsson, J. E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and

- length of schooling. *Intelligence*, 36, 143-52.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2014). *Testagem e avaliação psicológica. Introdução a testes e medidas*. Porto Alegre, AMGH Editora Ltda.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Damásio, B. F. (2012). Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica*, 11(2), 213-228.
- Dawson, M., Soulières, I., Gernsbacher, M. A., & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science*, 18(8), 657-662.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- Dijk, H. van, & Tellegen, P. J. (2004). *NIO Nederlandse Test voor Onderwijsniveau. Handleiding en Verantwoording [NIO – Teste holandês para o nível educacional. Manual e Relatório Científico]*. Amsterdam: Harcourt Test Publishers.
- Drunen, P. van (2003). *Tussen traditie en vernieuwing: zestig jaar SON-tests [Entre tradição e inovação: sessenta anos de testes do SON]*. Amsterdam: Boom Test Uitgevers.
- Elliott, C. D., Murray, D. J., & Pearson, L. S. (1982). *British ability scales: Manual*. Windsor: National Foundation for Educational Research.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Evers, A., Lucassen, W., Meijer, R. R., & Sijtsma, K. (2009). *COTAN beoordelingssysteem voor de kwaliteit van tests [COTAN assessment system for the quality of tests]*. Amsterdam: Nederlands Instituut van Psychologen.
- Flanagan, D. P., & Harrison, P. L. (2012). *Contemporary intellectual assessment. Theories, tests, and issues*. New York, NY: The Guilford Press.
- Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary Gf-Gc theory: Joint confirmatory factor analysis of the WJ-R and the KAIT in a non-white sample. *Journal of School Psychology*, 36(2), 151-182.
- Flores-Mendoza, C. E., & Nascimento, E. (2007). Condição cognitiva de crianças de zona rural. *Estudos em Psicologia-Campinas*, 24(1), 13-22.
- Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, 40, 155-171.

- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 2, 171-191.
- Flynn, J. R. (2009). *What is intelligence? Beyond the Flynn effect*. Cambridge, UK: Cambridge University Press.
- Geerlings, H., Laros, J. A., Tellegen, P. J., & Glas, C. A. (2011). Testing the difficulty theory of the SON-R 5½-17, a non-verbal test of intelligence. *British Journal of Mathematical and Statistical Psychology*, 67(2), 248-265. doi: 10.1111/bmsp.12017
- Grob, A., & Hagmann von Arx, P. (2017). *IDS-2. Intelligentie – en ontwikkelingschalen voor kinderen en jongeren. Handleiding [IDS-2. Escalas de inteligência e de desenvolvimento. Manual]*. Amsterdam: Hogrefe.
- Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179-203.
- Hammill, D. D., Pearson, N. A., & Voress, J. K. (1993). *DTVP-2 Developmental test of visual perception. Examiner's manual*. Austin, TX: Prod-Ed.
- Hammill, D. D., Pearson, N. A., & Wiederholt, J. D. (1996). *CTONI Comprehensive Test of Non-Verbal Intelligence. Examiner's Manual*. Austin, TX: Pro-Ed.
- Hennemann, M. (2009). *Ontwikkeling van de SON-I: onderzoek naar de moeilijkheidsopbouw van de makkelijke items van de subtests Patronen [Desenvolvimento do SON-I: pesquisa sobre a dificuldade de construção de itens fáceis do subteste Padrões]*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Hof, M., Eekhout, S., & Mengerink, E. L. (2006). *Ontwikkeling van de SON-I: Subtests Analogieën, Mozaïeken en Patronen [Desenvolvimento do SON-I: Subtestes Analogias, Mosaicos e Padrões]*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Hogan, T. P. (2006). *Introdução à prática de testes psicológicos*. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Horn, J. L. (1976). Human abilities: A review of research and theory in the early 1970s. *Annual Review of Psychology*, 27, 437-485.
- Horn, J. L. (1985). Remodeling old theories of intelligence: Gf-Gc theory. In B. B. Wolman (Eds.), *Handbook of intelligence* (pp. 267-300). New York: Wiley.

- Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (2nd ed.). New York: The Guilford Press.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 53-91). New York: Guilford.
- INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2016). *Censo escolar da educação básica 2016 – notas estatísticas*. Inep.
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*(1), 6-23.
- Jesus, G. R. (2009). *Normatização e validação do teste não-verbal de inteligência SON-R 2½-7[a] para o Brasil*. Tese de doutorado, Universidade de Brasília, Brasília.
- Joseph, R. M. (2011). The significance of IQ and differential cognitive abilities for understanding ASD. In D. A. Fein (Ed.), *The Neuropsychology of Autism*. (pp. 281-294). New York, NY: Oxford University Press.
- Judistira, E. M. (1996). *A preliminary validation research with the SON 5½-17 in China*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Kaufman, A. S. (2009). *IQ testing 101*. New York: Springer.
- Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47*, 635-650.
- Kittrell, C. L. (1999). *Continuous norming: Estimation of test statistics for normative purpose*. Tese de doutorado, Pasadena, California. doi: 10.13140/RG.2.1.1660.1120
- Kuipers, M. J. (2008). *Revisie van de SON-R 5½-17: Subtest Patronen*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Kuschner, E. S., Bennetto, L., & Yost, K. (2007). Patterns of nonverbal cognitive functioning in young children with autism spectrum disorders. *Journal of autism and development disorders, 37*(5), 795-807.
- Lai, M. C., Lombardo, M. V., Chakrabarti, B., & Baron-Cohen, S. (2013). Subgrouping the Autism “Spectrum”: Reflections on DSM-5. *PLoS Biology, 11*(4), e1001544. doi: 10.1371/journal.pbio.1001544.
- Lane, K. A. (2011). *Visual filtering in children with fetal alcohol spectrum disorder*. Tese de doutorado, McGill University, Montreal.

- 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.
- Laros, J. A., & Tellegen, P. J. (2001). *Cross-cultural research with the Snijders-Oomen non-verbal intelligence test*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report. [www.testresearch.nl].
- Laros, J. A., & Tellegen, P. J. (2004). Cultural bias in the SON-R test: Comparative study of Brazilian and Dutch children. *Psicologia: Teoria e Pesquisa*, 20(2), 103-111.
- Laros, J. A., Jesus, G. R., & Karino, C. A. (2013). Validação do teste não-verbal de inteligência SON-R 2½-7[a]. *Revista Avaliação Psicológica*, 12(2), 233-242.
- Laros, J. A., Almeida, G. O. N., Valentini, F., & Lima, R. M. F. (2015). Dimensionalidade e evidências de validade convergente do SON-R 6-40. *Temas em Psicologia*, 23(4), 929-945. doi: 10.9788/TP2015.4-10.
- Laros, J. A., Tellegen, P. J., Jesus, G. R., & Karino, C. A. (2015). *SON-R 2½-7[a] – Teste não-verbal de inteligência. Manual*. São Paulo: Hogrefe.
- Le Clercq, M., & Holvast, L. (1996). *The SON-R 5½-17 and the WISC-R applied to peruvian school children*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Lenhard, A., Lenhard, W., Suggate, S., & Segerer, R. (2018). A continuous solution to the norming problem. *Assessment*, 25(1), 112-125. doi: 10.1177/1073191116656437
- Letitre, S., Calkar, I., & Wemmenhove, E. (2007). *Onder de SON. Onderzoek naar de nieuwste versie van de SON, de SON-I [Pesquisa sobre a nova versão dos testes SON, o SON-I]*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Levy, S. E., Mandell, D. S., & Schultz, R. T. (2009). Autism. *Lancet*, 374(9701), 1627-1638. doi: 10.1016/S0140-6736(09)61376-3.
- Lima, R. M. F., & Laros, J. A. (2017). Evidências de validade convergente e discriminante dos escores do SON-R 6-40. *Psicologia: Teoria e Prática*, 19, 107-120.
- Linn, R. L. (2006). The standards for educational and psychological testing: Guidance in test development. In S. M. Downing, & T. M. Haladyna, (Eds.), *Handbook of test development* (pp. 27-38). New Jersey: Lawrence Erlbaum Associates.
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull Method for selecting the number of common factors. *Multivariate Behavioral Research*, 46(2), 340-364.
- Lorenzo-Seva, U. & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, 38(1), 88-91.

- Luteijn, F., & Barelds, D. P. F. (2004).). *Groninger Intelligentie Test-2: GIT2 [Teste de inteligência de Groningen-2: GIT2]*. Amsterdam: Pearson Assessment and Information B. V.
- 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.
- Macedo, E. C., Mecca, T.P., Valentini, F., Laros, J. A., Lima, R.M.F., & Schwartzman, J.S. (2013). Utilizando o teste não verbal de inteligência SON-R 2½-7[a] para avaliar crianças com Transtornos do Espectro do Autismo. *Revista Educação Especial, 26*(47), 603-618.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling, 11*(3), 320-341.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past present, and future. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, test and issues* (2nd ed.). New York: The Guilford Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1-10.
- Mecca, T. P., Lima, R. M. F., Laros, J. A., Macedo, E. C., & Lowenthal, R. (no prelo). Transtorno do espectro autista: Avaliação das habilidades cognitivas utilizando o teste não-verbal SON-R 6-40.
- Mecca, T. P., Orsati, F. T., & Macedo, E. C. (2014). Non-Verbal cognitive profile of young children with autism spectrum disorders. *Psychology, 5*, 1404-1417.
- Myreliid, A., Bergman, B., Stromberg, E. M., Jonsson, B., Nyberg, F., Gustafsson, J., & Annerén, G. (2010). Late effects of early growth hormone treatment in Down syndrome. *Acta Paediatrica, 99*(5), 763-769.
- Neisser, U., Boodoo, G., Bouchard, T. J Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101.
- Nisbett, R. E., Blair, J. A., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence. New findings and theoretical development. *American Psychologist, 67*(2), 130-159.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill, INC.

- Oosterhuis, H. (2017). *Regression-based norming for psychological tests and questionnaires*. Tese de doutorado, Tilburg University, Netherlands.
- Pasquali, L. (2009). Psicometria. *Revista da Escola de Enfermagem da USP*, 43, 992-999.
- Pasquali, L. (2011). *Psicometria. Teoria dos testes na Psicologia e na Educação*. Petrópolis, RJ: Editora Vozes.
- Pilati, R., & Laros, J. A. (2007). Modelos de equações estruturais em psicologia: conceitos e aplicações. *Psicologia: Teoria e Pesquisa*, 23(2), 205-216.
- PNUD (2013). *Atlas do desenvolvimento humano no Brasil*. IPEA.
- Primi, R. (2002). Inteligência fluida: definição fatorial, cognitiva e neuropsicológica. *Paidéia*, 12(23), 57-75.
- Primi, R. (2012). Psicometria: Fundamentos matemáticos da Teoria Clássica dos Testes. *Avaliação Psicológica*, 11(2), 297-307.
- Primi, R., & Nunes, C. H. S. (2010). O Satepsi: desafios e propostas de aprimoramento. In Conselho Federal de Psicologia (Orgs). *Avaliação Psicológica: diretrizes na regulamentação da profissão* (pp. 129-148). Brasília: CFP.
- Raven, J. C. (1962). *Coloured Progressive Matrices*. London: Lewis.
- Raven, J., Raven, J. C., & Court, J. H. (2003, updated 2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Pearson Assessment.
- Reynell, J. K. (1985). *Reynell Developmental Language Scale*. Windsor: Nfer-Nelson.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2015). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223-237. doi: 10.1080/00223891.2015.1089249
- Schelini, P. W. (2006). Teoria das inteligências fluida e cristalizada: Início e evolução. *Estudos de Psicologia*, 11(3), 323-332.
- Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review*, 25(1), 12-27. doi: 10.1016/j.hrmr.2014.09.004
- Schrank, F. A., & Flanagan, D. P. (2003). *WJ-III: Clinical use and interpretation*. Boston: Academic Press.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-337.

- Schroots, J. J. F., & Alphen de Veer, R. J. Van (1976). *LDT, Leidse diagnostische test. Deel 1/handleiding [Teste de diagnóstico LDT. Parte 1/manual]*. Amsterdam: Swets & Zeitlinger.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Seabra, A. G., Laros, J. A., Macedo, E. C., Abreu, N. (2014). *Inteligência e funções executivas: avanços e desafios para a avaliação neuropsicológica*. São Paulo, SP: Memnon Edições Científicas Ltda.
- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83-117.
- Sireci, S. G. (2013). Agreeing on validity arguments. *Journal of Educational Measurement*, 50(1), 99-104.
- Snijders-Oomen, N. (1943). *Intelligentieonderzoek van doofstomme kinderen [Pesquisa de inteligência com crianças surdas-mudas]*. Nijmegen: Berkhout.
- Snijders, J. Th., & Snijders-Oomen, N. (1958) eerste editie, (1970) tweede editie. *Snijders-Oomen niet-verbale intelligentieschaal SON-'58 [Escala de inteligência não-verbal Snijders-Oomen SON-'58]*. Groningen: Wolters-Noordhoff.
- Snijders, J. Th. & Snijders-Oomen, N. (1975). *Snijders-Oomen niet-verbale intelligentieschaal SON 2½-7 [Escala de inteligência não-verbal Snijders-Oomen SON 2½-7]*. Groningen: Wolters-Noordhoff.
- 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.
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- Stap, M. E. (2009). *Het functioneren van de nieuwe versie van de SON-I subtest Categorieën: Een vergelijking tussen culturen [O funcionamento da nova versão do subteste categorias do SON-I: Uma comparação entre culturas]*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Starren, J. (1975). *SSON 7-17. De ontwikkeling van een nieuwe versie van de SON voor 7-17 jarigen. Handleiding [SSON 7-17. O desenvolvimento de uma nova versão do SON para crianças de 7 a 17 anos. Manual]*. Groningen: Wolters-Noordhoff.
- Schumacker, R. & Beyerlein, S. (2000). Confirmatory factor analysis with different correlation types and estimation methods. *Structural Equation Modeling*, 7(4), 629-

- Tellegen, P. J., & Laros, J. A. (1993). The Snijders-Oomen Non-verbal intelligence tests: General intelligence tests or tests for learning potential? In J. H. M. Hamers, K. Sijtsma & A. J. J. M. Ruijsenaars, (Eds.), *Learning Potential Assessment: Theoretical, Methodological and Practical Issues* (pp. 267-283). Lisse: Swets & Zeitlinger.
- Tellegen, P. J., & Laros, J. A. (2008). *Evaluatie van de onderzoeksgegevens van de SON-I [Avaliação dos dados da pesquisa com o SON-I]*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Tellegen, P. J., & Laros, J. A. (2011). *SON-R 6-40. Snijders-Oomen Niet-verbale intelligentie test. Verantwoording [SON-R 6-40. Teste de inteligência não-verbal Snijders-Oomen. Relatório Científico]*. Göttingen, Germany: Hogrefe Verlag.
- Tellegen, P. J., & Laros, J. A. (2014). *SON-R 6-40. Snijders-Oomen non-verbal intelligence test: Research Report*. Göttingen, Germany: Hogrefe.
- Tellegen, P. J., & Laros, J. A. (2017). *SON-R 2-8. Snijders-Oomen niet-verbale intelligentietest [SON-R 2-8. Teste não-verbal de inteligência Snijders-Oomen]*. Amsterdam: Hogrefe.
- Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (1998). *Snijders-Oomen niet-verbale intelligentietest SON-R 2½-7. Handleiding*. Tweede gereviseerde editie [*Teste de inteligência não-verbal Snijders-Oomen SON-R 2½-7. Manual. Segunda edição revisada*]. Groningen: Hogrefe Verlag.
- Ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the exact and the approximate minimum rank of a covariance matrix. *Psychometrika*, *56*, 309-315.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209-220.
- Ullman, J. B. (2007). Structural Equation Modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics*, (pp. 676-780). Boston: Pearson Education.
- Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, *51*(4), 473-481.
- Valentin Kvist, A., & Gustafsson, J. E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence*, *36*, 422-436.
- Voncken, L., Albers, C. J., & Timmerman, M. E. (2017). Model selection in continuous test norming with GAMLSS. *Assessment*, 1-18. doi: 10.1177/1073191117715113

- Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4^a ed.). Baltimore: Williams & Wilkins.
- Wechsler, D. (2005a). *WISC-III NL. Wechsler Intelligence Scale for Children*. Third Edition Netherlands. London: The Psychological Corporation.
- Wechsler, D. (2005b). *Technische handleiding WAIS-III – Nederlandse bewerking. Herziene uitgave [Manual técnico do WAIS-III – adaptação holandesa. Edição revisada]*. Amsterdam: Harcourt Test Publishers.
- Wechsler, D. (2007). *Wechsler Intelligence Scale for Adults fourth edition – Integrated*. San Antonio, TX: Harcourt Assessment.
- Wechsler, D. (2009). *WPPSI-III-NL. Wechsler preschool and primary scale of intelligence*. Amsterdam: Pearson Assessment.
- Wechsler, D., & Naglieri, J. A. (2008). *WNV NL, Wechsler Non-verbal Scale of Ability. Technische Handleiding*. Amsterdam: Pearson Assessment.
- Wechsler, S. M., & Schelini, P. W. (2006). Bateria de habilidades cognitivas Woodcock-Johnson III: Validade de construto. *Psicologia: Teoria e Pesquisa*, 22, 287-295.
- Wu, Y. (2007). *Cognitive development of Chinese immigrant children in the Netherlands*. Heymans Instituut, Rijksuniversiteit Groningen: Internal Report.
- Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, 41, 86-94.
- Zhang, H., Yanqing, G. S., & Xiaoxun, T. (1997). Study on the adaptation of the Snijders-Oomen Non-verbal Intelligence Test. *Psychological Science China*, 20, 97-103.
- Zhu, J., & Weiss, L. (2005). The Wechsler scales. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 123-136). New York: The Guilford Press.
- Zimowski, M., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *Bilog-MG (version 3.0) [Computer program]*. Mooreville, IN: Scientific Software.
- Zumbo, B. D. (2014). What role does, and should, the test standards play outside of the United States of America? *Educational Measurement: Issues and Practice*, 33(4), 31-34.

Anexos

ANEXO 1 Carta apresentação da pesquisa SON-R 6-40 para escolas



Prezados,

Estamos realizando uma pesquisa nacional cujos objetivos são adaptar e obter evidências de validade do instrumento de avaliação da inteligência SON-R 6-40. Este instrumento é destinado à avaliação de pessoas entre 6 e 40 anos e pode ser administrado com ou sem uso da linguagem falada.

A pesquisa está vinculada à Universidade de Brasília (UnB) e ao Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações (PPG-PSTO), sendo coordenada pelo professor Dr. Jacob Arie Laros e pela doutoranda Renata Manuely Feitosa de Lima. O projeto foi aprovado pelo Comitê de Ética em Pesquisa do Instituto de Ciências Humanas da UnB (CEP/IH).

Os pais dos participantes deverão assinar um termo de consentimento. Além disso, a criança/adolescente selecionada também será questionada se deseja participar da pesquisa. Os pais ou a criança/adolescente poderão desistir da participação em qualquer momento da pesquisa sem nenhuma penalidade.

Em contrapartida, **a equipe de pesquisa se responsabiliza em oferecer:** (a) para a escola: parceria com a UnB - Instituto de Psicologia para disponibilizar aos pais uma avaliação de habilidades cognitivas dos seus filhos, bem como contribuir com uma pesquisa desenvolvida por uma Universidade pública do país; (b) para os pais e criança/adolescente: um relatório de desempenho, com sugestões de atividades para serem desenvolvidas a fim de potencializar o desenvolvimento cognitivo das habilidades mensuradas pelo teste. Caso haja alguma criança com desempenho discrepante, no relatório constará uma indicação para uma avaliação cognitiva mais aprofundada.

Dúvidas podem ser enviadas para:

- e-mail: jalaros@gmail.com; renatamanuely@gmail.com;

- telefone: (61) 8230-2899; (61) 3107-6902.

Atenciosamente,

Jacob Arie Laros
Renata Manuely Feitosa de Lima
Universidade de Brasília
Laboratório META/UnB – Métodos e Técnicas de Avaliação



Comitê de Ética em Pesquisa
Instituto de Ciências Humanas
Universidade de Brasília

ANÁLISE DE PROJETO DE PESQUISA

Título do Projeto: NORMATIZAÇÃO E VALIDAÇÃO DO TESTE NÃO-VERBAL DE INTELIGÊNCIA SON-R 6-40 PARA O BRASIL

Pesquisador responsável: FELIPE VALENTINI

Com base nas Resoluções 196/96, do CNS/MS, que regulamenta a ética da pesquisa em seres humanos, o Comitê de Ética em Pesquisa com Seres Humanos do Instituto de Ciências Humanas da Universidade de Brasília, após análise dos aspectos éticos, resolveu **APROVAR** o projeto intitulado “NORMATIZAÇÃO E VALIDAÇÃO DO TESTE NÃO-VERBAL DE INTELIGÊNCIA SON-R 6-40 PARA O BRASIL”.

O pesquisador responsável fica notificado da obrigatoriedade da apresentação de um relatório final sucinto e objetivo sobre o desenvolvimento do Projeto, no prazo de 1 (um) ano a contar da presente data (itens VII.13 letra “d” e IX.2 letra “c” da Resolução CNS 196/96).

Brasília, 13 de maio de 2011.

Debora Diniz - Coordenadora Geral - CEP/IH

ANEXO 3 Termo de informação aos pais ou responsáveis

Seu filho/tutelado foi convidado para participar da pesquisa “Normatização e validação do teste não-verbal de inteligência SON-R 6-40 para o Brasil”, que é coordenada pelo professor Dr. Jacob Laros (Universidade de Brasília – UnB) e pela estudante de doutorado Ms. Renata Manuely Feitosa de Lima. Esta pesquisa tem como objetivo a construção de um instrumento para avaliação do desenvolvimento cognitivo de brasileiros na faixa etária entre 6 e 40 anos. Destaca-se a relevância do estudo para aumentar a compreensão do desenvolvimento cognitivo da população brasileira nesta faixa etária.

Caso permita que seu filho/tutelado participe da pesquisa, ele responderá ao teste de inteligência SON-R 6-40 e a um questionário geral. O horário e local da sessão de avaliação serão combinados com o aplicador responsável pela sua região, que é um psicólogo devidamente treinado.

Os **riscos** envolvidos com a participação são mínimos, visto que eles responderão, basicamente, a problemas de raciocínio. Ao participar da pesquisa, seu filho receberá um relatório final individual com as conclusões da avaliação realizada, contendo orientações úteis sobre suas habilidades cognitivas, além de refletir sobre o próprio desenvolvimento cognitivo. Além disso, a participação poderá contribuir significativamente com as discussões científicas sobre o tema.

Todas as informações obtidas serão tratadas de forma sigilosa e os nomes dos participantes não serão divulgados em nenhum momento, sob hipótese alguma. A participação é **voluntária**, o que significa que o seu filho/tutelado poderá desistir a qualquer momento, sem que isso traga prejuízo ou penalidade.

Qualquer dúvida que você tiver a respeito desta pesquisa, poderá perguntar ao Dr. Jacob Laros e à Ms. Renata Manuely F. de Lima, pelos e-mails *jalaros@gmail.com*, ou *renatamanuely@gmail.com* ou telefones: (61) 31076902 e (61)8230-2899; ou pelo seguinte endereço: Campus Universitário Darcy Ribeiro, ICC, sala A1 061/4, Brasília, DF. Dúvidas a respeito da ética dessa pesquisa poderão ser encaminhadas ao CEP do Instituto de Ciências Humanas da UnB, localizado no Campus Universitário Darcy Ribeiro, ICC, sala B1 683.

PERMITO a participação do meu filho _____ na pesquisa.

Assinatura

Local e data

Pesquisadores responsáveis:
Renata Manuely Feitosa de Lima
CRP-01/17294
Jacob Arie Laros
Professor do Instituto de Psicologia - Universidade de Brasília
Departamento de Psicologia Social e do Trabalho

ANEXO 4

Termo de Consentimento

Este é um convite para você participar da pesquisa “Normatização e validação do teste não-verbal de Inteligência SON-R 6-40 para o Brasil”, que é coordenada pelo professor Dr. Jacob Laros (Universidade de Brasília – UnB) e pela Ms. Renata Manuely Feitosa de Lima. Esta pesquisa tem como objetivo a construção de um instrumento para avaliação do desenvolvimento cognitivo de brasileiros entre 6 e 40 anos.

Se quiser participar, você responderá ao teste de inteligência SON-R 6-40 e a um questionário geral. O horário e local da sessão de avaliação serão combinados com o aplicador responsável pela sua região, que é um psicólogo devidamente treinado.

Os **riscos** envolvidos com a participação são mínimos, visto que os instrumentos de coleta de dados correspondem apenas em resolver problemas de raciocínio. Ao participar da pesquisa você terá os seguintes **benefícios**: recebimento de um relatório final individual com as conclusões da avaliação realizada, contendo orientações úteis sobre suas habilidades cognitivas, oportunidade de refletir sobre o próprio desenvolvimento cognitivo e, além disso, a sua participação poderá contribuir significativamente com as discussões científicas sobre o tema.

Todas as informações obtidas serão tratadas de forma sigilosa e os nomes dos participantes não serão divulgados em nenhum momento, sob hipótese alguma. A participação é **voluntária**, o que significa que você poderá desistir a qualquer momento, sem que isso lhe traga nenhum prejuízo ou penalidade.

Qualquer dúvida que você tiver a respeito desta pesquisa, poderá perguntar ao Dr. Jacob Laros ou à Ms. Renata M. F. de Lima no endereço: Campus Universitário Darcy Ribeiro, ICC, sala A1 061/4, Brasília, DF; ou telefones: (61) 3107-6902, (61) 98230-2899; ou pelos e-mails *jalaros@gmail.com* ou *renatamanuely@gmail.com*. Dúvidas a respeito da ética dessa pesquisa poderão ser encaminhadas ao CEP do Instituto de Ciências Humanas da UnB, localizado no Campus Universitário Darcy Ribeiro, ICC, sala B1 683.

Eu _____ compreendi os objetivos da pesquisa “Normatização e validação do Teste não-verbal de Inteligência SON-R 6-40 para o Brasil” e como ela será realizada. Concordo em participar voluntariamente dessa pesquisa.

Assinatura

Local e data:

Pesquisadores responsáveis:
Renata Manuely Feitosa de Lima
CRP-01/17294

Jacob Arie Laros
Professor do Instituto de Psicologia - Universidade de Brasília
Departamento de Psicologia Social e do Trabalho

Anexo 5 – Folha de respostas do teste SON-R 6-40

Lembrete

Instruções Gerais

- Sempre utilize os exemplos A e B; o exemplo C (*Analogias e Categorias*) deve ser utilizado somente se o examinando não resolver o exemplo B sozinho.
- Preste auxílio e corrija somente os exemplos.
- Preste atenção para não fornecer informações extras àquelas contidas nas instruções verbais ou não-verbais.
- Não nomeie os desenhos utilizados.
- Utilize as instruções estendidas com os primeiros itens se for necessário.
- Informe ao sujeito, a cada item, se a solução está correta ou incorreta. Para tanto, espere até que o sujeito tenha completado sua resposta ou o limite de tempo tenha se esgotado.
- Não forneça nenhuma explicação adicional após o feedback de uma resposta incorreta.
- Pontue os escores (1=correto; 0= incorreto).

Subtestes com limite de tempo máximo (*Mosaicos e Padrões*)

- Comece a utilizar o cronômetro no exemplo B.
- Quando houver dúvida, pergunte ao sujeito se ele terminou o item.
- Solicite ao sujeito que pare quando o limite de tempo for alcançado. Caso ele esteja quase terminando, permita que ele o faça. Entretanto pontue o item como incorreto (0).
- Quando está claro que o sujeito não terá êxito em completar a tarefa é possível parar o cronômetro antes do tempo limite.

Procedimento adaptativo do teste

- *Escore*: o escore de cada série é igual ao número do último item administrado menos o número de erros na série.
- Em cada série, pare após dois erros (os erros não precisam ser consecutivos).
- Inicia a série-a com o item 1; as séries b e c devem ser iniciadas com o escore da série anterior, menos 2.
- *Exceção*: se os três primeiro itens (1a, 2a, 1b) estiverem incorretos, o subteste é encerrado.

Escores:	Bruto
1. Analogias	_____
2. Mosaicos	_____
3. Categorias	_____
4. Padrões	_____
Score Total	_____

SON-R 6-40

© P.J. Tellegen & J.A. Laros

Informações Gerais

Nome: _____ Código _____

Data da aplicação: ____/____/____

Data de nascimento: ____/____/____

Idade: _____ (anos e meses) Sexo: m / f

Local de Aplicação _____

Escolaridade _____ Profissão _____

Examinador: (código) _____ Nome: _____

Questionário Contextual

Cidade: _____ Estado: _____

Bairro: _____ Renda familiar: _____

Grau de escolaridade do pai _____ e da mãe _____

Ocupação do pai _____ e da mãe _____

Observações

(4=Muito baixa; 3=Baixa; 2=Bom; 1=Muito bom)

Motivação _____

Concentração _____

Cooperação _____

Compreensão das instruções _____

1. Analogias Início: __h: __min Término: __h: __min

- Não mencione o conceito no qual a transformação está baseada.

	1	2	3	4	5	6	7	8	9	10	11	12	último item	respostas incorretas	Escore
a													_____	- _____	= _____
	3	2	1	3	4	3	1	1	2	4	4	1			
b													_____	- _____	= _____
	2	3	3	1	4	2	3	2	4	1	4	2			
c													_____	- _____	= _____
	4	1	2	3	4	2	1	4	2	3	3	1			

1. Analogias

2. Mosaicos Início: __h: __min Término: __h: __min

Tempo máximo: 2 minutos por item

- A quantidade dos diferentes tipos de quadrado varia por item.
- Preste atenção para que apenas os compartimentos corretos estejam descobertos.
- Introduza os novos tipos de quadrados.
- Deixe que o sujeito devolva os quadrados depois de completar cada item.

	1	2	3	4	5	6	7	8	9	10	11	12	13	último item	respostas incorretas	escore
a														_____	- _____	= _____
	2 Tipos			4 Tipos de quadrados				6 Tipos de quadra dos								
b														_____	- _____	= _____
	2 Tipos			4 Tipos de quadrados				6 Tipos de quadrados								

2. Mosaicos

3. Categorias Início: __h: __min Término: __h: __min

- Dê feedback e pontue somente depois que o sujeito apontar as duas figuras.
- Não mencione o conceito no qual a similaridade está baseada.

	1	2	3	4	5	6	7	8	9	10	11	12	último item	respostas incorretas	escore
a													_____	- _____	= _____
	15	35	24	13	45	25	34	12	45	23	14	34			
b													_____	- _____	= _____
	24	34	25	14	24	35	13	15	23	25	14	23			
c													_____	- _____	= _____
	23	15	14	12	34	25	13	45	35	12	15	24			

3. Categorias

4. Padrões Início: __h: __min Término: __h: __min

Tempo máximo para os itens 1 - 8: 2 minutos por item

Tempo máximo para os itens 9 - 13: 3 minutos por item

- Dê ao sujeito lápis e borracha.
- Use a chave de correção se necessário.
- Para cada padrão, o cartão de papelão deve ser colocado abaixo da página do teste.

	1	2	3	4	5	6	7	8	9	10	11	12	13	último item	respostas incorretas	escore
a														_____	- _____	= _____
	Tempo máximo: 2 minutos								Tempo máximo: 3 minutos							
b														_____	- _____	= _____
	Tempo máximo: 2 minutos								Tempo máximo: 3 minutos							

4. Padrões
