



Universidade de Brasília

Instituto de Psicologia

Departamento de Psicologia Social e do Trabalho

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

**ADAPTAÇÃO E PROPRIEDADES PSICOMÉTRICAS
DOS SUBTESTES DE INTELIGÊNCIA DO IDS-2**

Talita de Araújo Alves

Orientador: PhD. Jacob Arie Laros

Brasília, DF

Julho de 2021



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

**ADAPTAÇÃO E PROPRIEDADES PSICOMÉTRICAS
DOS SUBTESTES DE INTELIGÊNCIA DO IDS-2**

Talita de Araújo Alves

Orientador: PhD. Jacob Arie Laros

Tese de doutorado apresentada ao Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações da Universidade de Brasília, como requisito parcial à obtenção do título de Doutora em Psicologia.

Brasília, DF

Julho de 2021

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

Banca Examinadora:

Prof. PhD. Jacob Arie Laros (*Presidente – Membro 1*)
Universidade de Brasília - UnB

Prof. Dr. Josemberg Moura de Andrade (*Membro 2*)
Universidade de Brasília - UnB

Profa. Dra. Monalisa Muniz Nascimento (*Membro 3*)
Universidade Federal de São Carlos - UFSCar

Profa. Dra. Mônia Aparecida Silva (*Membro 4*)
Universidade Federal de São João del-Rei - UFSJ

Profa. Dra. Cristiane Faiad de Moura (*Membro 5 - Suplente*)
Universidade de Brasília - UnB

“Let everything happen to you

Beauty and terror

Just keep going

No feeling is final”

— **Rainer Maria Rilke**

Agradecimentos

À minha família, mãe, pai, irmão, cunhada e sobrinhos, por todo amor, cuidado, torcida e compreensão. Sou grata por tudo que me ensinaram e que hoje é parte de quem eu sou. Aos pequenos Miguel e Pedro, por colorir e renovar nossos dias. Ao restante da minha família, avó, tios e primos, pelo carinho e compreensão.

Ao meu amado Hugo, por todo cuidado, carinho, suporte e atenção. Obrigado por estar comigo em todos os momentos e apoiar todas as minhas decisões. Seu amor e assistência foram essenciais para a conclusão desse projeto. Sou muito feliz em poder compartilhar minha vida com você.

Ao meu orientador, professor Laros, por quem tenho grande admiração. Obrigada por me acompanhar durante a trajetória da pós-graduação, por tudo que aprendi, pelo enorme privilégio de ser sua aluna e amiga, pela disponibilidade para ensinar e dedicação. Agradeço imensamente por despertar em mim o apreço à psicometria, com sua inteligência e zelo em tudo aquilo que faz.

Ao Fábio William, pela ética e cuidado em momentos tão importantes da minha vida. Obrigada por ser um profissional inspirador. Ao Dr. Marcelo Cavassani, pela ética, competência e cordialidade ao cuidar da minha saúde em um momento tão delicado.

Aos professores e membros da banca: Dr. Josemberg Andrade, Dra. Monalisa Muniz, Dra. Mônia Aparecida e Dra. Cristiane Faiad, por aceitarem o convite e pelas contribuições para o desenvolvimento desse trabalho. Eu fico imensamente feliz em poder contar com a participação de pesquisadores tão importantes da área de avaliação psicológica no Brasil.

Aos professores do PPG-PSTO que contribuíram diretamente para a minha formação científica: Jacob Arie Laros, Juliana Porto, Josemberg Andrade, Katia Puente-Palacios, Fábio Iglesias, Claudio Torres, Cristiane Faiad, Luiz Pasquali, Elaine Neiva e Ronaldo Pilati. Em especial agradeço à professora Cláudia Fukuda, por me apresentar à

psicometria e me orientar no início dessa jornada. E aos demais professores e cientistas brasileiros que me inspiram diariamente.

A todos os amigos queridos que a vida universitária e acadêmica me deu, em especial: Cássia, Hellen, Jesselyn, Gina, Francisco e Hannah, por todos os momentos de amizade, alegria e descontração vividos dentro e fora da Universidade. Aos atuais colegas de laboratório e aos que passaram pelo laboratório META. A Víthor Rosa Franco, pela amizade e incentivo, que sua dedicação à ciência, inteligência e carisma continuem inspirando por onde for.

Aos amigos que a vida me deu e aos colegas de profissão, Noemi, Beatriz, Rayssa, Jéssica, Luana, Allan, Luciana, Fernanda, Helena, Gabriela, Glaucia, André, Tiago, Aline, Flávia, Ingrid, Paula, Vitor, Thaís e tantos outros. Obrigada pelo apoio e amizade.

Aos meus alunos da UnB e do UNIEURO, Ingrid Fernandes, Amanda, Ana Beatriz, Anna Clara, Anna Luisa, Beatriz, Daniela, Emely, Ester, Igor, Ingrid, Juliana, Karoline, Lana, Maísa, Maria Danyelle, Marina, Rafaela, Raquel, Renata e Sarah, pelo interesse e participação nas pesquisas.

À Marianne Güntert e toda equipe da Editora Hogrefe Cetepp, pelo apoio às pesquisas brasileiras do IDS-2. Aos autores da versão original do teste, Alexander Grob e Priska Hagmann-von Arx, pelo interesse e apoio às pesquisas do IDS-2 no Brasil. Ao professor Leandro Malloy-Diniz e sua equipe de pesquisa, pela parceria e apoio nesse projeto.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, pelo apoio financeiro à pesquisa.

Por último, e não menos importante, à Charlotte, Lilith e Elvira, pelo afeto genuíno.

Sumário

Lista de Tabelas	8
Lista de Figuras	10
Resumo Geral	11
Apresentação	13
Manuscrito 1 - Adaptação dos subtestes de Inteligência do IDS-2 para o Brasil	26
Manuscrito 2 - Propriedades psicométricas dos itens e dos escores nos subtestes de Inteligência do IDS-2	45
Manuscrito 3 - Estrutura fatorial do conjunto de subtestes de inteligência no IDS-2	80
Considerações Finais	109

Lista de Tabelas

Apresentação

Tabela 1. <i>Distribuição dos subtestes de inteligência de acordo com os fatores do IDS-2</i>	18
-----------------------------------------------------------------------------------------------------	----

Manuscrito 1

Tabela 1. <i>Adaptação do subteste Repetir História</i>	36
---------------------------------------------------------------	----

Tabela 2. <i>Adaptação dos itens dos subtestes Repetir Números e Letras e Repetir Combinações de Números com Letras</i>	37
-------------------------------------------------------------------------------------------------------------------------------	----

Tabela 3. <i>Adaptação dos itens do subteste Nomear Categorias</i>	38
--------------------------------------------------------------------------	----

Manuscrito 2

Tabela 1. <i>Distribuição dos participantes por estado e cidades (N = 349)</i>	54
--------------------------------------------------------------------------------------	----

Tabela 2. <i>Índice de dificuldade (proporção de erros) dos itens de cinco subtestes de inteligência do IDS-2 segundo a TCT (N = 349)</i>	57
-------------------------------------------------------------------------------------------------------------------------------------------------	----

Tabela 3. <i>Poder discriminativo (correlação item total corrigida) dos itens de cinco subtestes de inteligência do IDS-2 segundo a TCT (N = 349)</i>	60
-------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Tabela 4. <i>Ajuste aos dados (-2 log likelihood) dos modelos de 1, 2 e 3 parâmetros de TRI para os cinco subtestes de inteligência do IDS-2</i>	62
--------------------------------------------------------------------------------------------------------------------------------------------------------	----

Tabela 5. <i>A dificuldade (parâmetro b) dos itens de cinco subtestes de inteligência do IDS-2 segundo o modelo 2-PL da TRI (N = 349)</i>	63
-------------------------------------------------------------------------------------------------------------------------------------------------	----

Tabela 6. <i>Poder discriminativo (parâmetro a) dos itens de cinco subtestes do IDS-2 segundo o modelo 2-PL da TRI (N = 349)</i>	65
----------------------------------------------------------------------------------------------------------------------------------------	----

Tabela 7. <i>Correlação de Spearman entre os índices de dificuldade e discriminação da TCT e TRI de cinco subtestes do IDS-2</i>	67
----------------------------------------------------------------------------------------------------------------------------------------	----

Tabela 8. <i>Características psicométricas dos escores brutos dos sete subtestes de inteligência do IDS-2 para seis grupos de idade</i>	70
-----------------------------------------------------------------------------------------------------------------------------------------------	----

Tabela 9. <i>Índice de fidedignidade (Alfa de Cronbach) dos escores nos cinco subtestes de inteligência do IDS-2 por grupo de idade</i>	71
-----------------------------------------------------------------------------------------------------------------------------------------------	----

Manuscrito 3

Tabela 1. <i>Média e desvio-padrão dos escores normatizados dos sete subtestes de inteligência do IDS-2 e do escore total para três grupos de idade</i>	90
Tabela 2. <i>Correlações entre os escores brutos e escores normatizados dos subtestes de inteligência do IDS-2 com a idade exata dos sujeitos (N = 349)</i>	92
Tabela 3. <i>Cargas fatoriais padrão e estrutura e comunalidades usando MRFA e Rotação Promin dos escores normatizados de sete subtestes de inteligência do IDS-2</i>	94
Tabela 4. <i>Cargas fatoriais usando análise fatorial de segunda-ordem (solução Schmid-Leiman) dos escores normatizados de sete subtestes de inteligência do IDS-2</i>	97
Tabela 5. <i>Índices de ajuste de três modelos de análise fatorial confirmatória dos escores normatizados de sete subtestes de inteligência do IDS- 2</i>	98
Tabela 6. <i>Confiabilidade composta (CC) e Variância Média Extraída (VME) dos fatores do modelo com melhor ajuste aos dados (M-3)</i>	101

Lista de Figuras

Manuscrito 3

Figura 1. <i>Resultado para modelo de mensuração com melhor ajuste aos dados na Análise Fatorial Confirmatória (M-3)</i>	99
--------------------------------------------------------------------------------------------------------------------------------	----

Resumo Geral

O IDS-2, *Intelligence and Development Scales*, é um instrumento de avaliação das funções cognitivas e do desenvolvimento destinado a sujeitos com idades entre 5 e 20 anos. A testagem da inteligência no IDS-2 compreende 14 subtestes que avaliam sete fatores baseados no modelo CHC de inteligência. O presente trabalho apresenta três manuscritos que reúnem estudos com os subtestes de inteligência do IDS-2 voltados para a população brasileira. O primeiro manuscrito descreve a adaptação dos subtestes de inteligência do IDS-2 para o contexto brasileiro. Os resultados desse estudo indicaram a necessidade de realizar algumas modificações em estímulos visuais, auditivos e de palavras dos itens, a fim de torná-los mais adequados à cultura brasileira. O segundo manuscrito apresenta as propriedades psicométricas dos itens e dos escores de cinco subtestes de inteligência do IDS-2. Os itens dos cinco subtestes de inteligência do IDS-2 foram analisados por meio dos modelos da TCT e da TRI, e foram estimados os parâmetros de dificuldade e discriminação dos itens. Os índices de dificuldade indicaram que os itens avaliam uma grande amplitude de níveis de habilidade, com índices adequados de poder discriminativo. O terceiro manuscrito descreve as evidências de validade obtidas com base na estrutura interna do teste e relata a estimação da fidedignidade dos escores. Os resultados da análise fatorial indicaram um modelo hierárquico com um fator geral de segunda ordem (Fator Geral de Velocidade e Memória) e três fatores de primeira ordem (Memória de Longo Prazo, Memória Viso-Espacial de Curto Prazo e Velocidade de Processamento), compatíveis com os pressupostos teóricos do modelo CHC. Os resultados dos três estudos apontam para o grande potencial do IDS-2 como ferramenta para o campo da avaliação psicológica no Brasil.

Palavras-chave: *Adaptação de testes psicológicos; IDS-2; Teste de inteligência; Propriedades Psicométricas; Evidências de validade.*

General Abstract

The IDS-2, Intelligence and Development Scales, is a broad scope instrument for the assessment of cognitive functions and development functions designed for subjects aged between 5 and 20 years. The cognitive functions consist of the function areas of Intelligence and Executive functions, while the developmental functions consist of the function areas of Psychomotor skills, Social and emotional competence, Scholastic skills and Attitude to work. The intelligence part of the IDS-2 consists of 14 subtests that assess seven factors from the CHC-model of intelligence. The present work contains three manuscripts that report and evaluate the results of studies with the IDS-2 intelligence subtests on a sample of Brazilian subjects. The first manuscript relates the adaptation of a number of the IDS-2 intelligence subtests to the Brazilian context. The results of this study indicated the necessity of some modifications of the items' visual, auditory and word stimuli in order to make them more suitable for the Brazilian culture. The second manuscript reports the estimation of psychometric properties of the items and scores of five IDS-2 intelligence subtests. The parameters of item difficulty and item discrimination of these subtests were estimated using both the CTT and the IRT model. The results indicated that the items are able to assess a wide range of skill levels, with adequate discriminative power. The third manuscript describes the validity evidence based on the internal structure of the test and reports the estimation of the reliability of the test scores. Results of factor analysis indicated a hierarchical model with a second order general factor (General Speed and Memory) and three first order factors (Long-term Memory, Short-term Visuospatial Memory and Processing Speed), consistent with the theoretical assumptions of the CHC-model. The results of the three studies indicate that IDS-2 represents a tool with a great potential for the field of psychological assessment in Brazil.

Keywords: Adaptation of psychological tests; IDS-2; Intelligence test; Psychometric properties; Validity evidence.

Apresentação

A inteligência está associada a variáveis importantes na vida das pessoas (Colom, 2006). O desempenho acadêmico, os resultados ocupacionais, os interesses profissionais, os cuidados em saúde e até o envolvimento em crimes possuem relação com a inteligência (Deary, 2012; Godoy et al., 2008; Neisser et al., 1996; Rindermann & Meisenberg, 2009; Rushton & Templer, 2009; Valentini et al., 2014). Na psicologia o estudo da inteligência é bastante antigo, sendo ainda uma das variáveis mais estudadas (Berg, 2000; Flores-Mendoza & Nascimento, 2001; Gottfredson, 1997; Neisser et al., 1996; Urbina, 2007). A avaliação da inteligência exerce um papel de destaque na tomada de decisões sobre aspectos importantes na vida dos indivíduos e seus impactos na sociedade. Ao longo do tempo a ciência psicológica tem acumulado informações sobre o tema e trabalhado no desenvolvimento e adequação de medidas a fim de subsidiar suas práticas profissionais e corroborar para o crescimento científico de outras áreas (Flores-Mendoza & Saraiva, 2018).

Os testes de inteligência estão inseridos no campo da avaliação psicológica, caracterizada como um processo sistematizado de levantamento e integração de informações (Conselho Federal de Psicologia – CFP, 2013). Os resultados dos testes compõem um conjunto de informações avaliativas, e são analisados com base no conhecimento psicológico dentro do contexto da demanda dos sujeitos e objetivos de avaliação (Segabinazi & Zamo, 2016). Os testes de inteligência representam ferramentas de trabalho e fontes de informações valiosas para os processos avaliativos, demonstrando empiricamente um alto valor preditivo (Flores-Mendoza et al., 2002; Flores-Mendoza & Saraiva, 2018).

A importância dos testes no processo avaliativo da inteligência ressalta a necessidade crescente da construção e adaptação de medidas que apresentem qualidade e adequação para a realidade do público brasileiro. No Brasil as diretrizes para a elaboração, uso e comercialização dos testes psicológicos são definidas pela Resolução nº 009/2018 do Conselho Federal de Psicologia- CFP. De acordo com a referida resolução os testes psicológicos devem apresentar, obrigatoriamente, condições mínimas exigidas a fim de demonstrar coerência técnico-científica para uso profissional. Os requisitos mínimos exigidos pela Resolução são: fundamentação teórica e definições dos construtos avaliados; objetivo do teste e contexto de utilização; pertinência teórica e qualidade técnica dos estímulos; evidências empíricas de validade e fidedignidade das interpretações dos escores; e apresentação dos procedimentos de aplicação e correção.

Em relação à fundamentação teórica, destacada entre os requisitos mínimos exigidos pela Resolução nº 009/2018 do CFP, entende-se que os testes estão diretamente relacionados às teorias que subsidiaram sua construção. Nesse sentido a evolução das teorias representa um grande desafio para os autores e usuários de testes, pois implica na reavaliação, adaptação e interpretação dos resultados dos instrumentos. No exemplo específico da inteligência, as proposições da teoria CHC (Cattell-Horn-Carroll) trouxeram um movimento expressivo de adaptação dos testes (Schelini & Wechsler, 2006). A teoria CHC representa a integração da teoria dos três estratos de Carroll (1993) e a teoria de Horn e Cattell (1966), e postula que a inteligência tem uma estrutura hierárquica de três níveis, no nível mais alto encontra-se um fator geral de inteligência ligado a dezesseis fatores amplos, situados no segundo nível, e no nível mais baixo estão cerca de 70 fatores específicos (Flanagan & Harrison, 2012). Devido à complexidade dessa teoria Ortiz (2015) afirma que, no cenário mundial, os instrumentos atuais de

inteligência avaliam apenas um número muito restrito de habilidades presentes no modelo CHC.

A análise de habilidades específicas do funcionamento intelectual confere robustez ao processo de avaliação. Pesquisadores e profissionais indicam a prática de *Cross Battery Assessment* para a obtenção de um retrato mais fiel do funcionamento intelectual dos indivíduos. O *Cross Battery Assessment* consiste na utilização de diversas medidas de avaliação da inteligência, a fim de obter dados do desempenho em habilidades específicas por meio dos instrumentos selecionados (Ortiz, 2015). Essa prática pode ser observada na avaliação neuropsicológica, em que os profissionais constroem uma bateria de testes com o intuito de avaliar o máximo de funções neuropsicológicas (Abreu et al., 2016).

A testagem de um número reduzido de habilidades pode ser recomendada para grupos especiais e situações que demandam tempo menor de aplicação (Alves, 2017). No Brasil, grande parte dos testes de inteligência está nesse tipo de classificação, que ora avaliam um fator geral de inteligência, ora fatores específicos. Além disso, os testes nacionais comumente apresentam apenas um tipo de tarefa, o que limita as interpretações realizadas com base nos resultados da testagem. De maneira particular, o público de crianças e jovens carece de baterias de avaliação da inteligência com foco em habilidades diversificadas.

O quantitativo restrito de instrumentos para a avaliação dos diferentes fatores de inteligência no cenário brasileiro sugere a demanda por novos testes e baterias de avaliação (Carreiro et al., 2014). Uma alternativa plausível à necessidade de novos testes de inteligência é a adaptação de medidas oriundas de diferentes culturas, uma vez que a literatura internacional indica a disponibilidade de instrumentos que contemplam os conteúdos de interesse (Borsa & Seize, 2017). Nesse sentido, o presente trabalho tem o intuito de contribuir para o campo da avaliação psicológica da inteligência no Brasil,

apresentando três estudos sobre os subtestes de inteligência do IDS-2, *Intelligence and Development Scales* (Grob & Hagmann-von Arx, 2018). O IDS-2 é uma bateria suíça de avaliação da inteligência e do desenvolvimento, e representa uma ferramenta com grande potencial para avaliação psicológica de crianças e jovens brasileiros.

O IDS-2 foi construído no departamento de Psicologia do Desenvolvimento e Personalidade da Faculdade de Psicologia da Universidade de Basel, na Suíça (Grob & Hagmann-von Arx, 2018). O instrumento é a segunda versão do teste IDS-1, o qual foi desenvolvido a partir da reformulação do Teste Kramer de Inteligência – KIT (Kramer, 1972). O IDS-1 foi lançado na Europa em 2009, e avalia crianças com idades entre 5 e 10 anos (Grob et al., 2009). O IDS-1 foi o primeiro teste em língua alemã a avaliar em conjunto a inteligência e o desenvolvimento, e atualmente é amplamente utilizado em clínicas psicológicas, psiquiátricas e pediátricas. Além disso, o IDS-1 foi adaptado e normatizado para outros países, dentre eles, a República Checa (Krejčířová et al., 2013) e a Itália (Ferri et al., 2015).

O IDS-1 se destaca pela utilização em diferentes contextos (escolar, clínica, pesquisas e outros). Estudos demonstram o uso dos subtestes do IDS-1 para grupos especiais, a exemplo das altas habilidades/superdotação (Hagmann-von Arx et al., 2008), nos transtornos do desenvolvimento da fala, das funções motoras e em outros transtornos como na dislexia e na discalculia. Na avaliação clínica, o uso do IDS é destacado em pessoas com TDAH (Transtorno do Déficit de Atenção e Hiperatividade), na Síndrome de Asperger, nos distúrbios do comportamento social e distúrbios emocionais (Gut et al., 2012). Pesquisas sobre as propriedades psicométricas do IDS indicaram índices altos de fidedignidade e de validade para os escores do instrumento (Meyer et al., 2009).

A segunda versão do IDS, o IDS-2, ampliou a faixa etária do instrumento a fim de contemplar o desenvolvimento desde os primeiros anos escolares até o início da vida

profissional. Dessa maneira, o IDS-2 é voltado para pessoas com idades dos 5 aos 20 anos (Grob & Hagmann-von Arx, 2018). No total o IDS-2 é formado por 30 subtestes, que avaliam o desenvolvimento geral dos sujeitos com foco em habilidades em áreas importantes do funcionamento, funções do domínio cognitivo e do desenvolvimento. No domínio cognitivo estão a inteligência e as funções executivas, enquanto o domínio do desenvolvimento engloba: habilidades motoras, competências socioemocionais, habilidades escolares e comportamento de trabalho. A aplicação do IDS-2 é realizada em sessões individuais e o tempo total de aplicação pode variar em função da idade do examinando e dos subtestes utilizados. Alguns subtestes do IDS-2 são destinados para idades específicas, e ainda há a possibilidade de o avaliador selecionar os subtestes que melhor atendem as necessidades de cada processo avaliativo (Grob & Hagmann-von Arx, 2018).

No IDS-2 a avaliação das habilidades motoras apresenta três subtestes: motricidade global, motricidade fina e visomotricidade. Os subtestes de competências socioemocionais analisam a habilidade de reconhecer emoções, de regular emoções e a avaliação de comportamentos socialmente competentes. As habilidades escolares são avaliadas por meio de quatro subtestes: raciocínio lógico-matemático, habilidades verbais, leitura e escrita. O comportamento de trabalho é composto por dois subtestes, conscienciosidade e motivação para aprender. Além disso, o IDS-2 também inclui a avaliação do examinador sobre a cooperação do sujeito durante a testagem (Grob & Hagmann-von Arx, 2018).

No domínio cognitivo o IDS-2 apresenta quatro subtestes de funções executivas (Fluência verbal, Atenção dividida, Nomear cores de animais e Percorrer rota uma vez) e 14 subtestes de inteligência, a saber: Copiar figuras, Repetir história, Assinalar duas características, Repetir números e letras, Reconhecer figuras, Completar matrizes,

Nomear categorias, Colocar discos, Descrever imagem, Riscar figuras, Repetir combinações de números com letras, Reconhecer figuras com rotação, Reconhecer figuras deslocadas e Nomear opostos. A teoria CHC foi a referência para o desenvolvimento dos subtestes de inteligência do IDS-2, os quais avaliam sete dimensões amplas do modelo: Processamento Visual; Memória de Longo Prazo; Velocidade de Processamento; Memória Fonológica de Curto Prazo; Memória Viso-espacial de Curto Prazo; Raciocínio Abstrato e Raciocínio Verbal (Grob & Haggmann-von Arx, 2018). A Tabela 1 apresenta a distribuição dos 14 subtestes de inteligência do IDS-2 em função dos fatores avaliados.

Tabela 1

Distribuição dos subtestes de inteligência de acordo com os fatores do IDS-2

Fator	Subtestes	Nº. subteste
Processamento Visual	Copiar Figuras	1
	Colocar Discos	8
Memória de Longo prazo	Repetir História	2
	Descrever Imagem	9
Velocidade de Processamento	Assinalar Duas Características	3
	Riscar Figuras	10
Memória Fonológica de curto prazo	Repetir Números e Letras	4
	Repetir Combinações de Números com Letras	11
Memória Viso-espacial de curto prazo	Reconhecer Figuras	5
	Reconhecer Figuras com Rotação	12
Raciocínio Abstrato	Completar Matrizes	6
	Reconhecer Figuras Deslocadas	13
Raciocínio Verbal	Nomear Categorias	7
	Nomear Opostos	14

Este trabalho faz parte do projeto que visa a adaptação, a obtenção de evidências de validade e a produção de normas brasileiras para o IDS-2. O projeto do IDS-2 no Brasil conta com o apoio da Hogrefe, principal editor europeu de materiais científicos em psicologia e psiquiatria, e tem o suporte da Hogrefe CETEPP no Brasil, que é parte do grupo editorial. Além disso, o projeto tem a parceria do grupo de pesquisa do professor Dr. Leandro Fernandes Malloy-Diniz do Departamento de saúde mental da Faculdade de Medicina da Universidade Federal de Minas Gerais - UFMG e do PhD. Jacob Arie Laros, professor titular e pesquisador colaborador sênior no Programa de Pós-Graduação em Psicologia Social, do Trabalho e das Organizações da Universidade de Brasília – UNB. O trabalho orientado pelo professor Laros tem foco na avaliação da inteligência, a qual é objeto de estudo deste trabalho. São apresentados três manuscritos, que descrevem estudos independentes com alguns dos subtestes de inteligência do IDS-2. O primeiro estudo objetivou adaptar quatro subtestes de inteligência do IDS-2 para o contexto brasileiro. No segundo estudo foram obtidas estimativas das propriedades psicométricas dos itens de cinco subtestes de inteligência do IDS-2. O último estudo objetivou obter evidências de validade e fidedignidade dos escores de sete subtestes de inteligência do IDS-2. Na sequência os subtestes de inteligência do IDS-2 são descritos em detalhes.

Os subtestes Copiar Figuras e Colocar Discos avaliam o fator Processamento Visual. No modelo CHC esse fator se refere à percepção visual, que indica a capacidade de perceber, armazenar e recuperar estímulos de natureza visual (Grob & Hagmann-von Arx, 2018). O subteste Copiar Figuras é composto por 20 itens organizados em ordem crescente de dificuldade. A aplicação é iniciada a partir do item de entrada destinado à faixa etária do examinando. A tarefa do subteste é reproduzir as figuras apresentadas utilizando peças em formato de retângulos e/ou triângulos. No subteste Colocar Discos o examinando deve posicionar discos metálicos sobre uma base magnética de acordo com o

modelo apresentado pelo examinador. O subteste é composto por 38 itens divididos em quatro figuras modelo. Os itens são representados por cada um dos discos metálicos. As figuras modelos são apresentadas de acordo com a faixa etária do examinando, assim sujeitos com idades de 5 a 8 anos realizam os itens das figuras um e dois (item 1 ao 12); os examinandos com idades entre 9 e 12 anos realizam os itens das figuras um, dois e três (item 1 ao 23); e os examinandos com idades entre 13 e 20 anos realizam os itens de todas as figuras (item 1 ao 38).

O fator Memória de Longo Prazo é definido na teoria CHC como a capacidade de armazenar e recuperar informações da memória de longo prazo (Flanagan & Harrison, 2012). No IDS-2 esse fator é representado pelos subtestes Repetir História e Descrever Imagem. No subteste Repetir História o sujeito ouve uma história e deverá recontá-la após o período mínimo de 20 minutos. A história apresenta duas partes e um total de 32 itens. A primeira parte da história é destinada a todas as idades do teste, 5 a 20 anos (item 1 ao 19), e a segunda parte da história é lida apenas para sujeitos com idade superior aos 9 anos (item 1 ao 32). Na tarefa do subteste Descrever Imagem o sujeito visualiza uma imagem e deverá descrevê-la após o período mínimo de 20 minutos. Para esse subteste também existem duas imagens, uma para sujeitos com idades entre 5 e 8 anos (item 1 ao 11), e outra para sujeitos com idade superior aos 9 anos (item 1 ao 21) (Grob & Hagmann-von Arx, 2018).

Os subtestes Assinalar Duas Características e Riscar Figuras avaliam o fator Velocidade de Processamento, em consonância com o fator de mesmo nome na teoria CHC, o qual indica a habilidade de executar tarefas cognitivas de forma automática e fluente (Flanagan & Harrison, 2012). No subteste Assinalar Duas Características o sujeito deve riscar figuras de papagaios com duas partes alaranjadas. Esse subteste apresenta duas folhas de exercícios, uma para sujeitos com idades de 5 a 9 anos, e outra

para sujeitos com idades de 10 a 20 anos. No subteste Riscar Figuras o sujeito deve riscar grupos de figuras com formas específicas na folha de exercício destinada à sua faixa etária (Grob & Hagmann-von Arx, 2018).

Os subtestes que avaliam Memória Fonológica de Curto Prazo são Repetir Números e Letras e Repetir Combinações de Números com Letras. Os dois subtestes são medidas do fator Memória de Curto Prazo da teoria CHC, o qual se refere à habilidade de manter e manipular informações imediatamente após obtê-las (Flanagan & Harrison, 2012). O subteste Repetir Números e Letras tem duas partes, a primeira parte apresenta 20 itens nos quais o sujeito deve repetir sequências de números ou letras na ordem que são ouvidos, enquanto na segunda parte o sujeito deve repetir a sequência na ordem inversa (Grob & Hagmann-von Arx, 2018). As mesmas tarefas são requeridas no subteste Repetir Combinações de Números com Letras, com a diferença de que os estímulos dos itens nesse último subteste são compostos por letras e números na mesma sequência.

Os subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação avaliam a Memória Viso-espacial de Curto Prazo, que em conjunto com os subtestes Repetir Números e Letras e Repetir Combinações de Números com Letras representam o fator Memória de Curto Prazo da teoria CHC. No subteste Reconhecer Figuras o sujeito deve memorizar a forma e a posição de algumas figuras para que na sequência possa reconhecê-las em meio a um conjunto maior de figuras. Os dois subtestes são compostos por 23 itens cada, organizados em ordem crescente de dificuldade. A aplicação é iniciada a partir do item de entrada destinado à faixa etária do examinando (Grob & Hagmann-von Arx, 2018). A diferença entre os subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação é que neste segundo subteste os estímulos sofrerão rotações.

No IDS-2 o fator Inteligência Fluida da teoria CHC é representado pelos subtestes Completar Matrizes e Reconhecer Figuras Deslocadas, que avaliam raciocínio

inferencial. O subteste Completar Matrizes é composto por 35 itens formados por figuras que apresentam um padrão de modificação. A tarefa do sujeito é escolher a opção que completa corretamente a sequência de figuras com base no padrão de modificação. O subteste Reconhecer Figuras Deslocadas é composto por 32 itens com figuras que representam um grupo ou conceito comum. A tarefa do sujeito é identificar a figura que não pertence ao grupo de imagens apresentadas (Grob & Hagmann-von Arx, 2018).

Por último são apresentados os subtestes de Raciocínio Verbal do IDS-2, Nomear Categorias e Nomear Opostos. Estes subtestes são medidas do fator Inteligência Cristalizada da teoria CHC, o qual abarca a aquisição do conhecimento verbal (Grob & Hagmann-von Arx, 2018). O subteste Nomear Categorias apresenta 34 itens formados por conjuntos de imagens ou palavras. A tarefa do sujeito é nomear corretamente o grupo apresentado. O subteste Nomear Opostos é composto por 34 itens em que o sujeito deve fornecer o nome oposto ao conceito das figuras ou palavras apresentadas.

Referências

- Abreu, N., Wyzykowski, A., Canário, N., Guimarães, P., & Reis, S. P. S. (2016). Como montar uma bateria para avaliação psicológica. In L. F. Malloy-Diniz, P. Mattos, N. Abreu, & D. Fuentes (Eds.), *Neuropsicologia: Aplicações clínicas* (pp. 107-123). Artmed.
- Alves, T. A., & Laros, J. A. (2017). Propriedades psicométricas do SON-R 6-40 em pessoas com deficiência intelectual. *Revista Psicologia Teoria e Prática, 19*(2), 151-163.
- Berg, C. A. (2000). Intellectual development in adulthood. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 117-137). Cambridge University Press.
- Borsa, J. C., & Seize, M. M. (2017). Construção e adaptação de instrumentos psicológicos: Dois caminhos possíveis. In B. F. Damásio & J. C. Borsa (Eds.), *Manual de desenvolvimento de instrumentos psicológicos* (pp. 15-38). Vetor.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Carreiro, L. R. R., Dias, N. M., Malloy-Diniz, L. F., Trevisan, B. T., Minervino, C. A. S. M., Roazzi, A., & Seabra, A. G. (2014). Testes de funções executivas no Brasil. In A. G. Seabra, J. A. Laros, E. C. Macedo, & N. Abreu (Eds.), *Inteligência e funções executivas: Avanços e desafios para a avaliação neuropsicológica* (pp. 84 – 111). Memnon.
- Colom, R. (2006). O que é inteligência. In C. Flores-Mendoza & R. Colom (Eds.), *Introdução à psicologia das diferenças individuais* (pp. 59-72). Artmed.
- Conselho Federal de Psicologia (2013). *Cartilha Avaliação Psicológica*. Retrieved from <http://satepsi.cfp.org.br/docs/cartilha.pdf>
- Deary I. J., (2012). Intelligence. *Annual Review of Psychology, 63*, 453- 482.
- Ferri, R., Rea, M., & Casagrande, M. (2015). *IDS – Intelligence and Development Scales*. Hogrefe.
- Flanagan, D. P., & Harrison, P. L. (2012). *Contemporary intellectual assessment: Theories, tests, and issues*. The Guilford Press.
- Flores-Mendoza, C. E., & Nascimento, E. (2001). Inteligência: O construto melhor investigado em psicologia. *Boletim de Psicologia, 51*(14), 37-64.
- Flores-Mendoza, C. E., Nascimento, E., & Castilho, A. V. (2002). A crítica desinformada aos testes de inteligência. *Estudos de Psicologia, 19*(2), 17-36.

- Flores-Mendoza, C., & Saraiva, R. (2018). Avaliação da inteligência: Uma introdução. In C. S. Hutz, D. R. Bandeira & C. M. Trentini (Eds.), *Avaliação da inteligência e da personalidade* (pp. 17-33). Artmed.
- Godoy, S., Noronha, A. P. P., Ambiel, R. A., & Nunes, M. F. O. (2008). Instrumentos de inteligência e interesses em orientação profissional. *Estudos de Psicologia*, 13(1), 75-81.
- Gottfredson, L. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24(1), 13-23.
- Grob, A., Meyer, C. S., & Hagemann-von Arx, P. (2009). *Intelligenz- und Entwicklungsskalen IDS für Kinder von 5–10 Jahren [Intelligence and Development Scales (IDS)]*. Huber Verlag.
- Grob, A., & Hagemann-von Arx, P. (2018). *Intelligenz- und Entwicklungsskalen für Kinder und jugendliche [Intelligence and Development Scales – 2 (IDS-2)]*. Hogrefe.
- Gut, J., Reimann, G., & Grob, A. (2012). Kognitive, sprachliche, mathematische und sozial-emotionale Kompetenzen als Prädiktoren späterer schulischer Leistungen: Können die Leistungen eines Kindes in den IDS dessen Schulleistungen drei Jahre später vorhersagen? *Zeitschrift für Pädagogische Psychologie*, 26, 213-220.
- Hagemann-von Arx, P., Meyer, C. S., & Grob, A. (2008). Assessing intellectual giftedness with the WISC-IV and the IDS. *Journal of Psychology*, 216(3), 173-180.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57(5), 253-270.
- Kramer, J. (1972). *Intelligenztest: Mit einer einföhrung in theorie und praxis der intelligenzprüfung*. Antonius.
- Krejčířová, D., Urbánek, T., Širůček, J., & Jabůřek, M. (2013). *IDS – Inteligenční a vývojová škála pro děti ve věku 5–10 let*. CZ Hogrefe.
- Meyer, C. S., Hagemann-von Arx, P., Lemola, S., & Grob, A. (2010). Correspondence between the general ability to discriminate sensory stimuli and general intelligence. *Journal of Individual Differences*, 31(1), 46-56.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77-101.
- Ortiz, S. O. (2015). CHC theory of intelligence. In S. Goldstein, D. Princiotta & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective and current concepts* (pp. 209-228). Springer.

- Rindermann, H., & Meisenberg, G. (2009). Relevance of education and intelligence at the national level for health: The case of HIV and AIDS. *Intelligence*, 37, 383–395.
- Rushton, J. P., & Templer, D. I. (2009). National differences in intelligence, crime, income, and skin color. *Intelligence*, 37, 341-346.
- Schelini, P. W., & Wechsler, S. M. (2006). Estudo da estrutura fatorial da bateria multidimensional de inteligência infantil. *Estudos de Psicologia*, 11(3), 232-332.
- Segabinazi, J. D., & Zamo, R. S. (2016). Psicodiagnóstico e inteligência. In C. Hutz, D. R. Bandeira, C. Trentini & J. S. Krug (Eds.), *Psicodiagnóstico* (pp. 258-273). Artmed.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Artmed.
- Valentini, F., Muniz, M., & Barbosa, A. A. G. (2014). Inteligência e desempenho acadêmico: Um enfoque das pesquisas brasileiras. In A. G. Seabra, J.A. Laros, E. C. Macedo, & A. Neander (Eds.), *Inteligência e funções executivas: Avanços e desafios para a avaliação neuropsicológica* (pp. 157-182). Memnon.

MANUSCRITO 1

Adaptação dos subtestes de inteligência do IDS-2 para o Brasil

Título em inglês

Adaptation of the intelligence subtests of the IDS-2 for Brazil

Título abreviado

Adaptação do IDS-2 para o Brasil

Resumo

Instrumentos para avaliar habilidades cognitivas frequentemente utilizam itens de natureza não-verbal. A adaptação transcultural de medidas não-verbais envolve diversos procedimentos, incluindo a avaliação dos especialistas e do público-alvo. Nesse contexto, o presente estudo objetivou adaptar quatro subtestes de inteligência do IDS-2 (*Intelligence and Development Scales*) para o Brasil. Participaram do estudo 37 pessoas ($N = 37$) com idades entre 5 e 21 anos ($M = 12,3$; $DP = 4,5$), residentes em Brasília-DF e Belo Horizonte-MG. O estudo adaptou os itens de quatro subtestes do IDS-2, a saber: Repetir História, Repetir Números e Letras, Nomear Categorias e Repetir Combinações de Números com Letras. Inicialmente a tradução para português da versão original do manual em alemão foi revisada por especialistas. Os subtestes foram submetidos a uma análise de adequação das instruções de aplicação e da compreensão dos estímulos junto ao público-alvo do instrumento. Após essa etapa os itens passaram por uma nova análise de especialistas a fim de realizar adequações culturais do conteúdo e, na sequência, foi realizada uma aplicação piloto junto ao público de interesse. Os resultados indicaram que, apesar dos esforços para a construção de itens de compreensão universal, um número pequeno de mudanças nos estímulos dos itens teve que ser realizado. Os estímulos alterados apresentavam conteúdos auditivos e/ou visuais que não são comuns para a realidade de crianças e jovens brasileiros. Esses achados reforçam a importância do processo de adaptação transcultural de medidas, especialmente para as medidas não-verbais.

Palavras-chave: *IDS-2; adaptação de testes não verbais; teste de inteligência.*

Abstract

Measures of cognitive abilities often use non-verbal items. Cross-cultural adaptation of non-verbal measures involves several procedures, including the assessment of experts and of the target group. In this context this study aimed to adapt four intelligence subtests of the IDS-2 (Intelligence and Development Scales) for Brazil. The study included 37 people (N = 37) aged between 5 and 21 years (M = 12.3; SD = 4.5), living in Brasília-DF and Belo Horizonte-MG. The study adapted items of four subtests from IDS-2, namely: Retelling a Story, Repeating Number and Letter Sequences, Naming Categories and Repeating Mixed Sequences of Numbers and Letters. Initially, the translation to Portuguese of the original German version of the subtests was reviewed by experts. The subtests also underwent an analysis of the adequacy of the administration guidelines and of whether the stimuli of the items were understood by the target group of the IDS-2. The items of the four subtests were analyzed by experts in order to make cultural adjustments to the content, after which the subtests underwent a new pilot administration. The results indicated that, despite efforts to construct measures with universal design, a small number of changes of the item stimuli had to be carried out. The altered stimuli presented auditory and visual contents that are not common to the reality of Brazilian children and teenagers. These findings reinforce the importance of the process of cross-cultural adaptation of measures, especially for non-verbal measures.

Key words: *IDS-2; adaptation of non-verbal tests; intelligence test.*

O processo de adaptação viabiliza a adequação cultural de instrumentos psicológicos quando se pretende utilizar um teste desenvolvido em uma outra cultura. Esse processo representa um dos caminhos para a garantia da testagem justa, e de práticas não discriminatórias de avaliação entre as diferentes culturas e populações (Gopaul-McNicol & Armour-Thomas, 2002; Ortiz et al., 2012). O princípio da justiça na testagem (*fairness*) é fundamental no processo de adaptação de medidas, pois protege os direitos dos examinandos em todos os aspectos e momentos da testagem. De acordo com esse princípio o processo de adaptação deve ter foco nas características próprias do público para o qual o instrumento é adaptado, buscando reduzir qualquer tipo de viés de testagem associado à cultura de construção do instrumento original (AERA et al., 2014). Em relação aos testes psicológicos a necessidade de adaptação transcultural reflete importantes implicações técnicas e éticas. Os testes psicológicos são ferramentas comumente empregadas nos processos de avaliação psicológica, cujos resultados são capazes de potencializar, ou até mesmo omitir, aspectos deficitários do funcionamento cognitivo, comportamental e emocional dos sujeitos, o que pode levar a conclusões e indicações errôneas e equivocadas.

As principais referências para a adaptação de medidas podem ser encontradas em trabalhos importantes da área: *Standards for Educational and Psychological Testing* (AERA et al., 2014), e os *Guidelines for Translating and Adapting Tests* (ITC, 2017). Os documentos apresentam orientações gerais sobre o processo de adaptação cultural dos testes: autorização do(s) autor(es) da versão original, tradução, formato de testagem e resposta, instruções, evidências de validade, fidedignidade e normas da versão adaptada (ITC, 2017). Esses documentos são guias fundamentais para a área de testagem e avaliação, entretanto, faltam diretrizes de natureza prática para os pesquisadores e profissionais que se interessam em trabalhar com medidas adaptadas.

No contexto brasileiro a Resolução nº 09/2018 do Conselho Federal de Psicologia - CFP estabelece as diretrizes para a realização da avaliação psicológica e regulamenta o SATEPSI – Sistemas de Avaliação de Testes Psicológicos, o qual é responsável pela avaliação técnica e científica dos testes encaminhados à apreciação da Comissão Consultiva em Avaliação Psicológica do Conselho (CCAP). Entre os critérios para a obtenção de um parecer favorável para a versão adaptada dos testes estão: a descrição dos critérios de tradução e adaptação do teste, e a equivalência da versão adaptada com o teste original. A comissão avalia o procedimento de tradução, os resultados do comitê de especialistas, a concordância entre os juízes e o estudo transcultural (Conselho Federal de Psicologia [CFP], 2018).

De modo geral, os procedimentos de adaptação indicados pela literatura mostram que apenas a tradução direta dos itens não é capaz de assegurar a equivalência de conteúdo dos instrumentos, pois questões linguísticas e de comunicação surgem a partir da discrepância de heranças culturais (Byrne et al., 2009; Smith et al., 2013). Alguns dos procedimentos de adaptação dos instrumentos envolvem a tradução e retro-tradução, a técnica bilingue, a abordagem de comitê e o estudo piloto (Hambleton, 2005; Sireci et al., 2006). A adaptação dos testes segue um processo complexo, e representa um conjunto de cuidados éticos, científicos e metodológicos a fim de garantir a adequação e a aplicabilidade das medidas em diferentes contextos culturais (Maneesriwongul & Dixon, 2004; Van de Vijver & Hambleton, 1996).

Cassepp-Borges, Balbinotti e Teodoro (2010) apresentam uma proposta de adaptação medidas que é composta pelas etapas de tradução, síntese e validação do conteúdo da versão preliminar do instrumento. Na etapa da tradução a versão original do instrumento poderá ser diretamente traduzida por profissional bilingue ou passar pelo procedimento de tradução reversa, em que o material é traduzido para o idioma alvo e

retraduzido para o idioma original. A etapa de síntese inclui o chamado método de comitê, o qual objetiva analisar e consolidar as traduções e produzir uma versão preliminar do instrumento. Por último é realizada a etapa de validação do conteúdo, em que juízes avaliam a pertinência do conteúdo do instrumento para medir o construto proposto.

Não há uma técnica perfeita para a adaptação dos instrumentos, apesar das sistemáticas propostas. Cha et al., (2007) recomendam utilizar técnicas combinadas que atendam às características específicas do estudo proposto. Borsa et al. (2012) apresentam uma proposta de seis etapas para o processo de tradução e adaptação de instrumentos para o contexto brasileiro. As etapas sugeridas incluem: a tradução do instrumento original, a síntese das versões traduzidas, a avaliação de juízes, a avaliação do instrumento pelo público-alvo, a tradução reversa e o estudo piloto. O ponto de destaque da proposta de Borsa et al. (2012) é a participação de amostras do público-alvo no processo de adaptação. Essa etapa tem ligação direta com a análise da compreensão dos itens por parte do público de interesse da adaptação do instrumento.

A perspectiva da participação do público-alvo no processo de adaptação pode ser especialmente útil em medidas de habilidades cognitivas. Essas medidas incluem testes de inteligência geral, funções relacionadas à atenção, memória, percepção visual, criatividade e raciocínio. Os itens desses instrumentos frequentemente utilizam estímulos visuais e de execução de tarefas, o que estende o processo de adaptação ao conteúdo dos itens, e a etapa de tradução pode estar mais restrita às instruções de aplicação. Nesse sentido é fundamental que a adaptação garanta a elegibilidade do teste e adequação ao público-alvo pretendido (ITC, 2017; Malda et al., 2008).

Na psicologia os testes de inteligência estão entre os instrumentos mais difundidos no campo da avaliação, sendo presentes nos mais diversos contextos culturais (Hogan,

2006; Flores-Mendoza & Saraiva, 2018). A inteligência representa um construto fundamental na psicologia, pois apresenta relações expressivas com o desenvolvimento humano, social e econômico dos sujeitos (Deary, 2012). Além disso, o processo de desenvolvimento da inteligência é influenciado por diferenças nos contextos socioculturais e oportunidades educacionais, ocasionando variações significativas entre os resultados da testagem. Nesse contexto a adaptação de testes de inteligência deve considerar características do próprio instrumento e da cultura alvo para determinar a extensão e o foco do processo de adaptação (Malda et al., 2008).

O presente estudo está inserido no contexto da adaptação cultural de testes de inteligência e objetiva adaptar quatro subtestes de inteligência do IDS-2 (*Intelligence and Development Scales*) para o contexto de crianças e adolescentes brasileiros. O teste IDS-2 é uma bateria suíça de avaliação da inteligência e do desenvolvimento para crianças e jovens com idades entre 5 e 20 anos (Grob & Hagmann-von Arx, 2018). O IDS-2 é composto por 30 subtestes que avaliam funções cognitivas (inteligência e funções executivas) e funções do desenvolvimento (habilidades psicomotoras, competências socioemocionais, competências escolares e atitude de trabalho) (Grob & Hagmann-von Arx, 2018). O IDS-2 é uma medida robusta e promissora para o Brasil, pois avalia diversos fatores de inteligência e apresenta tarefas inovadoras de avaliação, as quais não são identificadas nos testes atualmente disponíveis no contexto brasileiro (SATEPSI, 2020).

Método

Participantes

O estudo teve a participação de 44 sujeitos em diferentes etapas. A segunda etapa do estudo contou com a participação de duas psicólogas com experiência na área de avaliação psicológica, ambas do Distrito Federal - DF. Na terceira etapa participaram

sete sujeitos com idades entre 5 e 7 anos, dois do sexo masculino e cinco do sexo feminino, todos residentes do DF. A quarta etapa teve cinco participantes, dois homens e três mulheres, todos com formação em psicologia e atuando em projetos de adaptação e construção de medidas na Universidade de Brasília - UnB. Por último, participaram da quinta etapa do estudo 30 sujeitos ($n = 30$) com idades entre 5 e 21 anos ($M = 12,3$; $DP = 4,5$), sendo a maioria do sexo feminino (60%). Essa amostra incluiu 15 participantes residentes em Belo Horizonte - MG e 15 sujeitos do DF.

Instrumentos

Foram adaptados quatro subtestes de inteligência do IDS-2, a saber: Repetir História, Repetir Números e Letras, Nomear Categorias e Repetir Combinações de Números com Letras (Grob & Hagmann-von Arx, 2018).

O subteste Repetir História avalia o fator Memória de Longo Prazo. A tarefa do subteste consiste em ouvir e recontar uma história após o período mínimo de 20 minutos. A história é dividida em duas partes. A primeira parte é contada para examinandos de 5 a 20 anos de idade, enquanto a segunda parte é contada apenas para examinandos com idades superiores aos 9 anos. O subteste apresenta 32 itens relacionados ao conteúdo da história. Os itens 1 ao 19 são voltados para os examinandos com idades entre 5 e 8 anos, e os itens 20 ao 32 são direcionados para examinandos com idades a partir dos 9 anos (Grob & Hagmann-von Arx, 2018).

O subteste Nomear Categorias avalia o Raciocínio Verbal. O subteste tem 34 itens formados por grupos de imagens ou palavras. A tarefa do sujeito é analisar e nomear corretamente a categoria que representa o conjunto de palavras ou imagens apresentadas (Grob & Hagmann-von Arx, 2018).

Os subtestes Repetir Números e Letras e Repetir Combinações de Números com Letras avaliam a Memória de Curto Prazo. A tarefa do examinando nos dois subtestes é

ouvir e repetir a sequência de números ou letras que foi apresentada. Os subtestes são formados por 23 itens cada. Os dois subtestes se diferenciam pela característica dos itens, no subteste Repetir Números e Letras os itens são compostos por número ou letras, e no subteste Repetir Combinações de Números com Letras os itens apresentam sequências de números e letras juntos (Grob & Hagmann-von Arx, 2018).

Procedimentos

Os quatro subtestes de inteligência do IDS-2 foram submetidos a cinco etapas de adaptação, que incluíram: a) tradução dos itens originais em língua alemã para o português; b) revisão da tradução; c) estudo piloto com os itens originais traduzidos para o português; d) análise de juízes; e) estudo piloto com itens adaptados. As etapas de adaptação adotadas neste estudo foram construídas com base nas principais recomendações do campo de testagem e avaliação psicológica, tais como os *Standards for Educational and Psychological Testing* (AERA et al., 2014) e os *Guidelines for Translating and Adapting Tests* (ITC, 2017). Além disso, observou-se os procedimentos adotados em pesquisas anteriores de adaptação cultural de instrumentos de avaliação cognitiva (Borsa & Seize, 2017; Malda et al., 2008; Tellegen & Laros, 2004).

A tradução dos itens originais dos quatro subtestes de inteligência do IDS-2 foi realizada a partir do manual em alemão do teste por tradutor bilingue com expertise no processo de desenvolvimento do IDS-2 para a Alemanha, Suíça e Áustria. A tradução das instruções de aplicação e dos itens do instrumento foi revisada por dois psicólogos da área de avaliação psicológica. O objetivo da revisão foi analisar e adequar as instruções das tarefas para torná-las mais próximas da linguagem natural (Harkness, 2003).

Após as etapas de tradução e revisão do manual foi realizado um estudo piloto para verificar a clareza das instruções e a compreensão do conteúdo dos itens em uma amostra formada por sujeitos da população alvo da adaptação do instrumento. Nessa etapa sete

crianças e adolescentes com idades entre 5 e 20 anos responderam aos itens dos quatro subtestes de inteligência do IDS-2. Em sessões individuais os sujeitos forneciam respostas aos itens do teste e identificavam os estímulos e os conceitos apresentados. Além disso, os participantes relatavam de que maneira chegaram à resposta fornecida, e se haviam considerado alternativas de respostas. A testagem ocorreu de maneira não padronizada a fim de avaliar a compreensão dos examinandos sobre os materiais do teste e os procedimentos de testagem (Tellegen & Laros, 2004; Malda et al., 2008).

Os resultados do estudo piloto foram submetidos à apreciação de juízes. Essa etapa contou com a participação de cinco psicólogos com experiência em construção e adaptação cultural de medidas e objetivou avaliar a pertinência dos itens para medir o construto pretendido, a clareza e a objetividade das instruções de aplicação do teste, a adequação do material do teste para a faixa etária do instrumento e a adequação cultural dos itens. Durante a análise de juízes também foram obtidas sugestões de modificação dos itens que apresentaram divergência cultural.

Todos os itens que apresentaram discrepâncias semânticas foram adaptados para estímulos ou expressões mais adequados culturalmente. Os novos estímulos foram submetidos a um segundo estudo piloto. Os itens adaptados são descritos com mais detalhes na sessão de resultados.

Resultados

As adaptações realizadas no subteste Repetir História envolveram mudanças em palavras e expressões traduzidas da versão original em língua alemã. As modificações foram feitas na etapa de revisão da tradução. A Tabela 1 apresenta os termos originais e adaptados no subteste.

Tabela 1*Adaptação do subteste Repetir História.*

Termo original	Termo Adaptado	Etapa de adaptação
Nico	João	Revisão da tradução
Judith	Julia	Revisão da tradução
Bote	Barco	Revisão da tradução
Bebida gelada	Suco gelado	Revisão da tradução
Campainha (celular)	Alarme	Revisão da tradução
Nevando	Chovendo	Revisão da tradução

Os nomes dos personagens da história original foram substituídos, uma vez que os nomes da versão original não eram comuns à cultura brasileira. Para a realização da substituição foi feita uma consulta aos dados censitários do Instituto Brasileiro de Geografia e Estatística - IBGE do ano de 2010, os quais continham os 20 nomes mais frequentes no Brasil no ano do levantamento. A expressão “bebida gelada” foi modificada para “suco gelado”. A palavra bebida é indefinida e, devido a fatores culturais, pode receber a conotação de bebida alcoólica, o que não seria apropriado para a faixa etária do público-alvo do instrumento. A palavra bote foi modificada para barco em todos os momentos que aparece na história, pois a palavra barco é o termo genérico mais utilizado no Brasil para denominar embarcações de pequeno porte. O mesmo critério foi utilizado para modificar a palavra campainha, a qual foi substituída por alarme. Outra adaptação importante diz respeito ao uso do verbo nevar, evento climático que não é comum no território brasileiro, o qual foi substituído pelo verbo chover.

Tabela 2

Adaptação dos itens dos subtestes Repetir Números e Letras e Repetir Combinações de Números com Letras.

Item original	Item Adaptado	Etapas de adaptação
6, 3	6, 2	Revisão da tradução/ Análise de Juízes
L, F	L, H	Revisão da tradução/ Análise de Juízes
G, L, C	G, L, I	Revisão da tradução/ Análise de Juízes
6, 9, 1, 3	7, 9, 1, 3	Revisão da tradução/ Análise de Juízes
K, D, L, F, B	K, D, O, F, T	Revisão da tradução/ Análise de Juízes
W, L, B, K, G, F	A, L, B, K, G, I	Revisão da tradução/ Análise de Juízes
9, 2, 10, 8, 1, 3, 6	9, 2, 10, 8, 1, 3, 5	Revisão da tradução/ Análise de Juízes
D, F, C, L, W, K, B	D, F, C, L, O, K, B	Revisão da tradução/ Análise de Juízes
G, F, D, L, C, B, K, W	G, F, D, L, C, I, K, O	Revisão da tradução/ Análise de Juízes
3, 6, 5, 1, 8, 9, 2, 10, 4	3, 7, 5, 1, 8, 9, 2, 10, 4	Revisão da tradução/ Análise de Juízes
B, M, D, K, G, S, C, F, W	B, M, O, K, G, S, C, F, N	Revisão da tradução/ Análise de Juízes
G, K, B, F, D, L, W, M, C, S	G, K, B, F, D, L, A, M, C, S	Revisão da tradução/ Análise de Juízes
W, L, 5, 8, K, B	I, L, 5, 8, K, B	Revisão da tradução/ Análise de Juízes
9, F, 10, 8, W, 3, L	9, F, 10, 8, O, 3, L	Revisão da tradução/ Análise de Juízes
3, 6, 5, G, K, B, 2, D, 4	1, 6, 5, G, K, B, 2, D, 4	Revisão da tradução/ Análise de Juízes
8, 4, G, F, 2, 10, L, 7, K, W	8, 4, G, F, 2, 10, L, 7, K, T	Revisão da tradução/ Análise de Juízes

A Tabela 2 apresenta os resultados da adaptação do subteste Repetir Números e Letras e do subteste Repetir Combinações de Números com Letras. As adaptações foram realizadas com base nas sugestões obtidas nas etapas de revisão da tradução e análise de juízes. As modificações consideraram a pronúncia dos elementos nos itens. Os números e as letras que apresentavam semelhanças na pronúncia foram substituídos. Esse é o caso dos números 3 e 6, cuja proximidade da pronúncia poderia confundir o examinando. Outro aspecto observado diz respeito ao número de sílabas utilizado durante a pronúncia das letras e dos números. Buscou-se minimizar a inclusão de números e letras cuja pronúncia é mais extensa. O exemplo típico desse caso é a letra W, que foi substituída nos itens, pois sua pronúncia é mais longa: - dábliu. Os mesmos critérios de adaptação

utilizados no subtteste Repetir Números e Letras foram aplicados ao subtteste Repetir Combinações de Números com Letras.

O subtteste Nomear Categorias sofreu adaptações associadas à familiaridade cultural dos estímulos visuais e das palavras que compõem os itens. As informações obtidas por meio da etapa de estudo piloto com os itens originais subsidiaram as mudanças dos itens. Nessa etapa uma amostra de representantes do público-alvo do instrumento foi submetida a uma aplicação não padronizada do teste, em que se apontava o reconhecimento dos estímulos apresentados. A substituição das imagens e das palavras buscou manter a equivalência de conteúdo em relação aos estímulos originais. A Tabela 3 apresenta os resultados da adaptação dos itens do subtteste Nomear Categorias.

Tabela 3

Adaptação dos itens do subtteste Nomear Categorias

Item original	Item Adaptado	Etapa de adaptação
Imagem de uma luva	Imagem de uma saia	Estudo piloto de aplicação não padronizada
Corvo	Papagaio	Estudo piloto de aplicação não padronizada
Anão	Saci-Pererê	Estudo piloto de aplicação não padronizada
Elfo	Curupira	Estudo piloto de aplicação não padronizada
Unicórnio	Mula sem cabeça	Estudo piloto de aplicação não padronizada
Tsunami	Enchente	Estudo piloto de aplicação não padronizada
Avalanche	Terremoto	Estudo piloto de aplicação não padronizada

No subtteste Nomear Categorias o item que apresentava a imagem de uma luva foi modificado, pois três sujeitos da amostra avaliada identificaram a luva como um equipamento de proteção individual para o trabalho (EPI). O desenho da luva foi substituído pela imagem de uma saia, que passou a compor a categoria de vestuário

indicada pelo item. A palavra corvo foi substituída por papagaio, que é uma ave mais comum no território brasileiro. A categoria representada pelas palavras: anão, elfo e unicórnio não foi reconhecida pelos sujeitos da amostra do estudo piloto, então foi adaptada por figuras do folclore brasileiro. A familiaridade com os termos apresentados também foi o critério para a adaptação das palavras tsunami e avalanche, pois referem-se a eventos naturais que não ocorrem no Brasil.

Discussão

O objetivo do presente estudo foi adaptar quatro subtestes de inteligência do IDS-2 para o contexto brasileiro. Os subtestes adaptados foram: Repetir História, Repetir Números e Letras, Repetir Combinações de Números com Letras e Nomear Categorias (Grob & Haggmann-von Arx, 2018). Os itens dos quatro subtestes adaptados são formados por estímulos predominantemente verbais, em que a execução das tarefas está diretamente associada ao reconhecimento e compreensão do conteúdo apresentado. Nesse sentido as adaptações foram feitas com o intuito de garantir a compreensão dos estímulos e reduzir possíveis vieses de contexto, aproximando os itens do contexto brasileiro sem perder a equivalência com o conteúdo original dos itens.

Os subtestes Repetir Números e Letras e Repetir Combinações de Números com Letras avaliam a Memória Fonológica de Curto Prazo, que está associada à capacidade de processar informações auditivas, armazenando e recuperando dados na memória de curto prazo. A adaptação dos itens desses subtestes considerou aspectos fonológicos, como a quantidade de sílabas utilizadas durante a pronúncia dos números e das letras. Compreende-se que o processo de adaptação dos subtestes Repetir Números e Letras e Repetir Combinações de Números com Letras foi direcionado à teoria (Malda et al., 2008). De acordo com o modelo de memória de Baddeley a capacidade de armazenar

informações fonológicas está diretamente associada ao tamanho da pronúncia da palavra (Baddeley et al., 1975).

As adaptações realizadas nos subtestes Nomear Categorias e Repetir História foram direcionadas à familiaridade/reconhecimento e cultura (Malda et al., 2008). Em Nomear Categorias a resolução dos itens envolve a classificação de um grupo de estímulos pictóricos ou verbais, o que recorre às habilidades adquiridas por meio da educação formal e da cultura. A extensão do vocabulário dos sujeitos é fundamental para a solução dos itens. Nesses subtestes a falta de reconhecimento e familiaridade com as palavras geraria viés de testagem e prejudicaria o desempenho dos sujeitos avaliados, o que justificou a adaptação dos estímulos.

Os procedimentos adotados na adaptação dos quatro subtestes de inteligência do IDS-2 permitiram que os sujeitos, representantes do público-alvo da adaptação, pudessem declarar o entendimento sobre os estímulos apresentados e quais critérios foram utilizados para a resposta fornecida. Nesse tipo de procedimento o item é considerado adequado quando o público-alvo do teste demonstra a compreensão do material de testagem (DeMaio & Rothgeb, 1996; Willis, 2005). A exemplo desse aspecto destaca-se a utilização da imagem da luva no subteste Nomear Categorias, a qual não induziu ao erro de classificação das imagens no item, porém gerou a atribuição equivocada da função do objeto.

O IDS-2 apresenta 14 subtestes de avaliação da inteligência (Grob & Hagmann-von Arx, 2018). Desse número total, o presente estudo adaptou quatro subtestes, pois estudos piloto demonstraram que os estímulos dos demais subtestes se apresentaram adequados para o público-alvo. Os subtestes remanescentes avaliam raciocínio abstrato, velocidade, processamento viso espacial e memória viso espacial, e os itens são formados por figuras

e formas abstratas com baixo viés cultural, o que torna os subtestes mais flexíveis para a utilização em outros países e culturas.

Conclusão

A literatura sobre a adaptação de instrumentos aponta uma diversidade de procedimentos, porém não há consenso sobre as etapas necessárias para garantir que o processo de adaptação será suficiente para tornar uma medida adequada para uso em uma nova cultura (AERA et al., 2014; Cha et al., 2007; ITC, 2017). As etapas de adaptação dos quatro subtestes de inteligência do IDS-2 para o contexto brasileiro foram desenvolvidas com base nos relatos científicos de estudos com instrumentos de diferentes naturezas (Borsa et al., 2012; Esendemir & Bindak, 2019; Guo et al., 2013; Malda et al., 2008; Tellegen & Laros, 2004). Nesse sentido acredita-se que esse trabalho traz contribuições significativas, pois permite refletir sobre as práticas de adaptação transcultural de instrumentos de avaliação cognitiva, mais precisamente sobre a inteligência, em que as medidas apresentam diferentes formatos de avaliação.

Os resultados do estudo permitem afirmar que, mesmo que um instrumento tenha sido desenvolvido com o objetivo de acessar diversas culturas, não é correto renunciar a um processo rigoroso de verificação da adequação da medida junto à população alvo. A etapa de estudo piloto, em que há uma consulta ao público-alvo do teste, mostrou-se fundamental para a garantia da compreensão das tarefas e dos conteúdos do instrumento na amostra estudada (Borsa & Seize, 2017). Ressalta-se ainda a necessidade de avaliar os resultados obtidos a partir da aplicação do teste adaptado por meio de procedimentos estatísticos, a fim de reunir evidências empíricas da adequação do instrumento para a população e garantir o direito dos sujeitos a uma testagem culturalmente justa.

Referências

- American Educational Research Association [AERA]; American Psychological Association [APA]; & National Council on Measurement in Education [NCME] (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575-589.
- Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Adaptação e validação de instrumentos psicológicos entre culturas: Algumas considerações. *Paidéia (Ribeirão Preto)*, *22*(53), 423-432.
- Borsa, J. C., & Seize, M. M. (2017). Construção e adaptação de instrumentos psicológicos: Dois caminhos possíveis. In B. F. Damásio & J. C. Borsa, *Manual de desenvolvimento de instrumentos psicológicos* (pp. 15-38). Vetor.
- Byrne et al. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved educational and training in psychology. *American Psychological Association*, *39*(2), 94-105.
- Cassepp-Borges, V., Balbinotti, M. A. A., & Teodoro, M. L. M. (2010). Tradução e validação de conteúdo: Uma proposta para a adaptação de instrumentos. In L. Pasquali, *Instrumentação psicológica: Fundamentos e práticas* (pp. 506-520). Artmed.
- Cha, E. S., Kim, K. H., & Erle, J. A. (2007). Translation of scales in cross-cultural research: issues and techniques. *Journal of Advanced Nursing*, *58*(4), 386-395.
- Conselho Federal de Psicologia. *Resolução nº 009/2018*. Retrieved from <http://satepsi.cfp.org.br/docs/Resolucao-CFP-n-09-2018-com-anexo.pdf>
- Deary I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*, 453- 482.
- DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for cognitive and communicative processes in survey research* (pp. 177-196). Jossey-Bass.
- Esendemir, O., & Bindak, R. (2019). Adaptation of the developed to measure mathematical knowledge of teaching geometry in Turkey. *International Journal of Educational Methodology*, *5* (4), 547-565. <https://doi.org/10.12973/ijem.5.4.547>
- Flores-Mendoza, C., & Saraiva, R. (2018). Avaliação da inteligência: Uma introdução. In C. S. Hutz, D. R. Bandeira & C. M. Trentini (Eds.), *Avaliação da inteligência e da personalidade* (pp. 17-33). Artmed.

- Gopaul-Mcnicol, S., & Armour-Thomas, E. (2002). *Assessment and Culture: Psychological Tests with Minority Populations*. Elsevier.
- Grob, A., & Hagmann-von Arx, P. (2018). *Intelligenz- und Entwicklungsskalen für Kinder und jugendliche* [Intelligence and Development Scales – 2 (IDS-2)]. Hogrefe.
- Guo, Q., He, C., Wen, X., Song, L., Han, Z., & Bi, Y. (2013). Adapting the pyramids and palm trees test and the kissing and dancing test and developing other semantic tests for the Chinese population. *Applied Psycholinguistics*, 35(6), 1001-1019. <https://doi.org/10.1017/S0142716412000677>
- Hambleton, R.K. (2005) Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In: R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-3). Lawrence Erlbaum Associates.
- Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35-56). John Wiley & Sons.
- Hogan, T. P. (2006). *Introdução à prática de testes psicológicos*. LTC – Livros Técnicos e Científicos Editora S.A.
- International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (Second edition). <https://www.intestcom.org/>
- Malda, M., Van de Vijver, F. J. R., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*, 50(4), 451-468.
- Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A method review. *Journal of Advanced Nursing*, 48(2): 175-86. <https://doi.org/10.1111/j.1365-2648.2004.03185.x>
- Ortiz, S. O., Ochoa, S. H., & Dynda, A. M. (2012). Testing with culturally and linguistically diverse populations: moving beyond the verbal-performance dichotomy into evidence-based practice. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.) (pp.526-552). The Guilford Press.
- Sistema de Avaliação de Testes Psicológicos (2020). *Testes Favoráveis*. <https://satepsi.cfp.org.br/testesFavoraveis.cfm>
- Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2006). Evaluating guidelines for test adaptations: A methodological analysis of translation quality. *Journal of Cross-Cultural Psychology*, 37(5), 557-567. Doi:10.1177/0022022106290478
- Smith, P. B., Fischer, R., Vignoles, V. L., & Bond, M. H. (2013). *Understanding social psychology across cultures: Engaging with others in a changing world*, 2nd Edition. Sage Publications.

- Tellegen, P. J., & Laros, J. A. (2004). Cultural bias in the SON-R test: Comparative study of Brazilian and Dutch children. *Psicologia: Teoria e Pesquisa*, *20*(2), 103-111.
- Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*(2), 89-99. <https://doi.org/10.1027/1016-9040.1.2.89>
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.

MANUSCRITO 2

**Propriedades psicométricas dos itens e dos escores nos
subtestes de Inteligência do IDS-2**

Título em inglês

Psychometric properties of items and scores on the IDS-2 intelligence subtests

Título abreviado

Propriedades psicométricas do IDS-2

Resumo

A avaliação da qualidade dos itens de um instrumento envolve a aquisição de dados empíricos sobre o seu funcionamento junto à população. Nesse sentido, e considerando o contexto de adaptação brasileira do IDS-2, o presente estudo teve o objetivo de obter estimativas das propriedades psicométricas dos itens e dos escores de cinco subtestes de inteligência do IDS-2 para uma amostra brasileira. Participaram do estudo 349 crianças e adolescentes ($N = 349$) com idades entre 5 e 20 anos ($M = 11,9$, $DP = 4,6$), provenientes de cinco estados brasileiros e do Distrito Federal. Foram utilizados cinco subtestes de inteligência do IDS-2: Reconhecer Figuras, Repetir História, Colocar Discos, Reconhecer Figuras com Rotação e Descrever Imagem. Os parâmetros de dificuldade e discriminação dos itens dos cinco subtestes foram avaliados de acordo com a Teoria Clássica dos Testes (TCT) e com a Teoria de Resposta ao Item (TRI). A estimação do parâmetro de dificuldade dos itens de acordo com a TCT e TRI indicou resultados muito semelhantes. Por outro lado, foram observadas estimativas mais divergentes do poder discriminativo segundo a TCT e TRI. Foram encontrados coeficientes de fidedignidade (Alfa de Cronbach) dos escores dos subtestes entre 0,73 e 0,82 com um valor médio de 0,79. De maneira geral, os itens dos cinco subtestes estudados avaliam uma grande amplitude de níveis de habilidade do construto medido, e apresentam valores adequados de discriminação.

Palavras-chave: *IDS-2; propriedades psicométricas; teste de inteligência.*

Abstract

The evaluation of the quality of an instrument's items involves empirical data collection which provides information about its functioning in the target population. In this sense and considering the context of Brazilian adaptation of the IDS-2, the present study aimed to obtain estimations of the psychometric properties of the items and scores from five IDS-2 intelligence subtests in a Brazilian sample. The study included 349 children and adolescents ($N = 349$) aged between 5 and 20 years ($M = 11.9$, $SD = 4.6$), from five Brazilian states and the Federal District. Five IDS-2 intelligence subtests were administered: Placing Disks, Retelling a Story, Describing a Picture, Recognizing Figures and Recognizing Rotated Figures. The parameters of items difficulty and discrimination of the five subtests were estimated according to Classical Test Theory (CTT) and the 2 Parameter Logistic Model (2-PLM) of Item Response Theory (IRT). The estimation of the item difficulty parameter using TCT and TRI indicated very similar results. On the other hand, more divergent estimates of the discriminative power were obtained by TCT and TRI. Reliability coefficients (Cronbach's Alpha) of the subtest scores were obtained between .73 and .82 with a mean value of .79. In general, the results indicate that the items of the five intelligence subtests considered in this study are able to assess a wide range of skill levels of the measured construct, and also show adequate discrimination values.

Keywords: *IDS-2; psychometric properties; intelligence test.*

As evidências empíricas do funcionamento do teste junto à nova população são essenciais para a área de avaliação psicológica, pois impactam diretamente sobre a utilização dos resultados da testagem e nos processos de tomada de decisão acerca de aspectos importantes na vida das pessoas (*American Educational Research Association [AERA]; American Psychological Association [APA]; National Council on Measurement in Education [NCME]*, 2014).

No Brasil os testes psicológicos necessitam passar pela aprovação do Conselho Federal de Psicologia – CFP, por meio do Sistema de Avaliação de Testes Psicológicos – SATEPSI, que estabelece os requisitos para o reconhecimento e uso profissional dos testes psicológicos, conforme a Resolução CFP nº 009/2018 (CFP, 2021). Os instrumentos novos são submetidos à Comissão Consultiva em Avaliação Psicológica para a avaliação da qualidade técnico e científica. Entre os principais parâmetros avaliados encontram-se as evidências de validade obtidas a partir de fontes diversas, as análises psicométricas dos itens e as estimativas de precisão dos resultados nos testes (Andrade & Valentini, 2018).

A precisão é um dos parâmetros fundamentais no contexto da avaliação. Para que os resultados de um teste sejam considerados confiáveis é necessário avaliar sua consistência, a qual é estabelecida por meio de estimativas do grau de flutuação dos resultados. De acordo com a Teoria Clássica dos Testes - TCT as medidas estão sujeitas ao erro de mensuração, o que provoca inconsistências nos resultados e restringe a utilização do instrumento (Zanon & Filho, 2015). O erro de mensuração pode ocorrer por motivos diversos, desde condições físicas e psicológicas do sujeito durante a realização do teste até fatores associados ao ambiente de avaliação, vieses nos itens e procedimentos de aplicação e correção (Thompson, 2003). Dessa forma, a obtenção do erro de mensuração é fundamental, pois possibilita construir intervalos de confiança em que se

encontra o escore verdadeiro do sujeito (Lord & Novick, 1968; Andrade et al., 2010; Hogan, 2018).

Existem diferentes métodos para estimar a fidedignidade dos escores de um teste (AERA et al., 2014). O método de teste-reteste indica as possíveis variações nos escores associadas às condições do sujeito na testagem. Este método requer que o instrumento seja aplicado e reaplicado em dois momentos diferentes para a mesma amostra de sujeitos (Hogan, 2018). Espera-se uma correlação alta ($r \geq 0,70$) entre os dois momentos de aplicação, indicando a baixa flutuação nos resultados e a relativa estabilidade dos construtos psicológicos (Andrade & Valentini, 2018). Outro método para estimar a fidedignidade que requer apenas uma única aplicação do teste é baseado na consistência interna dos itens do teste. O coeficiente mais utilizado usando este método é o coeficiente alfa de Cronbach (Cronbach, 1951) e fornece uma indicação das flutuações associadas ao conteúdo do instrumento (Thompson, 2003). Independentemente do método de avaliação utilizado é desejável que os resultados das medidas apresentem altos índices de precisão, indicando que uma porção substancialmente menor do escore observado está associada ao erro de mensuração.

A abordagem da TCT é amplamente utilizada na literatura psicométrica, e se mantém em um lugar de importância ao lado de outros modelos, como a Teoria de Resposta ao Item - TRI. A TCT é criticada principalmente pela dependência que as estimativas possuem em relação à amostra (Embretson & Reise, 2000; Fan, 1998; Thorpe & Favia, 2012). Nesse sentido, os modelos da Teoria de Resposta ao Item surgem a partir do interesse e da necessidade de superar as limitações da Teoria Clássica dos Testes. Na TRI os parâmetros não enviesados de dificuldade (parâmetro b) e de discriminação (parâmetro a), podem, teoricamente, ser obtidos em amostras não representativas. Em outras palavras, na TRI os valores dos parâmetros tendem a ser mais independentes da

amostra (Embretson & Reise, 2000). É importante ressaltar que a independência da amostra na TRI implica que, se a amostra for muito grande, as estimativas dos parâmetros serão muito próximas dos valores verdadeiros, independentemente da forma como a amostra é desenhada (Verhelst, 2004). Uma outra diferença importante é que na TRI o item é considerado como a unidade básica de análise, em contraste com a TCT que trata o escore total no teste como a unidade de análise. Ressalta-se ainda que, apesar das diferenças, os dois modelos estão relacionados (Courville, 2004; Fan, 1998; MacDonald & Paunonen, 2002).

A TRI representa um conjunto de modelos matemáticos que viabiliza a avaliação das probabilidades de respostas dos indivíduos nos itens, em função de características dos itens e do traço latente (Andrade et al., 2010). Os modelos de TRI apresentam dois principais pressupostos: a unidimensionalidade e a independência local. O pressuposto da unidimensionalidade significa que um dado conjunto de itens mede apenas uma variável latente, o que pode ser conferido por meio de análise fatorial exploratória e da análise fatorial confirmatória. Já o pressuposto da independência local se constitui no fato de que os itens não estarão correlacionados se a variável latente for mantida constante (Conde & Laros, 2007). Outros pressupostos envolvem a escolha correta do modelo, avaliação de ajuste do modelo aos dados e o tamanho da amostra conforme a complexidade do modelo.

Para a compreensão da TRI a curva característica do item – CCI é um conceito fundamental, o qual representa uma função logística entre a resposta da pessoa ao item e o nível do construto avaliado, o *theta*. Na TRI as características técnicas dos itens podem ser avaliadas por meio de modelos matemáticos de um, dois ou três parâmetros (De Ayala, 2009; Edelen & Reeve, 2007; Baker, 2001). O modelo de um parâmetro (1-PL) estima a dificuldade dos itens, referindo-se à probabilidade de o sujeito fornecer a

resposta correta para o item (nos itens dicotômicos) em função do seu nível de habilidade. O parâmetro de dificuldade ou parâmetro b é a localização, e representa o ponto da curva característica em que a probabilidade de a resposta certa é de 50%. Quanto mais alto este parâmetro maior o teta (θ) que uma pessoa precisa ter para obter uma probabilidade de 50% de dar a resposta correta ao item (Baker & Kim, 2017; Verhelst, 2014). O modelo logístico de 2 parâmetros (2-PL) estima a dificuldade e a discriminação dos itens. O parâmetro de discriminação (a) indica o quanto um item é capaz de diferenciar indivíduos com níveis distintos da variável latente avaliada. Na CCI o parâmetro a é a inclinação da curva no valor do parâmetro de localização, e indica a extensão da relação do item com o construto. Para itens politômicos, os modelos de resposta diferem, mas incluem a especificação de localização e inclinação. Neste último caso, a curva característica corresponde a cada categoria de resposta. Por último o modelo de 3 parâmetros (3-PL) estima a dificuldade, a discriminação e a probabilidade de chute em itens com múltipla escolha.

Os modelos da TRI representam um importante avanço para as medidas em psicologia e para o campo educacional. A título de exemplo, observa-se o emprego das técnicas de TRI nas avaliações de desempenho educacional dos estudantes brasileiros: as provas do SAEB (Sistema de Avaliação da Educação Básica) e as provas do ENEM (Exame Nacional do Ensino Médio) (Andrade et al., 2010). Na avaliação psicológica da inteligência, o uso da TRI viabiliza análises detalhadas do funcionamento dos itens, contribuindo para o refinamento dos testes e aprimoramento das avaliações. Apesar dos avanços na aplicação da TRI nos instrumentos brasileiros não são todos os profissionais da avaliação que apresentam familiaridade com os significados e implicações práticas dessas análises, o que indica a necessidade de descrever esses conteúdos em detalhes nos manuais dos testes (Andrade & Valentini, 2018).

Os indicadores psicométricos dos itens são parte essencial do processo de avaliação de um instrumento, o qual busca atestar a qualidade da medida e sua adequação junto à população. Consequentemente essas exigências se aplicam aos testes que são importados de outros países e culturas. Nessas condições encontra-se o IDS-2 (*Intelligence and Development Scales*), que atualmente está em processo de adaptação, validação e normatização para o Brasil (Grob & Hagmann-von Arx, 2018). As Escalas de Inteligência e Desenvolvimento IDS-2 compõem uma bateria extensa de avaliação de funções cognitivas e desenvolvimentais para o público com idades entre 5 e 20 anos. O teste tem origem suíça, e foi desenvolvido do Departamento de Psicologia do Desenvolvimento e Personalidade da Faculdade de Psicologia da Universidade de Basel, sendo a ampliação do instrumento anterior, o IDS-1. O IDS-2 consiste em 30 subtestes, que avaliam seis domínios: (1) inteligência; (2) funções executivas (funções cognitivas); (3) psicomotricidade; (4) competências socioemocionais; (5) competências escolares e (6) comportamento de trabalho (Grob & Hagmann-von Arx, 2018).

No IDS- 2 a inteligência é avaliada a partir de sete fatores, conforme o modelo CHC (Cattell-Horn-Carroll), sendo eles: raciocínio fluido, raciocínio verbal, processamento visual, memória auditiva de longo prazo, memória auditiva de curto prazo, memória visual de curto prazo e velocidade de processamento (Schneider & McGrew, 2018). Trata-se de uma medida robusta da inteligência, que possibilita uma avaliação flexível e personalizada dos sujeitos, onde o profissional pode selecionar os subtestes conforme sua demanda. Ao considerar que os subtestes de inteligência do IDS-2 têm potencial para contribuir no campo da avaliação psicológica brasileira, este estudo apresenta como objetivo geral: Obter estimativas das propriedades psicométricas dos itens e dos escores de cinco subtestes de inteligência do IDS-2 para uma amostra brasileira de pessoas com idades entre 6 e 20 anos. Os objetivos específicos são: (1) obter

estimativas da fidedignidade dos escores nos cinco subtestes e (2) comparar os parâmetros psicométricos dos itens obtidos por meio da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI).

Método

Participantes

Participaram do estudo 349 crianças e adolescentes ($n = 349$), sendo 51,3% do sexo feminino ($n = 179$), com idades variando dos 5 aos 20 anos ($M = 11,9$, $DP = 4,6$). A escolaridade dos participantes variou desde a Educação Infantil ao Ensino Superior. A distribuição dos participantes de acordo com a escolaridade foi igual a: Educação Infantil (9,7%), Ensino Fundamental (58,2%), Ensino Médio (13,5%), e cursando o Ensino Superior (10,9%). A amostra também incluiu sujeitos que não estavam estudando numa escola no momento da coleta (2,3%) e que não declararam a escolaridade (5,4%). Os participantes eram provenientes dos seguintes estados brasileiros: Bahia (BA), Ceará (CE), Distrito Federal (DF), Goiás (GO), Minas Gerais (MG), Piauí (PI) e São Paulo (SP). Destaca-se que a maior parte dos sujeitos foram de cidades do Distrito Federal e do Goiás, 68,8% e 24,6%, respectivamente. Dentre os participantes apenas 2 sujeitos apresentaram histórico de tratamento psicológico e/ou psiquiátrico, e realizavam tratamento farmacológico (oxalato de escitalopram e cloridrato de metilfenidrato) no momento da aplicação do teste. A Tabela 1, abaixo, indica a distribuição de sujeitos por estado e cidades.

Tabela 1*Distribuição dos participantes por estado e cidades (N = 349)*

Localidade	%	Localidade	%
Bahia (BA)		Goiás (GO)	
Coribe	2,6	Águas Lindas	6,2
Ceará (CE)		Formosa	
Fortaleza	0,6	Luziânia	1,7
Distrito Federal (DF)		Novo Gama	
Águas Claras	4,0	Padre Bernardo	2,6
Plano Piloto	1,8	Valparaíso	7,2
Ceilândia	17,2	Minas Gerais (MG)	
Cidade Estrutural	2,0	Arinos	0,3
Guará	1,1	Piauí (PI)	
Itapoã	2,6	Corrente	0,3
Lago Norte	0,3	Esperantina	0,3
Paranoá	2,6	Riacho Frio	2,3
Park Way	0,6	São Paulo (SP)	
Planaltina	3,4	São Paulo	0,3
Recanto das Emas	8,3		
Riacho Fundo (I e II)	2,9		
Samambaia	10,5		
Santa Maria	4,9		
Sobradinho	0,6		
Taguatinga	5,1		
Vicente Pires	0,6		

Instrumentos

O presente estudo avaliou as propriedades psicométricas de cinco subtestes de inteligência do IDS-2 (*Intelligence and Development Scales*), a saber: Reconhecer Figuras, Repetir História, Colocar Discos, Reconhecer Figuras com Rotação e Descrever Imagem (Grob & Haggmann-von Arx, 2018). A descrição dos cinco subtestes analisados é apresentada a seguir:

1. Reconhecer Figuras: o subteste avalia a memória viso-espacial de curto prazo. A tarefa consiste em reconhecer figuras geométricas predeterminadas em um grupo de figuras. O subteste apresentou um índice de consistência interna (Alfa de Cronbach) igual a 0,87 para a faixa etária total do instrumento (5-20 anos) (Grob & Hagmann-von Arx, 2018).

2. Reconhecer Figuras com Rotação: o subteste também avalia o fator de memória viso-espacial de curto prazo. A tarefa é reconhecer figuras predeterminadas que sofreram rotação em um grupo de outras figuras. O subteste apresentou um índice de consistência interna (Alfa de Cronbach) igual a 0,87 para a faixa etária total do instrumento (5-20 anos) (Grob & Hagmann-von Arx, 2018).

3. Repetir História: avalia memória de longo prazo. A tarefa consiste em ouvir uma história e recontá-la após o período mínimo de 20 minutos. O subteste apresentou um índice de consistência interna (Alfa de Cronbach) igual a 0,93 para a faixa etária total do instrumento (5-20 anos) (Grob & Hagmann-von Arx, 2018).

4. Colocar Discos: avalia o processamento visual. A tarefa do sujeito é copiar formas utilizando pequenos discos de metal. O subteste apresentou um índice de consistência interna (Alfa de Cronbach) igual a 0,95 para a faixa etária total do instrumento (5-20 anos) (Grob & Hagmann-von Arx, 2018).

5. Descrever Imagem: avalia a memória de longo prazo. A tarefa consiste em visualizar uma imagem e descrever informações visuais sobre a imagem após o período mínimo de 20 minutos. O subteste apresentou um índice de consistência interna (Alfa de Cronbach) igual a 0,92 para a faixa etária total do instrumento (5-20 anos) (Grob & Hagmann-von Arx, 2018).

Procedimentos

Os dados foram coletados com a versão adaptada das escalas de inteligência do IDS-2 para o contexto brasileiro. Um grupo de aplicadores foi treinado para a coleta de dados, a qual ocorreu entre os anos de 2018 e 2020. Os participantes foram contatados e selecionados com base no critério de idade. O estudo obteve parecer de aprovação do Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais COEP-UFMG, e todos os participantes (ou responsáveis legais) assinaram o Termo de Consentimento Livre e Esclarecido (TCLE) ou o Termo de Assentimento Livre e Esclarecido (TALES) antes de responder ao instrumento. A coleta dos dados ocorreu individualmente nas casas ou nas escolas dos participantes, com duração média de 60 minutos.

Análise de dados

Para a análise dos dados foram utilizados os softwares: IBM SPSS *Statistics*, versão 23 e BILOG-MG versão 3.0. Os parâmetros de dificuldade e discriminação dos itens foram estimados com base na Teoria Clássica dos Testes (TCT) e com base na Teoria de Resposta ao Item (TRI). A fidedignidade dos escores foi estimada usando o método de consistência interna (Alfa de Cronbach).

Resultados

Os parâmetros de dificuldade e discriminação dos itens dos cinco subtestes de inteligência do IDS-2 foram inicialmente estimados com base na Teoria Clássica dos Testes (TCT). O parâmetro de dificuldade dos itens foi obtido por meio da proporção de sujeitos que indicaram a resposta errada para o item ($1-p$) (Crocker & Algina, 1986; Urbina, 2014;). Os índices de dificuldade dos itens segundo a TCT dos cinco subtestes de inteligência do IDS-2 encontram-se na Tabela 2.

Tabela 2

Índice de dificuldade (proporção de erros) dos itens de cinco subtestes de inteligência do IDS-2 segundo a TCT (N = 349)

Item	REF	RFR	REH1	REH2	COD1	COD2	COD3	DIM1	DIM2
1	0,01	0,01	0,23	0,61	0,08	0,42	0,59	0,41	0,45
2	0,02	0,01	0,40	0,70	0,18	0,39	0,62	0,30	0,53
3	0,02	0,01	0,15	0,74	0,12	0,43	0,66	0,56	0,79
4	0,01	0,02	0,33	0,69	0,24	0,54	0,74	0,41	0,49
5	0,01	0,02	0,40	0,87	0,14	0,55	0,77	0,07	0,08
6	0,04	0,04	0,15	0,48	0,15	0,70	0,79	0,34	0,56
7	0,08	0,25	0,46	0,67	0,23	0,73	0,77	0,58	0,62
8	0,17	0,19	0,62	0,53	0,15	0,72	0,75	0,62	0,70
9	0,12	0,20	0,60	0,47	0,15	0,79	0,77	0,61	0,63
10	0,28	0,22	0,45	0,44	0,16	0,83	0,75	0,42	0,48
11	0,10	0,14	0,59	0,62	0,25	0,81	0,78	0,26	0,32
12	0,13	0,17	0,58	0,62	0,11	0,57	0,76	0,31	0,46
13	0,14	0,24	0,11	0,42			0,80	0,50	0,57
14	0,59	0,58	0,21				0,82	0,37	0,47
15	0,47	0,53	0,25					0,64	0,76
16	0,55	0,64	0,40					0,46	0,57
17	0,78	0,85	0,30					0,54	0,61
18	0,83	0,88	0,34					0,52	0,60
19	0,87	0,93	0,28					0,53	0,54
20	0,93	0,95						0,45	0,45
21	0,94	0,96						0,71	0,81
22	0,96	0,97							
23	0,98	0,99							
M	0,39	0,42	0,46			0,52		0,50	

Notas: REF = Reconhecer Figuras; RFR = Reconhecer Figuras com Rotação; REH1 = Repetir História parte 1 (itens 1 a 19); REH2 = Repetir História parte 2 (itens 20 a 32); COD1 = Colocar Discos parte 1 (itens 1 a 12); COD2 = Colocar Discos parte 2 (itens 13 a 24); COD3 = Colocar Discos parte 3 (itens 25 a 38); DIM1 = Descrever Imagem parte 1 (itens 1 a 21); DIM2 = Descrever Imagem parte 2 (itens 22 a 42).

Os dados apresentados na Tabela 2 mostram que o subteste com a maior dificuldade, de acordo com os valores da média geral, foi o subteste Colocar Discos ($p = 0,52$), seguido por Descrever Imagem ($p = 0,50$), Repetir História ($p = 0,46$), Reconhecer Figuras com Rotação ($p = 0,42$), e Reconhecer Figuras ($p = 0,39$). Com a exceção de

alguns itens, os subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação apresentam uma tendência de aumento progressivo da dificuldade. O que chama a atenção é que para os subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação faltam itens na faixa de proporção de erro entre 30% e 40%.

Na Tabela 2 pode ser observado que o subteste Repetir História tem um padrão mais difuso de distribuição dos índices de dificuldade dos itens. Neste subteste não existe uma tendência de aumento progressivo da dificuldade nos itens. Os resultados do subteste Repetir História foram divididos em duas partes, a primeira parte (itens 1 a 19) é direcionada às idades de 5 a 20 anos, enquanto a segunda parte (itens 20 a 32) é voltada às idades de 9 a 20 anos. Na primeira parte do subteste os itens com as maiores dificuldades encontram-se na parte central do conjunto (itens 7 a 12). Os itens da segunda parte do subteste apresentaram um índice médio de dificuldade maior, com valores de proporção de erro superiores a 42%.

Os resultados do subteste Colocar Discos foram divididos em três partes, de acordo com a divisão das faixas etárias para a aplicação dos itens. A primeira parte do subteste (itens 1 a 12) é aplicada para todas as idades do instrumento, 5 a 20 anos. A segunda parte do subteste (itens 13 a 24) é acrescida à primeira parte para as idades de 9 a 12 anos, e a terceira parte do subteste (itens 25 a 38) é incluída para as idades de 13 a 20 anos. É possível observar que os itens mais fáceis são encontrados na primeira parte do subteste, com média de 0,16 para a proporção de erros, enquanto a segunda e terceira partes do subteste apresentaram maior dificuldade, com médias de 0,62 e 0,74, respectivamente.

Os resultados da Tabela 2 mostram que no subteste Descrever Imagem existe um padrão difuso de dificuldade dos itens, semelhante ao padrão observado no subteste Repetir História (também um subteste visando avaliar a memória de longo prazo). A

dificuldade média dos itens do subteste Descrever Imagem, entretanto, é maior em comparação com a dificuldade média dos itens do subteste Repetir História. Na primeira parte do subteste Descrever Imagem a média da dificuldade foi 0,46, e para a segunda parte foi 0,55. O que chama a atenção nesse subteste é que tanto na primeira quanto na segunda parte o item 5 se destaca por apresentar um valor muito baixo de proporção de erros (de respectivamente 0,07 e 0,08) em comparação aos demais itens do subteste.

Em geral, os resultados da Tabela 2 indicam que itens muito fáceis (proporção de erros $\leq 0,10$) e itens muito difíceis (proporção de erros $\geq 0,90$) ocorrem principalmente nos subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação. Nestes dois subtestes têm, respectivamente, 7 e 6 itens muito fáceis e 4 e 5 itens muito difíceis. Nos demais subtestes - Repetir História, Colocar Discos e Descrever Imagem – têm respectivamente 0, 1 e 2 itens muito fáceis e nenhum item muito difícil.

Em seguida o poder discriminativo dos itens foi analisado com base na Teoria Clássica dos Testes (TCT). Um item discrimina bem se as pessoas que têm alta proficiência na habilidade medida pelo teste obtêm, em média, um escore alto no item, enquanto a média do escore no item dos que têm pouca proficiência é baixa (Verhelst, 2014). Urbina (2014) considera o poder discriminativo de um item como uma evidência de validade do item.

O poder discriminativo dos itens pode ser obtido por meio da correlação item total corrigida e pode ser interpretado como um coeficiente de quão efetivamente o item discrimina entre pessoas que têm um escore total relativamente alto e aquelas que têm um escore total relativamente baixo (Crocker & Algina, 1986). Os coeficientes de discriminação dos cinco subtestes de inteligência do IDS-2 encontram-se na Tabela 3.

Tabela 3

Poder discriminativo (correlação item total corrigida) dos itens de cinco subtestes de inteligência do IDS-2 segundo a TCT (N = 349)

Item	REF	RFR	REH1	REH2	COD1	COD2	COD3	DIM1	DIM2
1	0,17	0,04	0,37	0,58	0,22	0,63	0,79	0,06	0,07
2	0,35	0,21	0,28	0,51	0,35	0,66	0,76	0,18	0,30
3	0,42	0,26	0,35	0,54	0,38	0,68	0,78	0,32	0,32
4	0,36	0,25	0,31	0,51	0,33	0,63	0,65	0,11	0,12
5	0,34	0,23	0,51	0,35	0,37	0,58	0,63	0,17	0,04
6	0,42	0,34	0,23	0,65	0,36	0,50	0,62	0,15	0,21
7	0,33	0,33	0,47	0,46	0,38	0,49	0,63	0,26	0,25
8	0,40	0,24	0,27	0,55	0,41	0,49	0,65	0,15	0,24
9	0,48	0,46	0,23	0,68	0,42	0,45	0,65	0,19	0,21
10	0,37	0,39	0,16	0,62	0,41	0,38	0,67	0,00	0,03
11	0,50	0,51	0,25	0,56	0,37	0,33	0,67	0,22	0,20
12	0,57	0,51	0,43	0,58	0,30	0,78	0,65	0,72	0,64
13	0,49	0,49	0,18	0,60			0,62	0,62	0,58
14	0,41	0,48	0,24				0,59	0,68	0,59
15	0,46	0,48	0,29					0,57	0,44
16	0,56	0,56	0,49					0,66	0,60
17	0,52	0,46	0,35					0,62	0,62
18	0,52	0,50	0,33					0,58	0,55
19	0,54	0,43	0,38					0,58	0,58
20	0,46	0,39						0,62	0,64
21	0,42	0,40						0,39	0,36
22	0,35	0,31							
23	0,29	0,19							
M	0,42	0,37	0,42			0,53		0,37	

Notas: REF = Reconhecer Figuras; RFR = Reconhecer Figuras com Rotação; REH1 = Repetir História parte 1 (itens 1 a 19); REH2 = Repetir História parte 2 (itens 20 a 32); COD1 = Colocar Discos parte 1 (itens 1 a 12); COD2 = Colocar Discos parte 2 (itens 13 a 24); COD3 = Colocar Discos parte 3 (itens 25 a 38); DIM1 = Descrever Imagem parte 1 (itens 1 a 21); DIM2 = Descrever Imagem parte 2 (itens 22 a 42).

Os dados da Tabela 3 mostram que o subteste que apresentou o maior índice médio de discriminação foi Colocar Discos, seguido por Reconhecer Figuras e Repetir História, que apresentaram a mesma discriminação média (0,42), e Reconhecer Figuras com Rotação e Descrever Imagem, que também apresentaram valores iguais para a média da

discriminação (0,37). De acordo com Hogan (2018) os valores de discriminação obtidos por meio da Teoria Clássica dos Testes são considerados bons entre 0,50 e 0,80, e podem ser considerados aceitáveis entre 0,30 e 0,50. Para Verhelst (2014) o valor de 0,30 também é reconhecido como uma referência mínima para o poder de discriminação do item. Em geral, os itens com discriminação abaixo de 0,30 estão no início e final dos subtestes, o que também está relacionado à variância, que tende a ser menor nesses itens. Destaca-se ainda o valor da discriminação do item 10 do subteste Descrever Imagem, que é igual a 0,00, indicando que, para a amostra estudada, esse item não discriminou o desempenho dos sujeitos em relação ao escore total.

Além das análises através da Teoria Clássica dos Testes (TCT), os itens dos subtestes também foram analisados usando a Teoria de Resposta ao Item (TRI). Antes de obter os parâmetros de dificuldade e discriminação dos itens, foi analisado qual modelo de TRI se ajustou melhor aos dados, o modelo 1-PL, 2-PL ou 3-PL. Os seguintes índices foram considerados: a diferença do índice de ajuste ($-2 \log likelihood$) entre os modelos e a divisão desta diferença entre os graus de liberdade. A Tabela 4 apresenta os resultados de ajuste aos dados aos três modelos de TRI para os cinco subtestes de inteligência do IDS-2.

Tabela 4

Ajuste aos dados (-2 log likelihood) dos modelos de 1, 2 e 3 parâmetros de TRI para os cinco subtestes de inteligência do IDS-2

	REF	RFR	REH	COD	DIM
1-PL	3.993	4.289	12.044	10.311	17.063
2-PL	3.818	4.139	11.536	9.578	15.624
Diferença 1	175	150	508	733	1.439
GL	23	23	32	38	42
Dif. 1 / GL	7,61	6,52	15,88	19,29	34,26
2-PL	3.818	4.139	11.536	9.578	15.624
3-PL	3.895	4.220	11.588	11.474	15.679
Diferença 2	-77	-81	-52	-1.896	-55
GL	23	23	32	38	42
Dif. 2 / GL	-3,34	-3,52	-1,37	-49,9	-1,31

Notas: 1-PL = modelo de 1 parâmetro; 2-PL = modelo de 2 parâmetros; 3-PL = modelo de 3 parâmetros; GL = graus de liberdade; REF = Reconhecer Figuras; REH = Repetir História; COD = Colocar Discos; RFR = Reconhecer Figuras com Rotação; DIM = Descrever Imagem.

Para entender os resultados da Tabela 4 é necessário explicar que o *-2 log likelihood* indica a falta de ajuste, assim valores mais altos são evidência de menor ajuste do modelo aos dados. As diferenças na Tabela 4 do *-2 log likelihood* entre os modelos 1-PL e 2-PL são positivas para todos os cinco subtestes, significando que o modelo 2-PL se ajusta melhor aos dados do que o modelo 1-PL. Para saber se a diferença é estatisticamente significativa temos que dividir a diferença sobre os graus de liberdade. Esta divisão também é chamada de razão crítica: valores da razão crítica maior do que 1,96 indicam significância estatística de 5%. Uma vez que todos os valores da divisão Dif. 1/GL são maiores que 1,96 podemos afirmar que o modelo 2-PL se ajusta melhor aos dados do que o modelo 1-PL. As diferenças entre o *-2 log likelihood* dos modelos 2-PL e 3-PL são todas negativas indicando que o ajuste do modelo 2-PL é melhor do que o ajuste do modelo 3-PL. Para três dos cinco subtestes a diferença foi significativa. Resumindo,

podemos afirmar que os resultados da Tabela 4 mostram claramente que o modelo 2-PL tem o melhor ajuste aos dados. A Tabela 5 apresenta os resultados da estimação do parâmetro de dificuldade dos itens dos cinco subtestes de inteligência do IDS-2 segundo o modelo 2-PL da TRI.

Tabela 5

A dificuldade (parâmetro b) dos itens de cinco subtestes de inteligência do IDS-2 segundo o modelo 2-PL da TRI (N = 349)

Item	REF	RFR	REH1	REH2	COD1	COD2	COD3	DIM1	DIM2
1	-4,08	-5,18	-1,38	0,40	-2,92	-0,18	0,34	-1,26	-0,72
2	-2,89	-3,66	-0,65	0,70	-1,72	-0,29	0,44	-1,88	0,20
3	-2,61	-3,36	-1,99	0,79	-1,84	-0,15	0,53	0,40	1,59
4	-3,02	-3,10	-0,91	0,69	-1,57	0,19	0,76	-0,97	-0,11
5	-3,09	-3,28	-0,35	1,80	-1,73	0,21	0,83	-4,02	-5,43
6	-2,44	-2,44	-2,68	0,00	-1,83	0,84	0,89	-1,57	0,48
7	-2,45	-1,36	-0,13	0,64	-1,36	0,98	0,87	0,62	0,94
8	-1,55	-2,08	0,82	0,15	-1,58	0,93	0,79	1,21	1,66
9	-1,57	-1,19	0,85	-0,01	-1,48	1,33	0,85	1,01	1,17
10	-1,04	-1,29	-0,49	-0,12	-1,50	1,71	0,76	-1,22	-0,26
11	-1,59	-1,30	0,73	0,43	-1,32	1,76	0,85	-2,18	-1,71
12	-1,26	-1,18	0,40	0,40	-2,38	0,28	0,81	-0,38	0,00
13	-1,44	-0,90	-3,66	-0,17			0,93	0,10	0,31
14	-0,37	0,29	-2,35				1,02	-0,22	0,02
15	-0,09	0,13	-1,66					0,49	0,92
16	0,21	0,44	-0,35					-0,01	0,31
17	0,87	1,26	-1,26					0,23	0,42
18	0,97	1,23	-1,08					0,15	0,40
19	1,11	1,58	-1,27					0,20	0,23
20	1,54	1,83						-0,02	-0,02
21	1,66	1,84						0,86	1,28
22	2,00	2,21							
23	2,31	3,00							
M	-0,79	-0,72		-0,37		-0,08		-0,16	

Notas: REF = Reconhecer Figuras; RFR = Reconhecer Figuras com Rotação; REH1 = Repetir História parte 1 (itens 1 a 19); REH2 = Repetir História parte 2 (itens 20 a 32); COD1 = Colocar Discos parte 1 (itens 1 a 12); COD2 = Colocar Discos parte 2 (itens 13 a 24); COD3 = Colocar Discos parte 3 (itens 25 a 38); DIM1 = Descrever Imagem parte 1 (itens 1 a 21); DIM2 = Descrever Imagem parte 2 (itens 22 a 42).

Na Teoria de Resposta ao Item o parâmetro b indica a dificuldade dos itens, a qual está diretamente associada ao nível de habilidade dos sujeitos (θ). Dessa forma, quanto maior o valor de b maior será o nível de habilidade requerido para obter uma probabilidade de 50% de dar a resposta correta ao item. Os resultados da Tabela 5 indicam que entre os cinco subtestes avaliados, o subteste mais difícil foi o subteste Colocar Discos, com uma média do parâmetro b de -0,08 e que o subteste Reconhecer Figuras foi o mais fácil, com média do parâmetro b igual a -0,79.

Os subtestes Repetir História, Colocar Discos e Descrever Imagem foram divididos em duas ou três partes, pois os itens não são igualmente apresentados para todas as idades do instrumento. Nota-se que a primeira parte desses três subtestes apresentaram mais itens com um parâmetro b negativo, ou seja, itens relativamente fáceis. Esses resultados estão de acordo com os critérios de desenvolvimento dos subtestes, uma vez que os itens da primeira parte são apresentados para sujeitos com grande amplitude da faixa etária (5-20 anos).

De maneira geral, os itens com parâmetros b positivos encontram-se entre os últimos itens dos subtestes (no caso de Reconhecer Figuras e Reconhecer Figuras com Rotação) e nas partes 2 ou 3 dos demais subtestes (Repetir História, Colocar Discos e Descrever Imagem), o que indica uma tendência à organização dos itens de acordo com o nível crescente de dificuldade.

Em geral, podemos verificar nos resultados da Tabela 5 que os itens muito fáceis (parâmetro $b \leq -2$) e os itens muito difíceis (parâmetro $b \geq +2$) ocorrem principalmente nos subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação. Nestes dois subtestes tem sete itens com um parâmetro $b \leq -2$ e 2 itens com um parâmetro $b \geq +2$. Nos demais subtestes - Repetir História, Colocar Discos e Descrever Imagem - foram observados respectivamente 3, 2 e 3 itens muito fáceis e nenhum item muito difícil.

A discriminação dos itens (parâmetro a) também foi estimada pelo modelo 2-PL da TRI.

Os valores estimados do parâmetro a encontram-se na Tabela 6.

Tabela 6

Poder discriminativo (parâmetro a) dos itens de cinco subtestes do IDS-2 segundo o modelo 2-PL da TRI (N = 349)

Item	REF	RFR	REH1	REH2	COD1	COD2	COD3	DIM1	DIM2
1	0,88	0,67	0,63	1,31	0,56	1,39	5,37	0,17	0,18
2	1,18	1,06	0,41	1,19	0,62	1,56	4,33	0,28	0,38
3	2,04	1,77	0,61	1,46	0,85	1,59	4,49	0,44	0,58
4	1,91	1,03	0,52	1,11	0,49	1,31	2,40	0,22	0,21
5	1,36	0,99	0,85	0,80	0,81	1,12	2,37	0,40	0,27
6	1,29	1,15	0,41	2,02	0,73	0,81	2,26	0,25	0,30
7	0,75	0,57	0,74	0,98	0,62	0,81	2,06	0,33	0,33
8	0,79	0,47	0,37	1,29	0,94	0,81	2,16	0,24	0,32
9	1,16	1,00	0,29	2,31	0,99	0,74	2,34	0,27	0,29
10	0,67	0,72	0,23	1,90	0,91	0,67	2,63	0,16	0,17
11	1,34	1,52	0,33	1,45	0,59	0,58	2,96	0,30	0,27
12	1,85	1,40	0,59	1,57	0,64	5,04	2,54	3,28	1,56
13	1,12	1,16	0,36				2,45	1,51	1,36
14	0,80	1,02	0,35				2,31	2,42	1,43
15	0,96	0,99	0,44					1,71	1,26
16	2,28	1,64	0,84					1,67	1,42
17	1,77	1,42	0,45					1,77	1,94
18	2,47	2,52	0,40					1,26	1,17
19	3,88	2,39	0,49					1,38	1,37
20	2,61	2,17						1,63	1,76
21	2,42	2,97						0,88	0,91
22	1,95	1,94							
23	2,40	1,71							
M	1,65	1,40	0,88			1,73		0,91	

Notas: REF = Reconhecer Figuras; RFR = Reconhecer Figuras com Rotação; REH1 = Repetir História parte 1 (itens 1 a 19); REH2 = Repetir História parte 2 (itens 20 a 32); COD1 = Colocar Discos parte 1 (itens 1 a 12); COD2 = Colocar Discos parte 2 (itens 13 a 24); COD3 = Colocar Discos parte 3 (itens 25 a 38); DIM1 = Descrever Imagem parte 1 (itens 1 a 21); DIM2 = Descrever Imagem parte 2 (itens 22 a 42).

A discriminação de itens é um coeficiente que indica a eficácia com que o item discrimina entre pessoas que têm uma proficiência relativamente alta e aquelas que têm uma proficiência relativamente baixa (Crocker & Algina, 1986). Segundo Baker (2001) valores baixos do parâmetro a são valores $< 0,65$, valores moderados do parâmetro a variam de 0,65 a 1,34 e valores bons são valores $\geq 1,35$.

Observando os valores dos parâmetros dos itens na Tabela 6 podemos afirmar que 46 dos 158 itens apresentaram valores de discriminação $< 0,65$. A maioria dos itens com valores baixos de discriminação pertencem aos dois subtestes de memória de longo prazo: Repetir História (16 itens com parâmetro $a < 0,65$) e Descrever Imagem (22 itens com parâmetro $a < 0,65$). O subteste Colocar Discos tem 6 itens com valores baixos para o parâmetro de discriminação. Nos dois subtestes de memória viso-espacial, os subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação têm, respectivamente, 0 e 2 itens com valores baixos no parâmetro de discriminação.

Considerando o valor médio do parâmetro a podemos observar que o subteste Colocar Discos é o subteste com maior poder discriminativo ($M=1,73$), seguido pelos dois subtestes de memória viso-espacial Reconhecer Figuras e Reconhecer Figuras com Rotação ($M=1,65$; $M=1,40$). Os dois subtestes de memória de longo prazo, Descrever Imagem e Repetir História, têm entre os cinco subtestes avaliados o menor poder discriminativo ($M=0,91$; $M=0,88$).

Depois de obter as estimativas de dificuldade e de discriminação dos itens dos cinco subtestes de inteligência do IDS-2, tanto pela TCT como pela TRI, comparou-se os índices obtidos pelas duas teorias. A comparação foi realizada para cada um dos cinco subtestes por meio da correlação Spearman entre os índices obtidos pela TCT e TRI.

Tabela 7

Correlação de Spearman entre os índices de dificuldade e discriminação da TCT e TRI de cinco subtestes do IDS-2

Subteste	Índice de dificuldade	Índice de discriminação
	r	r
REF	0,97	0,38
RFR	0,97	0,31
REH	0,96	0,97
COD	0,96	0,92
DIM	0,97	0,97

Notas: REF = Reconhecer Figuras; REH = Repetir História; COD = Colocar Discos; RFR = Reconhecer Figuras com Rotação; DIM = Descrever Imagem.

A Tabela 7 mostra uma correlação média de 0,97 entre as estimativas de dificuldade da TCT e TRI, indicando que, apesar das diferenças entre os modelos de mensuração, os índices de dificuldade dos itens obtidos são altamente equiparados. Por outro lado, as estimativas de discriminação dos itens apresentaram uma correlação média de 0,71, o que indica que existem diferenças entre os índices de discriminação obtidos pela TCT e TRI. Essas diferenças são mais acentuadas para os subtestes Reconhecer Figuras ($r = 0,38$) e Reconhecer Figuras com Rotação ($r = 0,31$).

Os resultados da Tabela 7 são consistentes com os achados de estudos de simulação de Fan (1998) e MacDonald & Paunonen (2002). Os resultados do estudo de Fan indicam que os índices de dificuldade da TCT e TRI são altamente comparáveis, com correlações entre os índices na faixa de 0,91 a 0,99. As correlações entre os índices de discriminação, entretanto, ficaram bem mais diversificadas (entre 0,26 e 0,95). As correlações mais baixas foram observadas entre as dificuldades da TCT e o modelo 3-PL da TRI. O estudo de simulação de MacDonald e Paunonen apontou correlações entre os índices de

dificuldade baseados na TCT e na TRI na faixa de 0,93 a 0,99. Os autores acharam uma gama muito mais ampla de correlações entre os índices de discriminação baseados na TCT e na TRI, entre 0,24 e 0,96. Os autores afirmam que as estimativas de discriminação dos itens baseadas na TCT foram muito imprecisas nos casos de conjuntos de itens muito diversificados em relação à dificuldade. Assim, MacDonald & Paunonen recomendam considerar o índice de discriminação da TRI ao invés da TCT para os casos de testes com itens muito diversificados quanto à dificuldade.

Depois das análises dos itens foi iniciada a análise dos escores totais nos cinco subtestes de inteligência do IDS-2. Foram calculados as médias e os desvios-padrão dos escores brutos para os cinco subtestes de inteligência do IDS-2. As médias e os desvios-padrão dos escores brutos e do escore bruto total são apresentados na Tabela 8.

Os resultados na Tabela 8 são mostrados para cinco grupos de idade. O primeiro grupo é formado pelas idades de 5 e 6 anos ($n = 48$), onde 56,3% é do sexo feminino. O segundo grupo é formado pelas idades de 7 e 8 anos ($n = 50$), em que 58% é do sexo feminino. O terceiro grupo inclui sujeitos com idades de 9 a 12 anos ($n = 99$), sendo 53,5% do sexo masculino. O quarto grupo inclui participantes com idades entre 13 e 15 anos ($n = 62$), onde 51,6% é do sexo feminino, e o quinto grupo inclui sujeitos com idades entre 16 e 20 anos ($n = 90$), em que 50% é do sexo feminino.

As médias dos escores brutos dos cinco subtestes de inteligência apresentam uma tendência de crescimento de acordo com o aumento da idade. Entretanto, no subteste Descrever Imagens essa tendência de crescimento não se aplica ao último grupo de idade. A mesma tendência de aumento das médias dos escores brutos com a idade pode ser observada nas médias do escore bruto total. A média do escore bruto total está aumentando mais entre o segundo grupo de idade (7 aos 8 anos) e o terceiro grupo de idade (9 aos 12 anos) e está aumentando menos entre o quarto grupo de idade (13 aos 15

anos) e o quinto grupo de idade (16 aos 20 anos). Ressalta-se que o aumento do escore bruto total do primeiro grupo para o segundo grupo de idade é menor do que o esperado.

Em geral, os resultados da Tabela 8 fornecem a primeira evidência de validade dos cinco subtestes do IDS-2. Hogan (2018) afirma que a curva de crescimento dos escores brutos em testes de inteligência é bastante inclinada durante aproximadamente os primeiros 12 anos de vida e que a curva de crescimento é muito menos inclinada entre 12 e 20 anos.

A inspeção dos desvios-padrão do escore bruto total dos cinco subtestes sugere que, com a exceção do segundo grupo de idade, também existe uma tendência de aumento da variabilidade dos escores brutos com o crescimento da idade. No aspecto da variabilidade o segundo grupo é atípico, com exceção do subteste Descrever Imagem em que o desvio-padrão é menor do que no primeiro grupo. Aparentemente o grupo de idade dos 7 aos 8 anos é composto por crianças com habilidades cognitivas relativamente inferiores, o que também explica o aumento modesto no escore bruto total entre o primeiro e o segundo grupo de idade.

Tabela 8

Características psicométricas dos escores brutos dos sete subtestes de inteligência do IDS-2 para seis grupos de idade

	Idade em anos					
	5-6 (N = 48)	7-8 (N = 50)	9-12 (N = 99)	13-15 (N=62)	16-20 (N=90)	5-20 (N=349)
Subteste	M (DP)	M (DP)	M (DP)	M (DP)	M (DP)	M (DP)
REF	10,5 (3,9)	11,6 (3,3)	14,3 (2,3)	14,9 (2,7)	16,6 (2,9)	14,0 (3,5)
RFR	9,9 (3,4)	11,2 (2,7)	13,3 (2,4)	14,7 (2,5)	15,0 (3,0)	13,2 (3,3)
REH	9,2 (4,0)	11,1 (3,3)	19,2 (6,0)	20,5 (6,8)	20,7 (4,9)	17,3 (7,0)
COD	7,1 (3,1)	9,2 (2,8)	15,2 (4,0)	25,9 (6,2)	27,1 (6,0)	18,2 (9,2)
DIM	10,2 (3,6)	12,0 (3,7)	22,8 (6,3)	26,3 (5,5)	25,8 (6,1)	20,9 (8,3)
EBT	46,9 (12,0)	55,1 (9,4)	84,8 (14,1)	102,3 (14,8)	104,6 (14,3)	83,6 (25,7)
Δ EBT	----	8,2	29,7	17,5	2,3	----

Notas: REF = Reconhecer Figuras; RFR = Reconhecer Figuras com Rotação; REH = Repetir História; COD = Colocar Discos; DIM = Descrever Imagem; ADC = Assinalar Duas Características; RIF = Riscar Figuras; EBT = Escore bruto total dos primeiros cinco subtestes.

A fidedignidade dos escores obtidos nos subtestes do IDS-2 foi estimada por meio do método de consistência interna, a partir do coeficiente Alfa de Cronbach. A Tabela 9 apresenta os resultados das estimativas de fidedignidade dos escores dos cinco subtestes de inteligência do IDS-2 para cinco grupos de idade.

Tabela 9

Índice de fidedignidade (Alfa de Cronbach) dos escores nos cinco subtestes de inteligência do IDS-2 por grupo de idade

Subtestes	Idade					Média
	5-6 (N = 48)	7-8 (N = 50)	9-12 (N = 99)	13-15 (N=62)	16-20 (N=90)	
REF	0,85	0,82	0,81	0,80	0,84	0,82
RFR	0,79	0,75	0,74	0,82	0,82	0,78
REH	0,82	0,71	0,85	0,89	0,82	0,82
COD	0,73	0,78	0,84	0,86	0,90	0,82
DIM	0,66	0,70	0,79	0,74	0,78	0,73
Média	0,77	0,75	0,81	0,82	0,83	0,79
EBT	0,89	0,85	0,91	0,90	0,91	---

Notas: REF = Reconhecer Figuras; RFR = Reconhecer Figuras com Rotação; REH = Repetir História; COD = Colocar Discos; DIM = Descrever Imagem; EBT = Escore Bruto Total dos cinco subtestes de inteligência do IDS-2

A inspeção da Tabela 9 mostra que os coeficientes de fidedignidade são mais baixos no grupo de idade de 7-8 anos. Isso pode ser explicado pela baixa variância no grupo de idade de 7-8 anos. Hogan (2018) afirma que um dos fatores que impacta mais o coeficiente de fidedignidade é a variância dos escores. No grupo de idade de 16-20 anos os coeficientes de fidedignidade são mais altos.

Observando a última coluna da Tabela 9 podemos ver que, em média, os subtestes com maior coeficiente de fidedignidade são Reconhecer Figuras, Repetir História e Colocar Discos (Alfa = 0,82). O subteste com a menor média dos coeficientes de fidedignidade nos cinco grupos é Descrever Imagem (Alfa = 0,73).

Discussão

O objetivo do presente estudo foi analisar as propriedades psicométricas dos itens de cinco subtestes de inteligência do IDS-2 para uma amostra brasileira. O IDS-2 tem um total de 14 subtestes para a avaliação da inteligência, os quais avaliam 7 dimensões da inteligência do modelo Cattell-Horn-Carroll (CHC), conforme proposto por Schneider e McGrew (2018). No presente estudo foi realizada a análise dos itens de cinco dos 14 subtestes de inteligência do IDS-2, a saber: Reconhecer Figuras, Reconhecer Figuras com Rotação, Repetir História, Colocar Discos, e Descrever Imagem. A seleção desses subtestes se deu por meio da análise das tarefas e estímulos apresentados. Comparou-se as tarefas dos subtestes de inteligência do IDS-2 com as tarefas apresentadas pelos demais instrumentos de avaliação da inteligência disponíveis no Brasil, e que possuíam parecer favorável pelo SATEPSI (Sistema de Avaliação de Testes Psicológicos) até o ano de 2018, quando a coleta de dados iniciou. O principal critério para a escolha dos cinco subtestes do IDS-2 foi o aspecto inovador das tarefas, ou seja, foram escolhidas as tarefas que possuíam um formato diferente de avaliação da inteligência e suas facetas.

Os itens dos cinco subtestes foram analisados em relação aos parâmetros de dificuldade e discriminação de acordo com a Teoria Clássica dos Testes (TCT) e de acordo com a Teoria de Resposta ao Item (TRI). Segundo Urbina (2014) as análises quantitativas de dificuldade e discriminação dos itens têm a importante função de demonstrar seu funcionamento quando eles são utilizados junto à população alvo. O objetivo da análise de itens é obter índices relativos de dificuldade e discriminação para o grupo alvo (Urbina, 2014). Os modelos da TRI se destacam como uma evolução dos procedimentos adotados pela psicometria clássica, oferecendo recursos que superam os limites da TCT. Entre as vantagens da TRI está a possibilidade de se obter estimativas sem vieses amostrais, uma vez que na TRI é possível obter estimativas dos itens por meio

de amostras não representativas, o que não ocorre na TCT, cujas estimativas dependem da representatividade da amostra (Embretson & Reise, 2000; Thorpe & Favia, 2012).

Os resultados obtidos para o parâmetro de dificuldade dos itens dos subtestes de inteligência do IDS-2 foram semelhantes para a TCT e a TRI. A aproximação entre as estimativas de dificuldade dos itens na TCT e na TRI já foi apontada na literatura (Fan, 1998; MacDonald & Paunonen, 2002; Courville, 2004). Em ambos os modelos as estimativas de dificuldade indicaram que o subteste mais difícil foi Colocar Discos e que o subteste Reconhecer Figuras foi o mais fácil. A tarefa do subteste Colocar Discos envolve habilidades associadas ao processamento visual dos sujeitos, os quais precisam dispor corretamente o material de testagem como uma cópia dos estímulos apresentados. Esse subteste se caracteriza pelo formato de avaliação mais inovador em relação aos demais subtestes de inteligência do IDS-2. Além disso, essa tarefa também se diferencia em relação aos demais instrumentos disponíveis para a avaliação da inteligência no Brasil (CFP, 2021). Os índices maiores de dificuldade dos itens do subteste Colocar Discos podem estar associados à falta de familiaridade da amostra em relação ao tipo de tarefa requerida (Roazzi & Souza, 2002; Byrne et al., 2009). Outro fato que precisa ser levado em consideração em relação à dificuldade de Colocar Discos é que ele é o único subteste avaliado que visa medir processamento visual. Pode ser que seja mais fácil construir itens difíceis para medir processamento visual do que itens difíceis para medir memória visoespacial e memória de longo prazo.

Os subtestes são de uma dificuldade adequada. A dificuldade dos cinco subtestes, indicada pela média da proporção de erros, ficou entre 0,30 e 0,52. Uma proporção média de erros substancialmente acima de 0,50 seria uma indicação de um teste muito difícil. Em relação à distribuição dos índices se observou a tendência de aumento progressivo da dificuldade dos itens nos subtestes, tanto para as estimativas da TCT quanto da TRI.

Nesse sentido, se destaca a importância da variância da amplitude dos índices de dificuldade dos itens, indicando que os itens avaliam diferentes níveis de habilidade, o que é desejado para os subtestes de inteligência do IDS-2, uma vez que o instrumento é voltado para uma faixa etária grande (5-20 anos). Além disso, a variabilidade dos índices de dificuldade evita efeitos indesejados como os efeitos de teto e de chão (Urbina, 2014). A característica do aumento progressivo da dificuldade dos itens também é importante para o processo de administração do instrumento, sobretudo para os subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação, pois impacta diretamente nas regras de interrupção da aplicação, evitando a fadiga e o cansaço dos examinandos e tornando a testagem mais adaptativa.

A discriminação dos itens dos subtestes é boa. A média do parâmetro de discriminação da TRI ficou entre 0,88 e 1,65. De Ayala (2009) afirma que valores razoavelmente bons para o parâmetro de discriminação na TRI variam de 0,80 a 2,50, e Baker (2001) afirma que valores moderados do parâmetro a variam de 0,65 a 1,34, e valores bons são valores $\geq 1,35$. As estimativas de discriminação dos itens apresentaram discrepâncias entre os resultados obtidos por meio da TCT e da TRI. Entretanto, vale ressaltar que os dois modelos convergem ao apontar o subteste Colocar Discos, por exemplo, com maior média de discriminação dos itens. As divergências entre as estimativas do parâmetro de discriminação dos itens indicaram que, sobretudo para os subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação, foi fundamental utilizar o modelo de mensuração da TRI, uma vez que os índices obtidos entre os dois modelos não são comparáveis. Além disso, as estimativas de discriminação obtidas por meio da TRI apresentam maior acurácia quando comparadas às estimas da TCT (MacDonald & Paunonen, 2002).

A análise de consistência interna dos itens dos cinco subtestes utilizados indicou valores satisfatórios. Em geral, os índices de precisão para os grupos de idade apresentaram estimativas maiores ao que pode ser considerado como o valor mínimo aceitável, 0,70 (George & Malley, 2002). Os maiores índices foram obtidos para os sujeitos dos grupos de idade 4 e 5, os quais correspondem as idades dos 13 aos 20 anos. Destaca-se que as estimativas mais baixas de fidedignidade foram obtidas nos primeiros grupos de idade, e estão relacionadas à menor variância dos itens, uma vez que os sujeitos desses grupos atingem o critério de interrupção do teste antes dos demais.

Conclusão

Os subtestes de inteligência do IDS-2 utilizados no presente estudo representam um avanço para o contexto da avaliação psicológica no Brasil, destacando-se o caráter inovador das tarefas propostas, que se diferenciam dos testes tradicionais de inteligência atualmente disponíveis no cenário brasileiro, o que também permite avanços nas pesquisas sobre as habilidades cognitivas da população de crianças e jovens brasileiros. O IDS-2 apresenta uma proposta de avaliação mais ampla das habilidades cognitivas associadas à inteligência. Esse tipo de ferramenta viabiliza uma avaliação mais personalizada, com foco nas forças e fraquezas dos sujeitos, além do monitoramento das mudanças que acompanham o desenvolvimento e os processos de intervenção na vida das pessoas dentro da faixa etária dos 5 aos 20 anos. Nesse sentido se destaca o fato de que as médias dos escores brutos dos cinco subtestes de inteligência apresentaram uma tendência de crescimento de acordo com o aumento da idade, o que pode ser considerada uma evidência da validade dos cinco subtestes (Hogan, 2018).

O tamanho e composição amostral podem ser considerados limitações importantes do estudo, algo que reduziu a variabilidade das respostas aos itens e limitou o quantitativo de informação obtida por meio da testagem. Em geral, as análises de dificuldade dos itens

dos quatro subtestes investigados indicaram a insuficiência de itens para avaliar a faixa intermediária de habilidade dos sujeitos, o que precisa ser investigado em novas amostras representativas da população. Ademais, o estudo identificou a similitude entre as estimativas dos modelos de medida da Teoria Clássica dos Testes e da Teoria de Resposta ao Item em relação às estimativas de dificuldade dos itens, o que pode justificar o uso da TCT na construção de medidas na atualidade, apesar das suas limitações. Por último ressalta-se a importância em avaliar as estimativas de discriminação dos itens por meio dos modelos da TRI, os quais se mostram, reiteradamente, mais adequados para o processo de avaliação da qualidade dos itens dos instrumentos.

Referências

- American Educational Research Association [AERA]; American Psychological Association [APA]; & National Council on Measurement in Education [NCME] (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Andrade, J. M., Laros, J. A., & Gouveia, V. V. (2010). O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Avaliação Psicológica*, 9(3), 421-435.
- Andrade, J. M., & Valentini, F. (2018). Diretrizes para a construção de testes psicológicos: a Resolução CFP nº 009/2018 em destaque. *Psicologia: Ciência e Profissão*, 38, 28-39. <https://dx.doi.org/10.1590/1982-3703000208890>.
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). Eric Clearinghouse on Assessment and Evaluation.
- Baker, F. B., & Kim, S. (2017). *The basics of item response theory using R*. Springer.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397 – 479). Addison-Wesley.
- Byrne et al. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved educational and training in psychology. *American Psychological Association*, 39(2), 94-105.
- Conde, F. N., & Laros, J. A. (2007) Unidimensionalidade e a propriedade de invariância das estimativas da habilidade pela TRI. *Avaliação Psicológica*, 6(2), 205-215.
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich College Publishers.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 197-334.
- Courville, T. G. (2004). *An empirical comparison of item response theory and classical test theory item/person statistics* [Unpublished doctoral dissertation]. Louisiana State University.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of life Research*, 16, 15-18. <https://doi.org/10.1007/s11136-007-9198-0>
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers.

- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/ person statistics. *Educational and Psychological Measurement*, 58(3), 357-381.
- George, D., & Malley, P. (2002). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.) Allyn & Bacon.
- Grob, A., & Hagemann-von Arx, P. (2018). *Intelligenz- und Entwicklungsskalen für Kinder und Jugendliche* [Intelligence and Development Scales - 2 (IDS-2)]. Hogrefe.
- Hogan T. P. (2018). *Psychological testing: A practical introduction* (4th ed.). Wiley.
- Laros, J. A., Tellegen, P. J., Jesus G. R., & Karino, C. A. (2015). *SON-R 2½-7[a] – Teste não-verbal de inteligência. Manual*. Hogrefe-CETEPP.
- Laros, J. A., Tellegen, P. J., Lima, R. M. F., & Valentini, F. (2021). *SON-R 6-40. Teste não-verbal de inteligência. I. Relatório científico*. Hogrefe-CETEPP.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- MacDonald, P., & Paunonen, S. V. (2002). A Monte Carlo comparison of item and person statistics based on Item Response Theory versus Classical Test Theory. *Educational and Psychological Measurement*, 62(6), 921-943. <https://doi.org/10.1177/0013164402238082>
- Resolução N° 009, de 25 de abril de 2018*. Estabelece diretrizes para a realização de Avaliação Psicológica no exercício profissional da psicóloga e do psicólogo, regulamenta o Sistema de Avaliação de Testes Psicológicos - SATEPSI e revoga as Resoluções n° 002/2003, n° 006/2004 e n° 005/2012 e Notas Técnicas n° 01/2017 e 02/2017. Brasília, DF: Conselho Federal de Psicologia.
- Roazzi, A. & Souza, B. C. (2002). Repensando a inteligência. *Paidéia*, 12(23), 31-55. <https://doi.org/10.1590/S0103-863X2002000200004>.
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (4th ed., pp. 71-163). The Guilford Press.
- Thompson, B. (2003). *Score reliability: Contemporary thinking on reliability issues*. SAGE Publications.
- Thorpe, G. L., & Favia, A. (2012). Data analysis using item response theory methodology: An introduction to selected programs and applications. *Psychology Faculty Scholarship*, 20. https://digitalcommons.library.umaine.edu/psy_facpub/20.
- Urbina, S. (2014). *Essentials of Psychological Testing* (2nd ed.). Wiley.
- Verhelst, N. D. (2014). Test theory: Some basic notions. *Educational and Science*, 39(172), 3-19.

- Zanon, C., & Filho, N. H. (2015). Fidedignidade. In C. S. Hutz, D. R. Bandeira & C. M. Trentini (Eds.), *Psicometria* (pp. 85-95). Artmed.
- Zimowski, M.F., Muraki, E., Mislevy, R.J. & Bock, R.D. (1996). *BILOG-MG: multiple-group IRT analysis and test maintenance for binary items*. Scientific Software International (SSI).

Manuscrito 3

Estrutura fatorial do conjunto de subtestes de inteligência no IDS-2

Título em inglês

Factor structure of the set of intelligence subtests in the IDS-2

Título abreviado

Estrutura fatorial do IDS-2

Resumo

A validade é uma propriedade essencial para o campo da avaliação e da testagem psicológica, pois está diretamente associada às interpretações realizadas com base nos resultados dos testes. Nesse contexto, o presente estudo objetiva obter evidências de validade dos escores de um conjunto de subtestes de inteligência do IDS-2. A amostra foi composta por 349 participantes, com idades variando dos 5 aos 20 anos ($M = 11,9$, $DP = 4,6$), sendo maioria do sexo feminino (51,3%). Foram utilizados sete subtestes do IDS-2: Colocar Discos, Repetir História, Descrever Imagem, Assinalar Duas Características, Riscar Figuras, Reconhecer Figuras e Reconhecer Figuras com Rotação. A partir do cálculo dos escores normatizados analisou-se a estrutura fatorial do instrumento por meio da Análise Fatorial Exploratória (AFE) e da Análise Fatorial Confirmatória (AFC). Os resultados indicaram um modelo hierárquico, com três fatores de primeira ordem e um fator de segunda ordem. Os fatores de primeira ordem foram interpretados como: fator de Memória Viso-espacial de Curto Prazo, fator de Memória de Longo Prazo e fator de Velocidade de Processamento, enquanto o fator de segunda ordem foi interpretado como um fator Geral de Velocidade e Memória. Os resultados são plausíveis com os pressupostos teóricos do modelo CHC de inteligência e fornecem indicadores iniciais da validade dos escores nos sete subtestes de inteligência do IDS-2 para a avaliação de amostras brasileiras.

Palavras-chave: *IDS-2; validade de construto; teste de inteligência.*

Abstract

Validity is an essential property in the area of psychological testing and assessment, being directly associated with the interpretations of the test results. In this context, the present study aimed to obtain evidence of the validity of the scores from a set of IDS-2 intelligence subtests. The sample consisted of 349 participants, with ages ranging from 5 to 20 years ($M = 11.9$, $SD = 4.6$), with the majority being female (51.3%). Seven IDS-2 subtests were administered: Placing Disks, Retelling a Story, Describing a Picture, Crossing out two Features, Crossing out Boxes, Recognizing Figures and Recognizing Rotated Figures. The factor structure of the set of seven intelligence subtests was analyzed using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The factor structure was analyzed using standardized scores. The results indicated a hierarchical model, with three first-order factors and one second-order factor. The first-order factors were interpreted as: Short-term Visuospatial Memory, Long-term Memory, and Processing Speed; the second-order factor was interpreted as a General Speed and Memory factor. The results are in accordance with the theoretical assumptions of the CHC-model of intelligence that served as a basis for the construction of the IDS-2 and provide initial evidence of the validity of the scores of the seven intelligence subtests of the IDS-2 for the Brazilian context.

Keywords: *IDS-2; construct validity; intelligence test.*

Em sua maioria os modelos de inteligência se diferem em relação às proposições estruturais e funcionais. Nesse contexto se destaca a interdependência entre os processos de mensuração e aprimoramento de métodos e técnicas de investigação científica e o avanço dos modelos teóricos acerca da inteligência humana. O cenário atual desse construto apresenta a coexistência de diferentes modelos teóricos e estruturais, e há relativo consenso sobre a estrutura fatorial da inteligência. A natureza múltipla das dimensões da inteligência e a mensuração dessas dimensões é reconhecida, e ao mesmo tempo discutida entre aqueles que questionam a validade das medidas frente ao processo de tomada de decisão na vida dos sujeitos (Canivez & Youngstrom, 2019; Dombrowski, et al., 2018).

Atualmente se destaca o modelo Cattell-Horn-Carroll (CHC) das habilidades cognitivas (Schneider & McGrew, 2018). Trata-se de uma integração funcional da teoria de inteligência, que apresenta uma estrutura hierárquica e multidimensional. Este modelo utiliza o acrônimo “CHC” como uma referência à ordem de contribuições dos seus principais autores: Raymond Cattell, John Horn e John Carroll. O modelo CHC foi proposto por McGrew em 1997, e organiza a inteligência em um modelo integrado que traz a junção entre a teoria da inteligência fluida (Gf) e inteligência cristalizada (Gc) de Cattell e Horn, e a teoria dos três estratos de Carroll (McGrew, 1997). O modelo CHC traz uma nova organização ao agregar novos fatores amplos ao segundo nível hierárquico e indicar novos padrões funcionais entre os fatores pré-existentes.

As habilidades amplas do modelo CHC estão categorizadas conceitualmente, e estão relacionadas à capacidade de raciocínio geral, conhecimento adquirido, habilidades sensoriais e psicomotoras, habilidades de memória e velocidade. A proposta atual do modelo apresenta 18 habilidades cognitivas, sendo elas: Raciocínio Fluido (Gf); Memória de Trabalho (Gwm); Eficiência de Aprendizagem (Gl); Fluência de Recuperação (Gr);

Velocidade de Processamento (Gs); Velocidade de Reação e Decisão (Gt); Velocidade Psicomotora (Gps); Conhecimento Compreensivo (Gc); Conhecimento de Domínio Específico (Gkn); Leitura e Escrita (Grw); Conhecimento Quantitativo (Gq); Processamento Visual (Gv); Processamento Auditivo (Ga); Habilidades Olfativas (Go); Habilidades Táteis (Gh); Habilidades Cinestésicas (Gk) e Habilidades Psicomotoras (Gp); (Schneider & McGrew, 2018).

A sistemática hierárquica proposta por essa teoria indica que os fatores de inteligência estão integrados funcionalmente, sendo alguns domínios ou habilidades mais abrangentes do que outros. De acordo com o modelo os fatores do estrato intermediário reúnem um conjunto de habilidades amplas e que surgem a partir da categorização de habilidades mais específicas, presentes no estrato inferior do modelo, as quais são avaliadas por meio das tarefas em testes de inteligência. Apesar da complexidade da teoria CHC, há flexibilidade e relativa permeabilidade para sua atualização, desde que observado um conjunto de critérios, dentre eles, a definição clara dos fatores e mensuração por meio do desempenho em testes, evidências de validade convergente e discriminante, validade incremental, correlação com funções cerebrais e relevância do novo fator para a vida dos sujeitos (Schneider & McGrew, 2018).

A proposição do modelo CHC representa um avanço significativo para o campo de estudo da inteligência. O modelo teve impacto sobre a construção e revisão dos instrumentos de avaliação da inteligência, e nos estudos que objetivam compreender o funcionamento cognitivo (Schelini, 2006; Wechsler & Nakano, 2016). Além disso, a utilização do modelo CHC permite a universalização dos termos empregados, viabilizando uma comunicação mais eficaz entre pesquisadores e profissionais. Nesse sentido, se destaca a relevância do modelo CHC para o uso clínico, militar,

organizacional e na avaliação neuropsicológica (Forgues, 2014; Schneider & Newman, 2015; Stough & Pase, 2015; Jewsbury & Bowden, 2016).

Quanto à instrumentalização do modelo CHC se destaca o uso das baterias de inteligência, as quais permitem avaliar uma abrangência maior de habilidades cognitivas. Entre as baterias de avaliação da inteligência se destaca o IDS-2 (*Intelligence and Development Scales*), que também avalia funções do desenvolvimento em sujeitos com idades entre 5 e 20 anos (Grob & Haggmann-von Arx, 2018). A avaliação da inteligência no IDS-2 é composta por 14 subtestes, os quais avaliam sete fatores de inteligência presentes no modelo CHC, sendo eles: Processamento Visual (PV); Memória de Longo prazo (ML); Velocidade de Processamento (VP); Memória Fonológica de curto prazo (MF); Memória Viso-espacial de curto prazo (MV); Raciocínio Abstrato (RA) e Raciocínio Verbal (RV).

O IDS-2 se encontra em fase de estudos para adaptar e obter evidências de validade dos escores dos subtestes junto à população brasileira. A validade dos escores dos instrumentos pode ser categorizada de acordo com as fontes de evidências. De acordo com os *Standards for Educational and Psychological Testing*, as principais fontes de validade são: 1) Evidências com fundamento no conteúdo do teste; 2) Evidências com base nas relações entre os escores do teste e outras variáveis; 3) Consequências da testagem; 4) Processos de resposta; e 5) Evidências com base na estrutura interna do teste (*American Educational Research Association [AERA]; American Psychological Association [APA]; National Council on Measurement in Education [NCME]*, 2014).

As evidências de validade com base na estrutura interna do teste, as quais são finalidade do presente estudo, indicam que o conjunto de itens estão relacionados entre si. No Brasil, o Conselho Federal de Psicologia – CFP, por meio da Resolução nº. 09 de 2018 destaca que entre os critérios de avaliação da qualidade dos estudos de validade

com base na estrutura interna do teste se destacam a utilização das técnicas de Análise Fatorial Exploratória (AFE) e Análise Fatorial Confirmatória (AFC). Esses estudos dão suporte à adequação das inferências realizadas com base nos escores da testagem (Urbina, 2014). Nesse sentido, o presente estudo tem o objetivo geral de: obter evidências de validade dos escores de um conjunto de subtestes de inteligência do IDS-2, considerando o contexto de adaptação do instrumento para a população brasileira. Os objetivos específicos são: 1) adquirir estimativas do desempenho de uma amostra brasileira por meio do cálculo dos escores normatizados; e 2) obter estimativas de precisão dos escores por meio dos métodos da confiabilidade composta (CC) e da variância média extraída (VME).

Método

Participantes

A amostra foi composta por 349 participantes. As idades dos participantes variaram dos 5 aos 20 anos ($M = 11,9$, $DP = 4,6$), com maioria do sexo feminino (51,3%). Os sujeitos eram em sua maioria estudantes (97,7%), vindos de seis estados brasileiros e do Distrito Federal (*para descrição detalhada dos participantes ver manuscrito 2, p. 54*).

Instrumentos

Foram utilizados sete subtestes do IDS-2 (Grob & Hagmann-von Arx, 2018). Os sete subtestes se referem à avaliação da inteligência e estão associados aos seguintes fatores da teoria de CHC (Schneider & McGrew, 2012): velocidade de processamento (VP), processamento visual (PV), memória viso-espacial de curto prazo (MV) e memória visual de longo prazo (ML).

O subteste de processamento visual que foi incluído neste estudo é Colocar Discos, no qual os sujeitos utilizam discos metálicos sobre uma base magnética com uma folha

branca para copiar a disposição dos círculos apresentados. Ao final da tarefa o examinando usa um lápis para circular os limites dos círculos indicados, e para a correção o avaliador utiliza um crivo para determinar a relação de sobreposição entre o desenho do sujeito e as guias do crivo de correção. Na pesquisa de normatização conjunta da Alemanha e Suécia o subteste Colocar Discos apresentou um Alfa de Cronbach de 0,95 (Grob & Hagmann-von Arx, 2018).

O subteste Repetir História foi desenvolvido para a avaliação do fator memória de longo prazo, de modo mais específico a memória auditiva ou fonológica. O sujeito escuta e repete uma história após o tempo mínimo de 20 minutos. O subteste Descrever Imagem também avalia a memória de longo prazo, entretanto se refere à memória visual. Neste subteste é apresentado um estímulo visual para o examinando, o qual deverá ser recordado posteriormente. Os escores dos dois subtestes de memória de longo prazo obtiveram um Alpha de Cronbach de 0,93 e 0,92, respectivamente (Grob & Hagmann-von Arx, 2018).

O fator velocidade de processamento foi representado por dois subtestes: Assinalar Duas Características e Riscar Figuras. Ambos os subtestes envolvem tarefas onde o sujeito deverá identificar e riscar figuras com características específicas em meio a figuras semelhantes. Por se tratar de subtestes de velocidade recorre-se ainda à indicação de limite de tempo para o desenvolvimento da tarefa. O Alfa de Cronbach dos dois subtestes foi igual a 0,97 e 0,96, respectivamente. O escore bruto total nos dois subtestes é determinado pelo somatório de figuras corretamente riscadas.

Os dois últimos subtestes incluídos no atual estudo foram desenvolvidos para a avaliação do fator memória viso-espacial de curto prazo. Em ambos os casos a tarefa consiste em memorizar estímulos apresentados e identificá-los entre outros estímulos imediatamente após sua apresentação. A diferença entre os dois subtestes reside no fato

de que para o subteste Reconhecer Figuras com Rotação o sujeito deverá identificar a mesma figura rotacionada. Para os dois subtestes a pesquisa de normatização Alemanha/Suécia indicou um Alfa de Cronbach de 0,87 (Grob & Hagmann-von Arx, 2018).

Procedimentos

A administração dos sete subtestes ocorreu de forma individual. Os aplicadores entraram em contato com os participantes e com os pais e/ou responsáveis a fim de apresentar o estudo e obter o consentimento para participação. Os participantes foram selecionados com base na idade e as aplicações ocorreram nas escolas e nas casas dos participantes em horários anteriormente acordados.

Análise de dados

Inicialmente os escores brutos dos subtestes e o escore bruto total foram transformados em escores normatizados. O modelo utilizado para a normatização dos escores foi a normatização contínua (*continuous test norming*) (Lenhard et al., 2018; Oosterhuis, 2007; Voncken et al., 2017; Tellegen & Laros, 2014). Este modelo utiliza regressões para modelar a transformação das distribuições de escores brutos para todos os grupos simultaneamente em função da idade dos sujeitos (Laros et al., 2021). Uma grande vantagem de normatização contínua é que o tamanho necessário para a amostra total e dos subgrupos é menor do que usando o modelo tradicional de normatização (Lenhard et al., 2019).

Para o cálculo das normas foi utilizada a idade exata dos participantes, obtida por meio da seguinte fórmula: idade exata = (ano da aplicação – ano do nascimento) + (mês de aplicação – mês do nascimento) ÷ 12 + (dia da aplicação – dia do nascimento) ÷ 365,25. Na sequência, a variável idade exata foi transformada em variável *y* com valores

entre -1 e + 1, onde $y = (\text{idade exata} - 13) \div 6$. A idade exata foi transformada em uma variável y com valores entre -1 e + 1 para evitar valores extensos, uma vez que nas análises de regressão a variável y é elevada à segunda e terceira potências, devido a relação curvilínea entre a idade e os escores brutos do teste (Kittrell, 1999). Como próximo passo foi realizada para cada subteste uma análise multivariada de regressão com variáveis independentes y , y^2 e y^3 e com variável dependente o escore total no subteste. Em seguida foi calculado o escore z para cada subteste, onde, $z = (\text{escore bruto do subteste} - (\text{constante} + \text{coeficiente de regressão} \times y - \text{coeficiente de regressão} \times y^2 + \text{coeficiente de regressão} \times y^3)) / (\sqrt{\text{variância}})$ (Laros et al., 2021). Finalmente os escores z dos subtestes foram transformados em escores da escala Wechsler, cuja média é 10 e o desvio padrão é 3. O escore total normatizado do teste foi obtido através do cálculo das covariâncias entre os subtestes, o que viabilizou o cálculo do escore total normatizado z , o qual foi transformado para a escala de QI, onde a média é 100 e desvio padrão é 15.

A Análise Fatorial Exploratória (AFE) foi conduzida por meio do programa FACTOR (versão 11.02.04 64 bits; Ferrando & Lorenzo-Seva, 2017). A matriz analisada foi a matriz de correlações policóricas. Para a extração dos fatores foi utilizado *Minimum Rank Factor Analysis* – MRFA (Ten Berge & Kiers, 1991). A vantagem principal do método MRFA é que ele permite o cálculo da porcentagem da variância comum explicada. Segundo Ferrando & Lorenzo (2018) a porcentagem da variância comum explicada pode ser considerada um critério auxiliar útil para julgar a adequação do modelo. Após a análise das correlações entre os fatores, foi realizada uma análise fatorial de segunda ordem, utilizando o método Schmid-Leiman (Thompson, 2004).

Utilizou-se o software Amos 18.0 (Arbuckle, 2009) do pacote estatístico SPSS para o desenho dos modelos e para a análise fatorial confirmatória. Foram seguidas as recomendações de Byrne (2010) e Thompson (2004). A Confiabilidade Composta (CC) e

Variância Média Extraída (VME) foram calculadas com base nos parâmetros estimados no programa Amos 18 seguindo as recomendações de Valentini & Damásio (2016).

Resultados

Os escores brutos dos sete subtestes de inteligência do IDS-2 e do escore total foram transformados em escores normatizados. Na Tabela 1 são apresentados os escores normatizados dos sete subtestes na escala Wechsler (M = 10; DP = 3) e o escore total normatizado na escala de QI (M = 100; DP = 15).

Tabela 1

Média e desvio-padrão dos escores normatizados dos sete subtestes de inteligência do IDS-2 e do escore total para três grupos de idade

Subteste	Idade em anos					
	5-20 (N = 349)		5-12 (N = 197)		13-20 (N = 152)	
	M	(DP)	M	(DP)	M	(DP)
Reconhecer Figuras	9,95	(2,82)	10,03	(2,24)	9,85	(3,44)
Repetir História	9,96	(2,94)	9,99	(2,67)	9,91	(3,26)
Colocar Discos	9,98	(2,90)	9,73	(2,63)	10,30	(3,19)
Reconhecer Figuras com Rotação	9,98	(2,88)	9,94	(2,79)	10,03	(3,00)
Descrever Imagem	10,00	(3,01)	9,95	(3,05)	10,07	(2,96)
Assinalar Duas Características	10,00	(2,97)	9,88	(2,84)	10,15	(3,13)
Riscar Figuras	9,97	(2,84)	9,84	(2,75)	10,14	(2,96)
Média	9,98	(2,91)	9,91	(2,71)	10,06	(3,13)
Escore total	99,97	(15,01)	99,37	(13,42)	101,10	(16,84)

Os resultados da Tabela 1 indicam que no grupo total de 5 a 20 anos as médias e os desvios-padrão dos escores normatizados dos subtestes têm valores que diferem minimamente dos valores da escala Wechsler de normatização (M=10; DP=3). A maior diferença pode ser observada no caso do subteste Reconhecer Figuras, onde a diferença da média é de 0,05 pontos e a diferença do desvio-padrão é de 0,18 pontos. É importante ressaltar que antes do arredondamento para números inteiros da escala Wechsler as médias e desvios-padrão dos escores normatizados no grupo total estavam exatamente 10 e 3 respectivamente. A média e o desvio-padrão do escore total no grupo de 5 a 20 anos se afastam apenas 0,03 e 0,01 pontos da escala de QI (M=100; DP=15). No grupo de idade de 5 aos 12 anos as diferenças absolutas entre as médias dos escores normatizados e a média de 10 da escala Wechsler variaram entre 0,01 e 0,27 e no grupo de idade de 13 aos 20 as diferenças variam entre 0,15 e 0,30. As diferenças absolutas entre os desvios-padrão dos escores normatizados e o desvio-padrão da escala Wechsler variaram no grupo de idade de 5 aos 12 anos entre 0,05 e 0,76 e no grupo de idade de 13 a 20 anos entre 0,00 e 0,44. O que chama a atenção é que os desvios-padrão dos escores normatizados no grupo de idade de 13 a 20 anos são mais altos do que no grupo de idade de 5 a 12 anos. Em geral podemos concluir que os escores normatizados dos sete subtestes do IDS-2 ficam com as médias e desvios-padrão bem perto da média e desvio-padrão da escala Wechsler (M = 10; DP=3) tanto no grupo geral como nos grupos de 5 a 12 anos e de 13 a 20 anos.

Na Tabela 2 são apresentadas as correlações dos escores brutos e dos escores normatizados dos sete subtestes do IDS-2 com a idade exata dos participantes.

Tabela 2

Correlações entre os escores brutos e escores normatizados dos subtestes de inteligência do IDS-2 com a idade exata dos sujeitos (N = 349).

Subteste	r EBT - idade	r ENT - idade
Reconhecer Figuras	0,51	-0,01
Repetir História	0,53	0,01
Colocar Discos	0,76	-0,00
Reconhecer Figuras com Rotação	0,51	0,01
Descrever Imagem	0,63	-0,02
Assinalar Duas Características	0,74	-0,02
Riscar Figuras	0,65	0,02
Escore Total	0,78	0,01

Notas: EBT = escore bruto total; ENT = escore normatizado total.

Os resultados da Tabela 2 mostram claramente correlações substanciais (de 0,51 a 0,76) entre os escores brutos dos sete subteste de IDS-2 e a idade dos participantes. Assim, uma parcela substancial da variância dos escores brutos dos subtestes é associada à idade dos sujeitos. Esta parcela de variância é variância confiável e estável, mas não se trata da variância que queremos medir, em outras palavras, não é variância pretendida. No fim das contas com um teste de inteligência não se pretende medir a idade de uma pessoa. Os escores brutos no teste não podem ser interpretados por conta da contaminação da variável idade. Por isso os escores brutos precisam ser transformados em escores normatizados, quando a variável idade é eliminada. A Tabela 2 mostra que nos escores normatizados a variável idade foi eliminada com sucesso: praticamente não existe mais relação entre os escores normatizados dos subtestes e a idade dos sujeitos (correlações entre -0,02 e +0,02). As pesquisas sobre a estrutura fatorial de um teste de inteligência deveriam ser realizadas utilizando os escores normatizados e não os escores brutos (Tellegen & Laros, 2014).

A análise fatorial exploratória dos sete subtestes de inteligência das Escalas IDS-2 foi realizada com os escores normatizados dos sete subtestes do IDS-2. O índice H foi obtido para avaliar a estabilidade dos fatores, sendo que valores de H maiores que 0,80 sugerem uma variável latente com maior probabilidade de ser estável entre os estudos (Ferrando & Lorenzo-Seva, 2018). Para avaliar a qualidade das estimativas dos escores fatoriais foi obtido o *Factor Determinacy Index* (FDI). Valores do FDI acima de 0,90 são recomendados (Ferrando & Lorenzo-Seva, 2018). Para avaliar o ajuste da solução fatorial aos dados foram obtidos o *Goodness of Fit Index* (GFI) e o *Root Mean Square of Residuals* (RMSR). A obtenção desses índices de ajuste da solução fatorial implica que no FACTOR é possível optar para trabalhar com o modelo de análise fatorial semi confirmatória (*Semi Confirmatory Factor Analysis* - Ferrando & Lorenzo-Seva, 2017). Além disso, a fidedignidade dos fatores foi estimada usando o coeficiente Alfa de Cronbach.

Tabela 3

Cargas fatoriais padrão e estrutura e comunalidades usando MRFA e Rotação Promin dos escores normalizados de sete subtestes de inteligência do IDS=2

Subteste	CFP			CFE			h^2
	F1	F2	F3	F1	F2	F3	
ADC	0,88	-0,03	-0,11	0,79	0,45	0,41	0,64
RIF	0,73	-0,02	0,01	0,72	0,42	0,44	0,52
REH	-0,09	0,86	0,00	0,43	0,81	0,31	0,66
DIM	0,17	0,45	0,01	0,45	0,56	0,30	0,32
RFR	0,00	0,00	0,87	0,53	0,37	0,87	0,76
REF	0,05	0,07	0,60	0,46	0,35	0,66	0,45
COD	0,29	0,15	-0,01	0,38	0,33	0,23	0,16
VCE	37,9%	27,2%	30,1%	Σ VCE = 95,2%			
VTE	27,2%	19,7%	21,6%	Σ VTE = 68,5%			
Índice H	0,80	0,74	0,80				
FDI	0,90	0,86	0,91				
Alfa de Cronbach	0,88	0,93	0,90				
GFI	0,99						
RMSR	0,02 (critério para um modelo aceitável < 0,05)						
Correlações entre fatores	F1	F2	F3				
F1	1						
F2	0,61	1					
F3	0,61	0,43	1				

Notas: CFP = carga fatorial padrão; CFE = carga fatorial estrutura; F1 = velocidade de processamento; F2 = memória de longo prazo; F3 = memória viso-espacial; h^2 = comunalidade; ADC = Assinalar Duas Características; RIF = Riscar Figuras; REH = Repetir História; DIM = Descrever Imagem; RFR = Reconhecer Figuras com Rotação; REF = Reconhecer Figuras; COD = Colocar Discos; VCE = Variância Comum Explicada; VTE = Variância Total Explicada; Índice H = Replicabilidade de construto; FDI = *Factor Determinacy Index*; GFI = *Goodness of Fit Index*; RMSR = *Root Mean Square of Residuals*.

Foi obtida uma solução fatorial com três fatores, que somados explicam 68,5% da variância total e 95,2% da variância comum. Os três fatores obtidos apresentam cargas fatoriais padrão altas nos subtestes que teoricamente pertencem ao fator, todas acima de 0,40 (Costello & Osbourne, 2005). O primeiro fator - Velocidade de Processamento - é

composto pelos subtestes Assinalar Duas Características (ADC) e Riscar Figuras (RIF), o segundo fator – Memória de Longo prazo - é composto por Repetir História (REH) e Descrever Imagem (DIM), e o terceiro fator – Memória Viso-Espacial de Curto Prazo - é formado pelos subtestes Reconhecer Figuras (REF) e Reconhecer Figuras com Rotação (RFR). O subteste Colocar Discos (COD) não apresentou cargas fatoriais significativas em nenhum dos três fatores obtidos.

As comunalidades mostradas na Tabela 3 foram calculadas a partir das cargas fatoriais padrão e cargas fatoriais estrutura segundo as recomendações de Thompson (2004). Os coeficientes de comunalidade indicam a proporção de variância de uma variável mensurada que os fatores, como um conjunto, podem reproduzir. Entretanto, os coeficientes de comunalidade também podem ser interpretados como o quanto da variância de uma variável medida foi útil para delinear os fatores extraídos (Thompson, 2004). Assim, podemos afirmar que o subteste Reconhecer Figuras com Rotação contribui mais na solução fatorial e o subteste Colocar Discos contribui menos.

A solução fatorial com três fatores conforme apresentada na Tabela 3 fornece uma forte evidência de validade baseada na estrutura interna dos sete subtestes de inteligência do IDS-2. O fato do subteste Colocar Discos ter uma comunalidade muito baixa é uma evidência de que este subteste está medindo um quarto e outro construto para além dos três fatores identificados.

De maneira geral os índices fatoriais relacionados com a qualidade da solução fatorial foram satisfatórios. O índice de replicabilidade dos fatores, índice H, apresentou o valor de 0,80 para o fator de Velocidade de Processamento (Assinalar Duas Características e Riscar Figuras), e para o fator de Memória Viso-Espacial de Curto Prazo (Reconhecer Figuras e Reconhecer Figuras com Rotação). O valor mais baixo do índice H (0,74) foi obtido para o fator Memória de Longo Prazo (Repetir História e Descrever

Imagem). O índice FDI (*Factor Determinacy Index*) apresentou valores dentro do esperado para os fatores de Velocidade de Processamento e Memória Viso-Espacial de Curto Prazo (0,90 e 0,91) respectivamente, e ligeiramente abaixo para o fator Memória de Longo Prazo (0,86). Os índices GFI e RMSR indicaram um bom ajuste da solução fatorial aos dados, com valores de 0,99 e 0,02 respectivamente. A consistência interna (Alfa de Cronbach) dos escores dos três fatores indicou valores altos, variando entre 0,88 para o fator Velocidade de Processamento e 0,93 para o fator Memória de Longo Prazo.

Os resultados da análise fatorial exploratória também permitiram verificar que os três fatores obtidos apresentam correlações maiores do que 0,30. Dessa maneira foi realizada uma análise fatorial de segunda ordem, utilizando o método Schmid-Leiman (Thompson, 2004).

O método Schmid-Leiman é utilizado para estimar um modelo bifactor exploratório, em que há a remoção da variância dos fatores de primeira ordem nos fatores de segunda ordem, ou seja, na solução Schmid-Leiman a variância dos fatores de primeira e segunda ordem não é compartilhada, o que, conseqüentemente, reduz os valores das cargas fatoriais dos fatores de primeira ordem (Canivez, 2016). A solução Schmid-Leiman com três fatores de primeira ordem e um fator de segunda ordem encontra-se na Tabela 4.

Tabela 4

Cargas fatoriais usando análise fatorial de segunda-ordem (solução Schmid-Leiman) dos escores normalizados de sete subtestes de inteligência do IDS-2

Subteste	F1	F2	F3	G
ADC	0,32	-0,03	-0,08	0,73
RIF	0,27	-0,02	0,01	0,67
REH	-0,03	0,66	0,00	0,48
DIM	0,06	0,35	0,01	0,46
RFR	0,00	0,00	0,66	0,57
REF	0,02	0,05	0,45	0,48
COD	0,11	0,12	-0,01	0,36

Fator	G
F1	0,93
F2	0,65
F3	0,66

F1 = velocidade de processamento; F2 = memória de longo prazo; F3 = memória viso-espacial; G = fator de segunda ordem; ADC = Assinalar Duas Características; RIF = Riscar Figuras; REH = Repetir História; DIM = Descrever Imagem; RFR = Reconhecer Figuras com Rotação; REF = Reconhecer Figuras; COD = Colocar Discos.

A Tabela 4 indica que de todos os sete subtestes os dois subtestes de velocidade (Assinalar Duas Características e Riscar Figuras) têm as cargas fatoriais mais altas no fator geral. Com exceção do subteste Colocar Discos os dois subtestes de velocidade também mostram as cargas mais baixas no seu fator de primeira ordem. O subteste com a carga mais baixa no fator geral é o subteste Colocar Discos. O único subteste que tem a carga fatorial mais alta no fator de primeira ordem em comparação ao fator de segunda ordem é o subteste Repetir História. Aparentemente, este subteste tem variância específica que não pode ser explicada pelo fator geral de segunda ordem.

O fator Velocidade de Processamento apresenta a maior carga fatorial no fator geral de segunda ordem, enquanto os demais fatores, Memória de Longo Prazo e Memória Viso-Espacial de Curto Prazo, apresentam as cargas mais baixas no fator de segunda ordem. É importante ressaltar aqui que o fator geral na Tabela 4 não pode ser interpretado como o QI do IDS-2, levando em consideração que quatro fatores não são representados no fator geral, a saber: Raciocínio Abstrato, Raciocínio Verbal, Memória Fonológica de Curto Prazo e Processamento Visual.

A análise fatorial confirmatória foi conduzida com o objetivo de avaliar qual o modelo de mensuração dos sete subtestes de inteligência do IDS-2 que melhor se ajusta aos dados. Foram comparados três modelos: o modelo unidimensional, o modelo com dois fatores (um fator de memória e um fator de velocidade de processamento) e o modelo com três fatores baseado na teoria de Cattell-Horn-Carroll (dois fatores de memória e um de processamento de velocidade). Os índices de ajuste dos modelos de mensuração dos sete subtestes de inteligência do IDS-2 são apresentados na Tabela 5.

Tabela 5

Índices de ajuste de três modelos de análise fatorial confirmatória dos escores normatizados de sete subtestes de inteligência do IDS-2

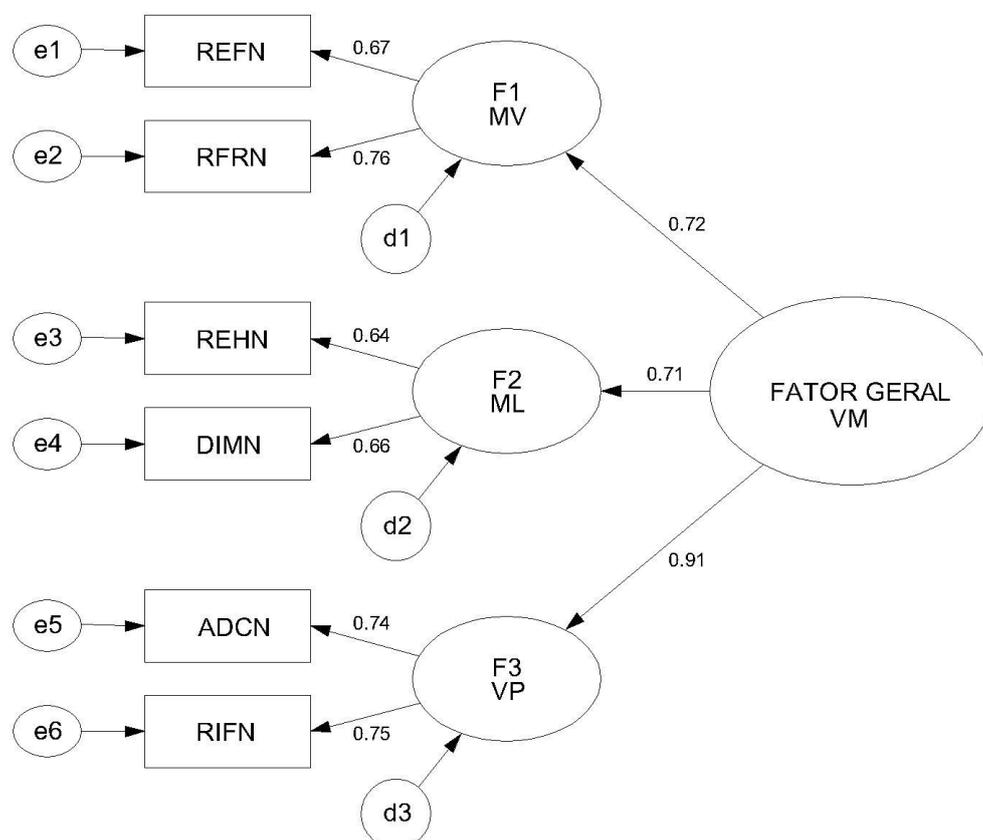
Modelo	χ^2	gl	$\Delta\chi^2$	Δgl	P	CFI	SRMR	RMSEA (IC 90%)
M-1	68,46	9	----	----	----	0,87	0,07	0,14 (0,11-0,17)
M-2	48,22	8	20,24	1	< 0,01	0,91	0,06	0,12 (0,09-0,15)
M-3	02,15	6	46,07	2	< 0,01	0,99	0,01	0,02 (0,01-0,04)

Notas: χ^2 = qui-quadrado; gl = graus de liberdade; $\Delta\chi^2$ = diferença de χ^2 entre dois modelos; Δgl = diferença de gl entre dois modelos; p = nível de significância da diferença de ajuste entre dois modelos; CFI = *Comparative Fit Index*; SRMR = *Standardized Root Mean Square Residual*; RMSEA = *Root Mean Square Error of Approximation*; IC 90% = o intervalo de confiança de 90% do RMSEA; M-1 = modelo unidimensional; M-2 = modelo com dois fatores e um fator de segunda ordem; M-3 = modelo com três fatores e um fator de segunda ordem

Na análise fatorial confirmatória utilizou-se o método de estimação *Maximum Likelihood*, e os seguintes índices recomendados por Byrne (2010) foram empregados para avaliar o ajuste dos modelos aos dados: o qui-quadrado (χ^2) acompanhado dos graus de liberdade (gl), o *Comparative Fit Index* (CFI), o *Standardized Root Mean Square Residual* (SRMR) e o *Root Mean Square Error of Approximation* (RMSEA) com o intervalo de confiança de 90%. A interpretação dos índices de ajuste foi baseada nos critérios sugeridos por Weston et al. (2008), que recomendam os seguintes valores para um bom ajuste: $CFI \geq 0,95$, $RMSEA \leq 0,06$ e $SRMR \leq 0,08$. De acordo com esses critérios o modelo que apresentou o melhor ajuste aos dados foi o modelo três (M-3). A Figura 1 abaixo apresenta o modelo final da Análise Fatorial Confirmatória.

Figura 1

Resultado para modelo de mensuração com melhor ajuste aos dados na Análise Fatorial Confirmatória (M-3)



As cargas fatoriais do modelo hierárquico M-3 apresentadas na Figura 1 mostraram que existe uma relação forte entre os fatores de primeira ordem com o fator geral de segunda ordem. O fator VP (Velocidade de Processamento) apresentou uma relação muito forte com o fator geral (carga fatorial de 0,91), enquanto os fatores ML (Memória de Longo Prazo) e MV (Memória Visual de Curto Prazo) apresentaram uma relação menor com o fator geral (carga fatorial de 0,71 e 0,72, respectivamente). Uma vez que a relação entre o fator Velocidade de Processamento e o fator geral é mais forte do que a relação entre os dois fatores de memória com o fator geral podemos interpretar o fator geral como um Fator de Velocidade e Memória. As cargas na Figura 1 indicam que o fator geral explica 82,8% da variância do fator VP, 51,8% da variância do fator MV e 50,4% da variância do fator ML. Ou seja, o fator geral não está conseguindo explicar 17,2% da variância do fator VP, e 48,2% e 49,6% dos fatores MV e ML. Podemos inferir dos resultados apresentados na Figura 1 que a velocidade de processamento é importante para resolver com sucesso as tarefas dos quatro subtestes de memória.

Por último, avaliou-se a precisão dos escores dos fatores obtidos pelo modelo da Análise Fatorial Confirmatória por meio dos índices de Confiabilidade Composta (CC) e Variância Média Extraída (VME) (Fornell & Larcker, 1981; Valentini & Damásio, 2016). De acordo com os resultados indicados na Tabela 6 observa-se que o Fator Geral de inteligência do modelo apresentou os maiores índices de CC (0,83) e VME (0,62). Por outro lado, o Fator Memória de Longo Prazo apresentou índices mais baixos, CC (0,60) e VME (0,42). De modo geral, os índices de precisão obtidos estão dentro dos pontos de corte estabelecidos na literatura, indicando valores adequados de precisão (Fornell & Larcker, 1981; Bagozzi & Yi, 1988). A única exceção foi observada para o Fator Memória de Longo Prazo, o qual apresentou uma Variância Média Extraída (VME) igual a 0,42. Entretanto, e conforme discutido por Valentini e Damásio (2016), os pontos de

corde desses dois índices devem ser relativizados, uma vez que eles são influenciados pelas cargas fatoriais e pelo quantitativo de itens do fator.

Tabela 6

Confiabilidade composta (CC) e Variância Média Extraída (VME) dos fatores do modelo com melhor ajuste aos dados (M-3)

Fator	CC	VME
Geral	0,83	0,62
Velocidade de Processamento (VP)	0,71	0,56
Memória Viso-Espacial (MV)	0,68	0,51
Memória de Longo Prazo (ML)	0,60	0,42

Resumindo, podemos afirmar que os resultados da análise fatorial confirmatória são consistentes com os resultados obtidos na análise fatorial exploratória, e podem ser interpretados como uma evidência da validade baseada na estrutura interna dos sete subtestes de inteligência do IDS-2.

Discussão

O objetivo desse estudo foi obter evidências de validade dos escores de um conjunto de subtestes de inteligência do IDS-2. Foram utilizados sete subtestes de inteligência do IDS-2: Reconhecer Figuras, Reconhecer Figuras com Rotação, Colocar Discos, Descrever Imagem, Repetir História, Assinalar Duas Características e Riscar Figuras (Grob & Hagmann-von Arx, 2018). Os escores brutos nos subtestes foram transformados em escores normatizados de QI, o que permitiu observar a redução do efeito da idade sobre os escores, chegando a valores próximos de 0. Esse efeito se dá em decorrência do método de normatização utilizado: *continuous norming* (Zachary & Gorsuch, 1985). O método de normatização contínua possui algumas vantagens em relação aos modelos tradicionais. É possível calcular os escores normatizados com todos

os grupos de idade ao mesmo tempo, tornando o escore bruto uma função da idade do sujeito. Além disso, o método de normatização contínua possibilita a utilização de amostras menores de cada grupo etário e maior acurácia nas estimativas de parâmetros populacionais.

Os resultados da análise fatorial exploratória indicaram uma solução com três fatores. O primeiro fator é formado pelos subtestes Assinalar Duas Características e Riscar Figuras, que avaliam o fator Velocidade de Processamento. O segundo fator é composto pelos subtestes Repetir História e Descrever Imagem, que avaliam o fator Memória de Longo Prazo, e o terceiro fator é formado pelos subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação, que avaliam o fator Memória Viso-Espacial de Curto Prazo. Essa estrutura fatorial replica o padrão de agrupamento dos sete subtestes nos estudos de normatização conjunta entre Suíça, Alemanha e Áustria (N = 1672) (Grob & Hagmann-von Arx, 2018).

Os fatores indicados na análise fatorial são compatíveis com as propostas do modelo teórico CHC (Cattell-Horn-Carroll). De acordo com o modelo CHC a inteligência é definida em termos de habilidades cognitivas organizadas em uma estrutura hierárquica de três estratos (níveis). O fator Velocidade de Processamento é localizado no segundo nível da estrutura hierárquica do modelo CHC, e indica a habilidade de executar tarefas cognitivas repetitivas de forma automática, fluente e rápida. A velocidade perceptiva é uma habilidade mais específica que está intimamente relacionada à velocidade de processamento, indicando a fluência e a velocidade para comparar e/ou diferenciar estímulos de natureza visual (Schneider & McGrew, 2018). As tarefas propostas pelos subtestes Assinalar Duas Características e Riscar Figuras do IDS-2 são caracterizadas pelos processos de busca e reconhecimento de estímulos visuais (figuras com cores ou

formatos específicos) em meio a estímulos visuais semelhantes (Grob & Hagmann-von Arx, 2018).

O fator Memória Viso-Espacial de Longo Prazo, representado na solução fatorial do IDS-2 pelos subtestes Repetir História e Descrever Imagem, é definido na teoria CHC como a capacidade de armazenar e evocar informações através de períodos (de minutos até anos) (Schneider & McGrew, 2018). No IDS-2 os subtestes Repetir História e Descrever Imagem avaliam estímulos diferentes: auditivo (Repetir História) e visoespacial (Descrever Imagem), porém ambos exigem que as informações sejam recuperadas após o período mínimo de 20 minutos transcorridos após a apresentação dos estímulos. O agrupamento dos subtestes em um único fator indica que os subtestes avaliam o construto Memória de Longo Prazo, apesar das diferenças nos estímulos.

O fator Memória Viso-Espacial de Curto Prazo pode ser definido como a habilidade de manter informações de natureza visoespacial na memória de curto-prazo (Schneider & McGrew, 2012). Atualmente a memória visoespacial de curto prazo é considerada uma habilidade específica da memória de trabalho (Schneider & McGrew, 2018). No IDS-2 esse fator é representado pelos subtestes Reconhecer Figuras e Reconhecer Figuras com Rotação. No subteste Reconhecer Figuras o sujeito observa uma figura e na sequência reconhece essa figura em meio a outras imagens. A mesma tarefa é proposta no subteste Reconhecer Figuras com Rotação, com a diferença de que as figuras serão rotacionadas (Grob & Hagmann-von Arx, 2018).

Os três fatores apresentaram uma correlação média entre si de 0,55, indicando a interdependência entre os fatores de inteligência. A correlação entre os fatores também permitiu explorar uma solução fatorial com um fator de segunda ordem. Os resultados obtidos por meio da análise fatorial exploratória são compatíveis com o modelo de mensuração que obteve o melhor ajuste aos dados na análise fatorial confirmatória. Em

ambos os casos se obteve uma estrutura hierárquica com um fator geral de segunda ordem e três fatores de primeira ordem: Velocidade de Processamento, Memória de Longo-Prazo e Memória Viso-Espacial de Curto Prazo. É importante destacar que o fator geral encontrado no presente estudo, não pode ser confundido com o fator geral (g) de Spearman proposto em 1904. Para se obter um fator geral de inteligência é necessário incluir outros fatores de inteligência mais centrais no construto de g, como Raciocínio Abstrato, Raciocínio Verbal e Processamento Visual.

Conclusão

Os resultados do presente estudo estão em concordância com os achados das pesquisas internacionais com o IDS-2 (Grob & Hagmann-von Arx, 2018). A estrutura fatorial do instrumento revela a natureza complexa e multifacetada do construto inteligência. Além disso, os resultados vão ao encontro às proposições mais atuais do modelo teórico CHC, que representa um importante avanço no estudo das habilidades cognitivas, permitindo a revisão e o desenvolvimento de novas ferramentas de avaliação. Destaca-se ainda a importância de se considerar as diferentes facetas da inteligência humana nos processos de testagem e avaliação, o que possibilita a identificação de déficits específicos e a intervenção mais eficaz sobre os processos cognitivos. Essa característica é um destaque importante na avaliação da inteligência por meio do IDS-2, e que tem sido uma recomendação para o levantamento de perfis cognitivos no âmbito clínico e de pesquisa (Schneider & Flanagan, 2015).

É importante destacar que o subteste Colocar Discos não apresentou cargas fatoriais significativas nos três fatores de primeira ordem. A tarefa do subteste Colocar Discos envolve o reconhecimento visual e a repetição do padrão de disposição espacial das peças apresentadas (discos metálicos), o que difere esse subteste dos demais avaliados nesse estudo. As habilidades requeridas para a tarefa de Colocar Discos estão associadas ao

fator Processamento Visual do modelo CHC, que, de modo simplificado, reflete a capacidade dos sujeitos em perceber e manipular imagens mentais para solucionar problemas (Schneider & McGrew, 2018). Esse achado pode ser considerado um indicador de que o subteste avalia este outro fator, e não os fatores relacionados neste estudo.

Por último se entende a necessidade de conduzir mais estudos com todos os 14 subtestes de inteligência do IDS-2, incluindo a obtenção de evidências empíricas de validade dos escores com base em outras fontes de validade. É fundamental aplicar todo o conjunto de subtestes de inteligência do IDS-2 em amostras maiores e mais representativas da população brasileira a fim de poder fazer afirmações mais bem fundadas sobre os diversos fatores do funcionamento intelectual e a relação entre eles.

Referências

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Arbuckle, J. L. (2009). *Amos 18 User's Guide*. Amos Development Corporation.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94. doi:10.1007/BF02723327
- Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment from three perspectives. *Journal of Personality Assessment*, *85*(1), 17-32.
- Canivez, G. L. (2016). Bifactor modeling in construct validation of multifaceted tests: Implications for multidimensionality and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247–271). Hogrefe.
- Canivez, G. L., & Youngstrom, E. A. (2019). Challenges to the Cattell-Horn-Carroll Theory: Empirical, clinical, and policy implications. *Applied Measurement in Education*, *32*(3), 232-248.
- Costello, A. B., & Osbourne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7), 1-9. doi.org/10.7275/jyj1-4868
- Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). Hierarchical exploratory factor analyses of the Woodcock-Johnson IV Full Test Battery: Implications for CHC application in school psychology. *School Psychology Quarterly*, *33*, 235-250. doi.org/10.1037/spq0000221
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equations models with unobservable variables and measurement error. *Journal of Marketing*, *18*(1), 39-50. doi:10.2307/3151312
- Ferrando, P.J., & Lorenzo-Seva, U. (2017). Programa FACTOR at 10: Origins, development and future directions. *Psicothema*, *29*(2), 236-240. doi: 10.7334/psicothema2016.304
- Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, *78*(5), 762–780. doi.org/10.1177/0013164417719308
- Forgues, S. (2014). *Aptitude testing of military pilot candidates*. [Unpublished master's thesis]. Queen's University.
- Grob, A., & Hagemann-von Arx, P. (2018). *Intelligenz- und entwicklungs-skalen für kinder und jugendliche* [Intelligence and Development Scales - 2 (IDS-2)]. Hogrefe.

- Jewsbury, P. A., & Bowden, S. C. (2016). Construct validity of fluency and implications for the factorial structure of memory. *Journal of Psychoeducational Assessment*, 35(5), 460-481.
- Kittrell, C. L. (1999). *Continuous norming: Estimation of test statistics for normative purposes* [Unpublished doctoral dissertation]. University of Pasadena.
- Laros, J.A., Tellegen, P., Lima, R.M.F., & Valentini, F. (2021). *SON-R 6-40: Teste não-verbal de inteligência. Parte I. Relatório científico*. Editora Hogrefe CETEPP.
- Lenhard, A., Lenhard, W., & Gary, S. (2018). Continuous norming of psychometric tests: A simulation study of parametric and semiparametric approaches. *PLoS ONE*, 14(9). doi.org/10.1371/journal.pone.0-2222279.
- 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
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 151-179). Guilford Press.
- Oosterhuis, H. E. M. (2007). *Regression based norming for psychological tests and questionnaires* [Unpublished doctoral dissertation]. University of Tilburg.
- Schelini, P. W. (2006). Teoria das inteligências fluida e cristalizada: Início e evolução. *Estudos de Psicologia*, 11(3), 232-332.
- Schneider, W. J., & Flanagan, D. P. (2015). The relationship between theories of intelligence and intelligence tests. In S. Goldstein, D. Princiotta & J. A. Naglieri (Eds.), *Handbook of intelligence: Evolutionary theory, Historical perspective and current concepts* (pp. 317-340). Springer.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D.P. Flanagan & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3th ed., pp. 99-144). The Guilford Press.
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (4rd ed., pp. 71-163). The Guilford Press.
- 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.
- Stough, C., & Pase, M. P. (2015). Improving cognition in the elderly with nutritional supplements. *Current Directions in Psychological Science*, 24(3), 177-183.
- Tellegen, P. J., & Laros J. A. (2014). *SON-R 6-40. Snijders-Oomen Non-verbal intelligence test. Volume I: Research report*. Hogrefe.

- 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.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.
- Urbina, S. (2014). *Essentials of psychological testing* (2nd ed.). Wiley.
- Valentini, F., & Damásio, B. F. (2016). Variância média extraída e confiabilidade composta: Indicadores de precisão. *Psicologia: Teoria e Pesquisa*, *32*(2) 1-7.
- Voncken, L., Albers, C. J., & Timmerman, M. E. (2017). Model selection in continuous test norming with GAMLSS. *Assessment*, *26*(7), 1329-1346. doi.org/10.1177/1073191117715113
- Wechsler, S. M., & Nakano, T. de C. (2016). Cognitive assessment of Brazilian children and youth: Past and present perspectives and challenges. *International Journal of School and Educational Psychology*, *4*(4), 215-224.
- Weston, R., Gore, P. A., Chan, F., & Catalano, D. (2008). An introduction to using structural equation models in rehabilitation psychology. *Rehabilitation Psychology*, *53*(3), 340-356.
- Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: implications for the WAIS-R. *Journal of Clinical Psychology*, *41*(1), 86-94.

Considerações Finais

No Brasil existe a escassez de instrumentos para a avaliação da inteligência do público infantil, campo em que os testes de inteligência exercem um papel importante, o que indica a necessidade de novos instrumentos de medida desse construto (Ferreira et al., 2010). A avaliação de pré-escolares, por exemplo, tem apenas um instrumento disponível: o teste não-verbal de inteligência SON-R 2½-7[a], que é voltado para a faixa etária de crianças a partir dos 2 anos e meio de idade até os 7 anos e 11 meses (Laros et al., 2015).

Nos processos diagnósticos a avaliação da inteligência auxilia na identificação de níveis de funcionamento cognitivo e no planejamento de intervenções e reabilitação. A avaliação da inteligência durante o desenvolvimento permite a identificação precoce de limitações e o início imediato das intervenções. É durante o desenvolvimento que há maturação e aumento exponencial de novas habilidades. Além disso, a plasticidade neuronal é maior durante a infância, tornando a intervenção mais eficaz e o prognóstico mais positivo (Garlick, 2003; Ferreira et al., 2010).

A testagem da inteligência e suas facetas também pode ser empregada com a finalidade de estabelecer funções preservadas. Nesse sentido, a utilização de testes que avaliam diferentes dimensões cognitivas pode ser considerada um padrão-ouro, pois possibilita o mapeamento cognitivo e a construção de intervenções mais adequadas (Mecca et al., 2014). Uma das grandes vantagens da testagem com o IDS-2 é a avaliação de diferentes fatores da inteligência, o que adiciona informações ao processo avaliativo, tornando-o mais individualizado (Grob & Haggmann-von Arx, 2018). O IDS-2 representa uma ferramenta em potencial para a área de avaliação psicológica no Brasil. Ressalta-se a necessidade de realizar mais estudos com o instrumento em amostras maiores e mais representativas da população brasileira.

Referências

- Ferreira, F. O., Coutinho, G., Freitas, P. M., Malloy-Diniz, L. F., & Haase, V. G. O. (2010). Exame neuropsicológico na idade pré-escolar. In L. F. Malloy-Diniz, D. Fuentes, P. Mattos, & N. Abreu, (Eds.), *Avaliação Neuropsicológica* (pp. 210-220) Artmed.
- Garlick, D. (2003). Integrating brain science research with intelligence research. *Current Directions in Psychological Science*, 12(5), 185-189. <https://doi.org/10.1111/1467-8721.01257>
- Grob, A., & Hagmann-von Arx, P. (2018). *Intelligenz- und Entwicklungsskalen für Kinder und Jugendliche* [Intelligence and Development Scales - 2 (IDS-2)]. Hogrefe.
- Laros, J. A., Tellegen, P. J., Jesus G. R., & Karino, C. A. (2015). *SON-R 2½-7[a] – Teste não-verbal de inteligência. Manual*. Hogrefe-CETEPP.
- Mecca, T. P., Orsati, F. T., & Macedo, E. C. (2014). Non-verbal cognitive profile of young children with autism spectrum disorders. *Psychology*, 5(11), 1404-1417. doi: 10.4236/psych.2014.511151.