



UNIVERSIDADE FEDERAL DE SANTA CATARINA
Centro de Filosofia e Ciências Humanas
Departamento de Psicologia
Programa de Pós-Graduação em Psicologia

CASSANDRA MELO OLIVEIRA

**CONSTRUÇÃO E BUSCA DE EVIDÊNCIAS DE VALIDADE DE
UM BANCO DE ITENS DE PERSONALIDADE PARA
TESTAGEM ADAPTATIVA DESENVOLVIDO A PARTIR DOS
PRINCÍPIOS DO DESENHO UNIVERSAL**

Orientador: Prof. Dr. Carlos Henrique Sancineto da Silva Nunes

FLORIANÓPOLIS, SC

2017

CASSANDRA MELO OLIVEIRA

**CONSTRUÇÃO E BUSCA DE EVIDÊNCIAS DE VALIDADE DE
UM BANCO DE ITENS DE PERSONALIDADE PARA
TESTAGEM ADAPTATIVA DESENVOLVIDO A PARTIR DOS
PRINCÍPIOS DO DESENHO UNIVERSAL**

Tese apresentada como requisito parcial à obtenção de grau de Doutora em Psicologia, Programa de Pós-Graduação em Psicologia, Mestrado/Doutorado, Centro de Filosofia e Ciências Humanas da Universidade Federal de Santa Catarina.

Orientador: Prof. Dr. Carlos Henrique Sancineto da Silva Nunes

Oliveira, Cassandra Melo

Construção e busca de evidências de validade de um banco de itens de personalidade para testagem adaptativa desenvolvido a partir dos princípios do Desenho Universal / Cassandra Melo Oliveira ; orientador, Carlos Henrique Sancineto da Silva Nunes, 2017.
173 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Filosofia e Ciências Humanas, Programa de Pós-Graduação em Psicologia, Florianópolis, 2017.

Inclui referências.

1. Psicologia. 2. Testagem Universal. 3. Personalidade. 4. Banco de Itens. 5. Testagem Adaptativa. I. Nunes, Carlos Henrique Sancineto da Silva. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Psicologia. III. Título.


Cassandra Melo Oliveira

Construção e busca de evidências de validade de um banco de itens de personalidade para testagem adaptativa desenvolvido a partir dos princípios do desenho universal


Tese aprovada como requisito parcial à obtenção do grau de Doutor em Psicologia, Programa de Pós-Graduação em Psicologia, Centro de Filosofia e Ciências Humanas da Universidade Federal de Santa Catarina.


Florianópolis, 24 de Novembro de 2017.


Dr. Carlos Henrique Sancineto da Silva Nunes
(Coordenador - PPGP/UFSC)


Dr. Carlos Henrique Sancineto da Silva Nunes
(PPGP - UFSC - Orientador)

Fernanda M. Lopes
Dra. Fernanda Machado Lopes
(PPGP-UFSC - Examinadora)

Videoconferência 
Dr. Ricardo Primi
(USF - Examinador)

Videoconferência 
Dr. José Maurício Haas Bueno
(UFPe - Examinador)

Dra. Carolina Baptista Menezes
(PPGP-UFSC - Examinadora suplente)

Dr. Lucas de Oliveira Carvalho
(USF - Examinador suplente)

Para os meus amores, Adolfo e Sofia.

Agradecimentos

A todos os que fizeram parte da minha trajetória acadêmica e de vida.

Ao meu orientador Carlos Henrique Sancineto da Silva Nunes agradeço pelo muito que aprendi nestes sete anos (mestrado e doutorado) e pela confiança depositada em mim.

Ao professor Adriano Henrique Nuernberg agradeço a colaboração e as boas ideias que me auxiliaram no processo de construção desta pesquisa desde o mestrado até a construção da tese.

A minha família.

Ao meu amor e a minha pequena que desde a barriga participou da escrita desta tese.

Aos amigos, tantos, em várias partes do Brasil e do mundo.

A todas as pessoas com deficiência.

Aos membros da Associação Catarinense para Integração do Cego (ACIC) pela colaboração incansável que tornou este trabalho possível.

A CAPES pela bolsa que foi essencial para que eu pudesse me dedicar com afinco a este trabalho.

A religiosidade e a Deus.

SUMÁRIO

RESUMO

INTRODUÇÃO	1
Artigo 1. Avaliações de Testagem Universal: O Modelo.....	13
Artigo 2. Construção de um Banco de Itens de Personalidade e Testagem Universal em Testagem Adaptativa.....	48
Artigo 3. Dimensionalidade de um banco de itens de Personalidade e Testagem Universal.....	75
Artigo 4. Calibração dos Itens de um Banco de Itens de Personalidade e Testagem Universal.....	97
Artigo 5. Funcionamento Diferencial dos itens de um teste para avaliação da personalidade baseado na Testagem Universal.	134
CONSIDERAÇÕES FINAIS	160
REFERÊNCIAS	165

RESUMO

Na Testagem Universal pessoas com e sem deficiência podem realizar os instrumentos de forma acessível procurando, assim, abarcar a diversidade humana. Esta tese primou pela elaboração de um banco de itens em testagem adaptativa, a qual é uma tarefa de alta complexidade que demanda amplo conhecimento técnico e teórico. Somou-se a tal fato a elaboração de um banco de itens segundo os princípios da Testagem Universal no construto personalidade, Cinco Grandes Fatores. Os objetivos foram: 1) Desenvolver um banco de itens de personalidade e Testagem Universal; 2) Delinear um roteiro para elaboração de testes segundo os princípios da Testagem Universal abrangendo a conceituação do teste, a escolha do número e tipo de itens, as análises e correções, a aplicação em campo, às análises qualitativas e quantitativas e a revisão final; 3) Descrever os desdobramentos alcançados pela Testagem Universal; 4) Construir um instrumento para calibração dos itens; 4) Verificar a dimensionalidade de um banco de itens de Personalidade e Testagem Universal; 5) calibrar os itens de um banco de itens de Personalidade e Testagem Universal; e, 6) Verificar a invariância dos parâmetros dos itens entre os indivíduos com e sem deficiência de um banco de itens de Personalidade e Testagem Universal. Os principais resultados nesta tese apontam para o desenvolvimento da proposta de um Modelo Integrativo de Testagem Universal. Foi construído um instrumento para calibração dos itens de personalidade e Testagem Universal, sendo este implementado em seu formato segundo os mesmos princípios. Foi perpetrada as análises de juízes com a reescritura, a exclusão de itens e a mudança de faceta – na medida em que cada um destas ações fez-se necessário. Os itens foram implementados no programa Concerto seguindo o Modelo de BIB escolhido para organização do banco. Ainda, os 525 itens deste estudo foram resumidos em cinco fatores por meio da análise fatorial. Os itens do Banco de Itens tiveram seus parâmetros estimados utilizando o modelo de resposta gradual ou de Samejima, separadamente, em cada um dos cinco fatores resultando em: 64 itens de Abertura, 58 itens de Extroversão, 82 itens de Neuroticismo, 76 de Realização e 37 de Socialização para o BI. Na análise do Funcionamento Diferencial dos Itens consideraram-se como itens indicativos de exclusão aqueles nos quais o DIF apresentou valores de moderados a altos ($|DIF| \geq 0,64$), contudo, ainda, submeteram-se os itens a reanálises de Testagem Universal (DIF entre as pessoas com e sem deficiência) e quando constado o não atendimento a algum dos princípios da TU, o item também foi indicado para exclusão.

Conclusões: No estudo da dimensionalidade os Cinco Grandes Fatores da Personalidade: Abertura, Realização, Extroversão, Socialização e Neuroticismo, apresentaram resultados semelhantes à literatura científica sedimentada da área. Acresce-se que após a calibração, 317 itens foram considerados aptos a compor o Banco de Itens de Personalidade e Testagem Universal. No tocante à aplicação da Testagem Universal verificou-se que os resultados foram bastante promissores na busca pela invariância dos itens uma vez que poucos itens apresentaram valores de DIF de moderados a altos. Assim, diante dos resultados alcançados, conclui-se que esta tese alcançou as primeiras evidências de validade de um Banco de Itens de Personalidade e Testagem Universal para Testagem Adaptativa.

Palavras-chave: Testagem Universal; Personalidade; Banco de Itens; Testagem Adaptativa.

ABSTRACT

In universal testing, people with and without disabilities can carry out instruments in an accessible way, thus seeking to encompass human diversity. This thesis was based on the elaboration of a database of adaptive testing items, which is a highly complex task that requires extensive technical and theoretical knowledge. It was added to this fact the elaboration of a bank of items according to the principles of the Universal Testament in the construct personality, Big Five. The objectives were: 1) To develop a bank of personality items and Universal Testing; 2) Outline a road map for the preparation of tests according to the principles of universal testing, including the concept of the test, the choice of number and type of items, analyzes and corrections, field application, qualitative and quantitative analysis and final review; 3) Describe the results achieved by universal testing; 4) Construct an instrument to calibrate items; 4) Check the dimensionality of a bank of Personality and Universal Test items; 5) calibrate the items of a bank of Personality and Universal Test items; and, 6) Verify the invariance of the parameters of the items between the individuals with and without deficiency of a bank of Personality and Universal Test items. The main results were: In this thesis was developed the proposal of an Integrative Model of Universal Testing. An instrument for the calibration of personality and universal testing items was constructed, and this one was implemented, in its format, according to these same principles. Judges' analysis of rewriting, item exclusion and facet change was perpetrated - when each of these aspects became necessary. The items were implemented in the Concerto program following the BIB Model chosen for the organization of the bank. Moreover, the 525 items of this study were summarized in 5 factors by means of factorial analysis. The Item Bank items had their parameters estimated using the Samejima step-by-step model, separately, in each of the five factors resulting in: 64 BI opening items, 58 BI extroversion items, 82 items of Neuroticism, 76 of Achievement and 37 of Socialization. In the analysis of the Differential Functioning of the Items, items in which DIF presented moderate to high values ($|DIF| \geq 0.64$) were considered as items indicative of exclusion; however, items were also submitted to Universal Testing reanalysis (DIF between people with and without disabilities) and when not complying with some of the TU principles, the item was also indicated for exclusion. Conclusions: In the study of dimensionality, the Big Five: Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism, presented similar results

to the sedimented scientific literature of the area. In addition, after the calibration, 317 items were considered fit to compose the Personality Bank and Universal Testing Items. Regarding the application of the Universal Testing, it was verified that the results were very promising in the search for the invariance of the items since few items presented moderate to high DIF values. Thus, in view of the results achieved, it is concluded that this thesis reached the first evidence of validity of a Bank of Personality Items and Universal Testing for Adaptive Testing.

Keywords: Universal Testing; Personality; Bank of Items; Adaptive Testing.

RESUMEN

En la Testagem Universal personas con y sin discapacidad pueden realizar los instrumentos de forma accesible buscando, así, abarcar la diversidad humana. Esta tesis primó por la elaboración de un banco de ítems en prueba adaptativa, la cual es una tarea de alta complejidad que demanda amplio conocimiento técnico y teórico. Se sumó a tal hecho la elaboración de un banco de ítems según los principios del Testagem Universal en el constructo personalidad, Cinco Grandes Factores. Los objetivos fueron: 1) Desarrollar un banco de ítems de personalidad y Testagem Universal; 2) Delinear un guión para la elaboración de pruebas según los principios de la prueba universal que abarque la concepción del test, la elección del número y tipo de ítems, los análisis y correcciones, la aplicación en campo, los análisis cualitativos y cuantitativos y la revisión final; 3) Describir los desdoblamientos alcanzados por la prueba universal; 4) Construir un instrumento para la calibración de los elementos; 4) Verificar la dimensionalidad de un banco de elementos de Personalidad y Testamento Universal; 5) calibrar los elementos de un banco de elementos de Personalidad y Testament Universal; y, 6) Verificar la invariancia de los parámetros de los ítems entre los individuos con y sin deficiencia de un banco de ítems de Personalidad y Testagem Universal. Los principales resultados fueron: En esta tesis se desarrolló la propuesta de un Modelo Integrativo de Testagem Universal. Se construyó un instrumento para la calibración de los elementos de personalidad y testación universal, siendo éste, implementado, en su formato, según estos mismos principios. Se perpetró a los análisis de jueces con la reescritura, la exclusión de ítems y el cambio de faceta - cuando cada uno de estos aspectos se hizo necesario. Los ítems fueron implementados en el programa Concerto siguiendo el modelo de BIB elegido para la organización del banco. Además, los 525 ítems de este estudio fueron resumidos en 5 factores por medio del análisis factorial. Los ítems del Banco de ítems tuvieron sus parámetros estimados utilizando el modelo de respuesta gradual o de Samejima, por separado, en cada uno de los cinco factores resultando en: 64 ítems de Apertura para el BI, 58 ítems de Extroversión en el BI, 82 ítems de Neuroticismo , 76 de Realización y 37 de Socialización. En el análisis del Funcionamiento Diferencial de los ítems se consideraron como ítems indicativos de exclusión aquellos en los cuales el DIF presentó valores de moderados a altos ($DIF \geq 0,64$), sin embargo, se sometieron los ítems a reanálisis de Testagem Universal (DIF entre las personas con y sin discapacidad) y cuando constató el no atendimento a

alguno de los principios de la TU, el ítem también fue indicado para exclusión. Conclusiones: En el estudio de la dimensionalidad Los Cinco Grandes Factores de la Personalidad: Apertura, Realización, Extroversión, Socialización y Neuroticismo, presentaron resultados semejantes a la literatura científica sedimentada del área. Además, después de la calibración, 317 ítems fueron considerados aptos para componer el Banco de ítems de Personalidad y Testagem Universal. En cuanto a la aplicación del Testagem Universal se verificó que los resultados fueron bastante prometedores en la búsqueda de la invariancia de los ítems, ya que pocos ítems presentaron valores de DIF de moderados a altos. Así, ante los resultados alcanzados, se concluye que esta tesis alcanzó las primeras evidencias de validez de un Banco de ítems de Personalidad y Testagem Universal para Testagem Adaptativa.

Palabras clave: Prueba Universal; Personalidad; Banco de artículos; Prueba adaptativa.

INTRODUÇÃO

“Respeito à Diversidade”, “Viver sem limites”, “Superação”, “A luta por mobilidade com acessibilidade continua”, “Na Luta por mais acessibilidade e inclusão”. As frases apresentadas são algumas das “palavras de ordem” encontradas em diversas notícias que abordam a luta das pessoas com deficiência por acessibilidade. Ultrapassar barreiras arquitetônicas, tecnológicas e sociais faz parte do dia-a-dia das pessoas com deficiência na busca pela inclusão social. Desvencilhar-se de um mundo instituído para as pessoas sem deficiência e procurar novas possibilidades de adaptação, transformação e vivência é uma constante na luta pela acessibilidade. As barreiras abrangem os mais diversos campos, dentre eles o do instrumental psicológico, os quais com frequência são desenvolvidos e apresentados em formatos inacessíveis ao público com deficiência.

A diversidade corporal humana é pouco considerada no processo de desenvolvimento dos instrumentos de avaliação utilizados por psicólogos. Os formatos comumente implementados nos testes psicológicos inviabilizam que indivíduos com deficiência os realizem com condições semelhantes às de pessoas sem deficiência (Thurlow, Lazarus, Chistensen & Shyyan, 2016; Thurlow, Lazarus, Albus & Hodgson, 2010). É necessário aumentar o tempo de realização dos testes para as pessoas com deficiência? Como uma pessoa cega responde a um teste no formato lápis e papel? O formato informatizado é acessível? Que recursos teóricos, técnicos e tecnológicos quando aplicados potencializam a acessibilidade dos testes psicológicos? Estas são algumas das perguntas norteadoras para o desenvolvimento de instrumentos de acessibilidade plena.

A acessibilidade plena consiste na possibilidade do maior número de pessoas possível utilizarem um produto apropriando-se aos seus objetivos e usos autonomamente. Contudo, diante da diversidade humana torna-se um grande desafio atender a um público deveras amplo e com características tão variadas quanto às pessoas com deficiência, os idosos, as crianças, etc. Assim, aporta-se no conceito do Desenho Universal que visa a criar espaços e produtos capazes de atender a esta diversidade humana com acessibilidade plena e que, para tanto, apresenta sete princípios: 1) o uso equitativo, 2) o uso flexível, 3) o uso simples e intuitivo, 4) a informação perceptível, 5) a tolerância ao erro, 6) o esforço físico mínimo e 7) o uso e acesso ao tamanho e espaço (Story, Mueller & Mace, 1998). Tais princípios resumem as

características que produtos ou espaços devem possuir para serem considerados de Desenho Universal.

O conceito de Desenho Universal surgiu na área da arquitetura e a partir desta alcançou a educação e a avaliação psicológica. O Desenho Universal busca desenvolver produtos, edificações e objetos que possam ser utilizados por um número maior de pessoas incluindo pessoas com deficiências, mas não apenas estas como crianças, idosos e pessoas sem qualquer deficiência (Story, 2001). Quando aplicada à testagem psicológica objetiva avaliar uma população ampla almejando acessibilidade populacional plena ou máxima possível ao teste. Os princípios do Desenho Universal são aplicáveis à construção e adaptação de instrumentos psicológicos, tendo muito a contribuir no processo de desenvolvimento de itens não tendenciosos - livres de vieses (*bias*), na elaboração das instruções dos testes e, de maneira geral, na qualidade do formato dos instrumentos (formato dos itens, legibilidade, apresentação intuitiva e de fácil percepção). Na sua aplicação aos instrumentais do psicólogo e a avaliação psicológica surge a Testagem Universal (Oliveira, 2013).

O desenvolvimento e a adaptação de testes psicológicos, bem como todos os aspectos que os tangem, torna-se seara científica de interesse da Testagem Universal. A Testagem Universal respeita todos os pressupostos teóricos instituídos no tocante aos testes psicológicos preocupando-se com o formato dos testes e sua influência na realização do mesmo. Foca as adaptações e os recursos de tecnologias assistivas quanto proporcionadores de acessibilidade analisando sua repercussão na qualidade dos instrumentos (Ketterlin-Geller, 2005; Thompson, Johnstone & Thurlow, 2002). Os princípios do Desenho Universal, na sua aplicação à testagem, foram sistematizados em sete princípios da Testagem Universal por Thompson, Johnstone e Thurlow (2002): 1) População de avaliação ampla e inclusiva, 2) Definição precisa do construto, 3) Itens acessíveis e não tendenciosos, 4) Testes flexíveis a acomodações, 5) Instruções e procedimentos simples, claros e intuitivos, 6) Texto compreensível e de leitura fácil, e, 7) Máxima legibilidade. Um teste será considerado de Testagem Universal se atender a estes sete princípios total ou parcialmente.

No desenvolvimento de testes psicológicos de Testagem Universal a serem utilizados pelas pessoas com deficiência faz-se imprescindível a utilização de tecnologias assistivas as quais criam um lócus de acessibilidade. A tecnologia assistiva, quando bem empregada, viabiliza a realização dos instrumentos por pessoas com deficiência de forma equitativa às pessoas sem deficiência. Porém, tais recursos exigem

que o instrumento seja planejado desde o princípio com vistas a garantir a exequibilidade técnica do mesmo. Os testes informatizados se tornam grandes aliados na construção de testes de Testagem Universal por permitirem que os instrumentos possuam maior flexibilidade agregando várias possibilidades de usos e recursos adicionais para as pessoas com deficiência (Thurlow, Lazarus, Albus & Hodgson, 2010).

Os testes informatizados acrescentam inúmeras vantagens frente aos tradicionais (lápis e papel). Os mesmos comportam a criação de um ambiente que proporciona maiores possibilidades de variação e conseqüentemente maior acessibilidade, isto é, maior flexibilidade no uso, escolha de formatos, tamanhos de fontes, além do fato de que admite o acoplamento de Tecnologias Assistivas que venham a complementar e potencializar o Desenho Universal. Adiciona-se que nestes instrumentos é possível agregar ferramentas capazes de calcular com precisão o tempo de testagem, possibilidade de repetição ilimitada das instruções, isto é, de acordo com as necessidades do testando, além de permitir correções informatizadas e a utilização de testagem adaptativa (*computerized adaptive testing - CAT*)¹ (Pietro, 2010; Thurlow, Lazarus, Albus & Hodgson, 2010; Pasquali, 2007).

Na testagem adaptativa computadorizada (CAT) o computador deixa de ser apenas aplicador do teste e passa a ser executor do mesmo – para tanto é necessário que exista um banco de itens também informatizado. Utilizando-se programação computacional é possível escolher itens entre os existentes no banco que estejam em consonância com a habilidade do sujeito que está realizando o teste. Assim, dependendo do nível de habilidade do indivíduo ele realizará um teste potencialmente diferente. Tal fator permite que sejam bem avaliados indivíduos nos diversos níveis de proficiência (elevado, médio e baixo) de forma ótima e não apenas com um conjunto fixo de itens preparados para a avaliação de uma faixa pré-definida como comumente ocorre nos testes tradicionais. A testagem adaptativa permite avaliar variados níveis de proficiência com máxima precisão em um curto período de tempo, e, deste modo, o testando não precisa submeter-se a um elevado número de itens (Pasquali, 2007; Eggen & Verschoor, 2006). Quando as vantagens dos testes informatizados e da testagem adaptativa (CAT) são somadas ao conceito do Desenho Universal potencializam-se os atributos e permite-se que indivíduos com e sem deficiência beneficiem-se deste instrumental.

¹ A sigla em inglês CAT foi utilizada significando testagem adaptativa por tratar-se de termo difundido em artigos em português e em *softwares*.

A testagem adaptativa é uma aplicação da teoria de resposta ao item (TRI). O desenvolvimento da Psicologia enquanto campo de estudo científico requer diante de problemas de alta complexidade a concomitante aplicação de métodos estatísticos robustos que permitam a compreensão do fenômeno estudado com maior qualidade e detalhamento. A Teoria da Resposta ao Item (TRI) é um conjunto de modelos matemáticos que permitem de forma probabilística uma estimativa do traço latente estudado. Assim, é possível estimar a probabilidade de um indivíduo acertar a um item em função dos parâmetros do item e da magnitude do traço latente. (Castro, Trentini & Riboldi, 2010; Andrade, Tavares & Valle, 2000). A utilização da TRI corrobora o atendimento dos pressupostos da Testagem Universal, pois proporciona a realização de estudos que potencializam a qualidade dos instrumentos psicológicos permitindo medidas mais eficazes e com níveis aceitáveis de vies.

A expressão vies (*bias*) foi cunhada em um momento na evolução dos testes psicológicos, no qual foi constatado que o conteúdo dos itens era muito importante para compreensão do mesmo. Fatores culturais influenciam o entendimento que se tem de um item, sobretudo no que se refere à língua já que esta é rica de expressões próprias de um povo, possui seus regionalismos e representa valores culturais ligados a um tempo e/ou lugar – o que pode afetar o conteúdo dos testes. Com a finalidade de evitar que um item se torne injusto para um público, ou seja, que pessoas com o mesmo nível de aptidão ou proficiência em um construto, mas, por exemplo, de culturas diversas, possuam diferentes probabilidades de responder a um item corretamente foram desenvolvidas técnicas estatísticas que detectam este problema (Pasquali, 2007; Zumbo, 2007). Acresce-se a este fato que nos testes de Testagem Universal almeja-se uma população ampla que inclua pessoas com e sem deficiência e, assim sendo, o estudo minucioso do conteúdo e do vies dos itens, bem como seus desdobramentos, faz-se imprescindível.

A análise estatística da função diferencial dos itens (DIF)², a qual é uma derivação do processo de calibração da TRI, é capaz de detectar o vies dos itens e, possui tamanha gravidade no desenvolvimento de instrumentos de Testagem Universal que o terceiro

² Usou-se a sigla em inglês – DIF – devido à sua disseminação na literatura científica nacional.

princípio, *itens acessíveis e não tendenciosos*, a destaca como de máxima relevância para detectar a tendenciosidade ou não dos itens. A não tendenciosidade reflete a escolha do texto dos itens e instruções, a qual deve primar pelos direitos humanos sendo livre, por exemplo, de quaisquer tipos de preconceitos. No caso das pessoas com deficiência, deve-se considerar as variações perceptivas e corporais uma vez que itens que desconsiderem tais aspectos facilmente incorrerão em erros. Independentemente do construto avaliado (personalidade, interesse, depressão, entre outros), o desenvolvedor deverá apropriar-se da cultura inerente à deficiência estudada. Contudo, a qualidade e a não tendenciosidade dependem não somente do conhecimento das deficiências, mas também do domínio teórico sobre o construto avaliado no instrumento.

A definição precisa do construto escolhido para o teste é condição *sine qua non* à elaboração de instrumentos de Testagem Universal. Na presente pesquisa o construto de interesse será a personalidade abordada na perspectiva da teoria fatorial e dos traços por meio do Modelo dos Cinco Grandes Fatores, quais sejam: Extroversão, Socialização, Neuroticismo, Realização e Abertura. Tal abordagem considera a personalidade como uma estrutura relativamente estável que quando avaliada por meio de instrumentos é capaz de identificar tendências comportamentais dos indivíduos. Contudo, a personalidade não é imutável, estando sujeita a variações devido a mudanças no meio social e ao desenvolvimento físico e psicológico dos indivíduos. Acresce-se que Os Cinco Grandes Fatores da Personalidade (*Big Five*) é um modelo empírico que deriva da análise fatorial remetendo a uma solução de explicação estatística dimensional que pressupõe que os indivíduos em geral apresentam algum nível em cada fator, sendo os níveis extremos tipicamente mal adaptativos (Silva & Nakano, 2011; Nunes, Hutz & Nunes, 2010).

O modelo dos Cinco Grandes Fatores não foi desenvolvido a partir de uma teoria, ele surgiu de estudos empíricos que uniram a Análise Fatorial, a Teoria dos Traços e a hipótese léxica (Nunes, Hutz, & Nunes, 2010; Pasquali, 2000). Existem diversas nomenclaturas para cada fator, porém a descrição de cada qual apresenta certa regularidade. Os Cinco Grandes Fatores de Personalidade – *Five Factor Model* são (Hutz *et al.*, 1998; McCrae & Oliver, 1990; Nunes, *et al.*, 2010):

1. Extroversão: Altos níveis neste fator descrevem um indivíduo ativo, assertivo, espontâneo, entusiasmado, falante, otimista, afetuoso, despreocupado, impulsivo e sociável.

2. Socialização: Altos níveis neste fator descrevem um indivíduo generoso, confiável, bondoso, prestativo, amigável, empático e altruísta.
3. Neuroticismo/Estabilidade Emocional – Neuroticismo: Altos níveis neste fator descrevem um indivíduo ansioso, tenso, sensível, instável, hostil, deprimido, vulnerável e baixa auto-estima./Estabilidade Emocional: Altos níveis neste fator descrevem um indivíduo confiante, emocionalmente estável, com alta autoestima, de humor positivo, relaxado e que lida bem com estresse.
4. Realização: Altos níveis neste fator descrevem um indivíduo organizado, persistente, eficiente, motivado, decidido, trabalhador, pontual, escrupuloso, ambicioso e perseverante.
5. Abertura: Altos níveis neste fator descrevem um indivíduo curioso, criativo, imaginativo, original e flexível.

Os instrumentos baseados no Modelo dos Cinco Grandes Fatores que estão aprovados para uso segundo o Sistema de Avaliação de Testes Psicológicos – SATEPSI, última atualização em julho de 2011, são o NEO PI-R (Inventário de Cinco Fatores NEO revisado), o ICFP-R (Inventário dos Cinco Fatores de Personalidade Revisado), a EFex (Escala Fatorial de Extroversão), a EFN (Escala Fatorial de Ajustamento Emocional/Neuroticismo), a EFS (Escala Fatorial de Socialização) e a BFP (Bateria Fatorial de Personalidade) (Andrade, J., 2008; CFP, 2011).

Evidências têm sido encontradas que reforçam a confirmação da hipótese transcultural do *Big Five* (McCrae & Terracciano, 2004). A transculturalidade do Modelo apoia o seu uso no desenvolvimento de instrumentos de Testagem Universal, já que atende ao primeiro princípio – *população ampla e inclusiva*. Assim, a presente pesquisa pretende desenvolver e buscar as primeiras evidências de validade de um teste psicológico informatizado de personalidade (modelo dos cinco grandes fatores) aplicando os princípios do Desenho Universal e a testagem adaptativa.

À guisa de alcançar uma população ampla e inclusiva almeja-se nesta pesquisa uma amostra de pessoas com deficiência visual e/ou motora além de pessoas sem deficiência.

As deficiências que comporão este estudo serão as deficiências visuais e as motoras. É considerada como deficiência visual a diminuição da resposta visual na qual mesmo após tratamento clínico ou cirúrgico ou uso de lentes corretivas perdura um comprometimento irreversível. A diferenciação entre deficientes visuais cegos e portadores de visão subnormal depende de duas escalas oftalmológicas: a acuidade

visual e o campo visual. Na cegueira ocorre uma alteração grave ou total em funções elementares da visão tais como: cor, distância, forma e tamanho. Para fins práticos considera-se cego o indivíduo que mesmo possuindo visão subnormal não é capaz de ler utilizando nenhum recurso óptico e que necessita do Braille ou de tecnologias assistivas para tal fim. Já o indivíduo com visão subnormal é aquele capaz de ler com o auxílio de potentes recursos ópticos ou através de impressões ampliadas (Amorin, 2006; Conde, 2005; Sá, Campos & Silva, 2007).

Existem várias definições teóricas para a cegueira, pois esta reúne indivíduos com diferentes graus de visão residual. A cegueira quando ocorre desde o nascimento é considerada congênita e, posteriormente a este, cegueira adquirida. A cegueira total – ou amaurose – pressupõe a completa perda da visão. Nestes casos nem a percepção luminosa existe. Assim, a cegueira é uma deficiência visual, mas existem os mais variados graus de deficiência visual e sua expressão também varia enormemente dependendo de fatores que vão além do simplesmente fisiológico (Amorin, 2006; Conde, 2005; Sá, Campos & Silva, 2007).

A deficiência visual não pode ser reduzida a um fenômeno fisiológico, já que há toda uma complexidade de fatores que a perpassam: sociais, afetivos, econômicos, culturais, políticos, artísticos, educacionais e tecnológicos. Indivíduos que não dispõem da visão, mas que receberam cuidados, educação e oportunidades de participar da vida social e cultural possuem um desenvolvimento bastante diferenciado e superior àqueles que não tiveram as mesmas oportunidades (Amorim, 2006; Kastrup, Sampaio, Almeida & Carijó, 2009).

Acresce-se que a maneira como o indivíduo cego percebe a deficiência visual irá influenciar na sua capacidade de desempenhar as atividades do dia-a-dia. Tal percepção está relacionada à sua história de vida e aprendizagem. Modelos familiares super-protetores influenciam negativamente o desempenho social do indivíduo, enquanto famílias que permitem e estimulam a independência do membro cego proporcionam seu desenvolvimento pessoal e social de forma mais efetiva (Amorim, 2006; Kastrup, Sampaio, Almeida & Carijó, 2009).

A deficiência física ou motora abrange tanto distúrbios da estrutura anatômica quanto alterações funcionais que interferem na movimentação e/ou locomoção dos indivíduos. Existem diferentes classificações em decorrência de aspectos como as causas, a topografia e os graus de acometimento. Quanto à natureza são: 1) As deficiências Musculares e/ou Neuromusculares que consistem em fraqueza muscular, paralisia ou falta de coordenação; 2) As deformações ósseas que afetam

principalmente os membros superiores e inferiores, a espinha e as articulações; e, 3) As limitações do vigor, da vitalidade e/ou da agilidade que implicam em vigor, vitalidade e agilidade altamente reduzidos (Brasil, 2006b).

As causas são provenientes de fatores genéticos, complicações durante a gestação, doenças ou acidentes. Assim, dividem-se em pré, peri e pós-natais: 1) As pré-natais abrangem problemas durante a gestação, como remédios tomados pela mãe e causas genéticas; 2) As perinatais, que ocorrem no período que antecede o parto até sete dias após o mesmo, envolvem, por exemplo, a prematuridade, o cordão umbilical enrolado ao corpo do bebê e problemas respiratórios; e, 3) As pós-natais, as quais abrangem nos bebês causas como as lesões cerebrais, a parada cardíaca e a meningite, enquanto que nos jovens e adultos algumas das causas comuns são as lesões medulares, os aneurismas e os acidentes vasculares cerebrais. Atualmente, com os modernos recursos de tecnologia assistiva disponíveis para este público, presencia-se a um aumento significativo na qualidade de vida destes indivíduos, os quais, dependendo do grau de comprometimento e comorbidades, atuam na sociedade em praticamente todos os espaços (Brasil, 2006b).

Portanto, para atender a este público amplo foi necessário agregar ao teste recursos de tecnologia assistiva, tais como: recursos em áudio, leitores de tela e auxílio de equipamento adaptativo – os quais passaram por pré-testes e estudos piloto. Ainda, complementa-se que os recursos de tecnologia assistiva tiveram que ser compatíveis com as ferramentas de programação computacional do instrumento para calibrar os itens e espera-se que também o sejam com o almejado instrumento de testagem adaptativa (ainda não abarcado neste estudo). Ao final desta tese alcançaram-se as primeiras evidências de validade do instrumento de calibração e concomitantemente do Banco de Itens para testagem adaptativa.

A validade de um teste é o grau com que as evidências coletadas corroboram a interpretação feita a partir dos resultados obtidos por meio de provas empíricas cientificamente consistentes. Desta forma, no processo de validação de um determinado teste as hipóteses interpretativas iniciais são avaliadas buscando-se evidências empíricas de validade. Tais evidências acumuladas apoiam a sua interpretação e seu uso em contextos e propósitos específicos. A qualidade do teste deve ser examinada diante do que ele pretende medir e de suas interpretações observando-se o grau em que a análise e comparação de seus elementos contribuem para que seja afirmada a sua validade – não

se trata do teste ser válido ou não, mas se as evidências são suficientes para comprovar a validade do instrumento diante de seus objetivos e das inferências feitas por meio dos resultados obtidos no teste (Alves, Souza & Baptista, 2011; American Educational Research Association *et al.*, 1999; Urbina, 2007).

A definição de validade evoluiu da perspectiva tripartite de Anastasi e Urbina (2000) a qual estava subdividida em validade de conteúdo, de critério e de construto para um conceito mais amplo. Nesta nova definição há a análise das evidências de validade dando-lhe flexibilidade, pois estudos são capazes de a todo tempo comprovar ou refutar a validade das interpretações de um teste, além do que o acúmulo de novos estudos de validade permite o aumento da credibilidade no tocante à validade do mesmo. Esta definição traz ainda um grande avanço quando considera como evidências de validade, além dos aspectos inerentes ao próprio teste e sua qualidade científica, as consequências da utilização destes instrumentos em longo prazo através de suas repercussões na sociedade, portanto, vislumbrando o uso ético dos testes psicológicos (Alves *et al.*, 2011; Primi, Muniz & Nunes, 2009).

Frente à nova definição de validade propôs-se que em vez de tipos de validade, fossem definidos fontes de evidência de validade. Segundo American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999), Urbina (2007), Primi, Muniz & Nunes (2009), Alves, Souza & Baptista (2011), tais fontes de evidência de validade são:

1. *Evidências Baseadas nas Relações com Outras Variáveis ou Evidências de Validade Baseadas nas Relações com Variáveis Externas*: Refere-se à verificação empírica da associação ou não associação entre os escores do teste a ser avaliado e de outras variáveis que medem o mesmo construto, construtos relacionados ou construtos diferentes.
2. *Evidências baseadas no conteúdo*: Levanta dados sobre os itens dos testes e se estão logicamente sendo capazes de medir as características psicológicas que se pretende avaliar.
3. *Evidências baseadas no processo de resposta*: Busca dados sobre os processos mentais que ocorrem durante a realização dos itens avaliando se estão em consonância com os objetivos do teste.
4. *Evidências baseadas nas consequências da testagem*: É o aspecto mais inovador da estruturação da validade em

evidências de validade, pois avalia a repercussão do teste na sociedade verificando se sua utilização está alcançando os objetivos desejados de acordo com a finalidade para a qual foi desenvolvido.

5. *Evidências baseadas na estrutura interna*: Levanta dados sobre a relação entre o teste e seus itens. Comumente realiza o estudo das estruturas de covariância entre partes do teste ou entre testes e subtestes. Para tanto são frequentemente utilizadas a análise fatorial, a análise da consistência interna e a modelagem com equações estruturais. No presente estudo será aplicada a Função Diferencial do Item e a Função Diferencial do Teste.

Testes com altos padrões de qualidade refletem a procura dos profissionais de Psicologia pela elevação da qualidade da atuação profissional. Assim, este estudo pretende contribuir para evolução da avaliação psicológica e sua qualidade técnica, bem como posteriores consequências práticas e éticas na atuação do profissional de Psicologia como promotor do bem-estar e da saúde. Destaca-se a contribuição desta pesquisa para que pessoas com deficiência possam beneficiar-se dos instrumentos psicológicos promovendo, assim, os direitos humanos. Portanto, tal tese de pesquisa está em consonância com a Área 3/Linha 3 do Programa de Pós-Graduação em Psicologia da Universidade Federal de Santa Catarina (PPGP/UFSC).

Acresce-se que testagem adaptativa figura entre as aplicações psicométricas mais inovadoras no campo dos testes psicológicos. A sua aplicação ocasiona um salto qualitativo na avaliação dos indivíduos. Logo, este estudo se coaduna com os avanços científicos na área da Psicometria pretendendo contribuir para o seu desenvolvimento. Agrega, ainda, ao uso da testagem adaptativa a aplicação dos pressupostos da Testagem Universal, o que reflete o ineditismo deste trabalho em decorrência da ausência de testes psicológicos no Brasil que utilizem a testagem adaptativa e o Desenho Universal simultaneamente.

O presente estudo faz parte das pesquisas desenvolvidas pelo Laboratório de Pesquisa em Avaliação Psicológica – LPAP/UFSC, estando vinculado aos projetos de construção, adaptação, validação e normalização de medidas psicológicas com o uso de novas tecnologias deste laboratório. Estes projetos têm por objetivo desenvolver, adaptar, validar e normatizar, para uso no Brasil instrumentos de avaliação de construtos sócio-emocionais com o uso de novas tecnologias. Tem-se trabalhado com o desenvolvimento de medidas que adotam o modelo

computadorizado adaptativo, bem como testes computadorizados desenvolvidos a partir dos princípios do Desenho Universal, estudos em andamento em pessoas com e sem deficiências. Dentre os construtos abarcados neste projeto constam personalidade, resiliência, *flow*, *mindfulness*, entre outros.

Para relatar esta tese, o texto foi organizado em cinco artigos, a saber:

1) O primeiro artigo descreve os desdobramentos teóricos alcançados pelo desenho universal, ressaltando o desenvolvimento de instrumentos em Psicologia – Testagem Universal. É delineado um roteiro para elaboração de testes segundo os princípios da Testagem Universal abrangendo a conceituação do teste, a escolha do número e tipo de itens, as análises e correções, a aplicação em campo, às análises qualitativas e quantitativas e a revisão final.

2) Já o segundo artigo versa sobre a construção do banco de itens para testagem adaptativa, a qual é uma tarefa de alta complexidade técnica e teórica. Neste artigo, descreve-se o desenvolvimento dos itens para três facetas em cada um dos Cinco Grandes Fatores segundo os princípios da Testagem Universal, a análise de juízes perpetrada, a escolha do Modelo BIB aplicado e, ainda, o desenvolvimento de um instrumento informatizado para coleta de calibração dos itens.

3) O terceiro artigo estuda a dimensionalidade do item por meio da análise fatorial com vistas a mensurar a adequação dos itens ao construto estudado.

4) No quarto artigo é apresentada a calibração nos itens, ou seja, os parâmetros dos itens são estudados.

5) Já o quinto e último artigo utiliza o Funcionamento Diferencial do Item para verificar a invariância da medida nos grupos com e sem deficiência. Tal aspecto é fundamental para analisar a aplicação da Testagem Universal nos itens para as pessoas com e sem deficiência.

OBJETIVOS

Geral

Buscar evidências de validade de um Banco de Itens de Personalidade para Testagem Adaptativa Desenvolvido a partir dos Princípios da Testagem Universal.

Específicos

- Delinear um roteiro para elaboração de testes segundo os princípios da Testagem Universal (Artigo 1);
- Desenvolver um banco de itens de personalidade e Testagem Universal (Artigo 2);
- Construir um instrumento para calibração dos itens segundo os princípios da Testagem Universal (Artigo 2);
- Verificar a dimensionalidade de um banco de itens de Personalidade e Testagem Universal (Artigo 3);
- Analisar a acessibilidade plena do instrumento segundo os princípios da Testagem Universal (Artigo 1 e Artigo 5);
- Comparar os resultados dos diferentes grupos no que se refere aos Cinco Fatores de Personalidade e a Testagem Universal (Artigo 5); e,
- Avaliar a qualidade dos parâmetros psicométricos dos itens e dos fatores, segundo os pressupostos da teoria de resposta ao item (Artigo 4 e Artigo 5).

Artigo 1. Avaliações de Testagem Universal: O Modelo.

Resumo

No presente artigo são descritos os desdobramentos alcançados pelo desenho universal, ressaltando o desenvolvimento de instrumentos em Psicologia – Testagem Universal. É delineado um roteiro para elaboração de testes segundo os princípios da Testagem Universal abrangendo a conceituação do teste, a escolha do número e tipo de itens, as análises e correções, a aplicação em campo, às análises qualitativas e quantitativas e a revisão final. Na Testagem Universal pessoas com e sem deficiência podem realizar os instrumentos de forma acessível procurando, assim, abarcar a diversidade humana. Uma das principais contribuições da Testagem Universal é a redução da necessidade de adaptações pós-elaboração do instrumental e consequente redução de custos com novos estudos. A tecnologia se torna uma grande aliada na aplicação dos princípios da Testagem Universal bem como, especificamente, as tecnologias assistivas – ambas são recursos potentes viabilizadores de acessibilidade em projetos de Testagem Universal. Complementarmente, são apresentados estudos internacionais e um estudo brasileiro de aplicação da Testagem Universal com resultados bastante promissores. Ainda, alguns desdobramentos dos modelos de Desenho Universal são apresentados seguidos da perspectiva da diversidade cognitiva em Testagem Universal e a possibilidade de mediação da interferência destas demandas no processo de testagem. Finaliza com a proposta de um Modelo Integrativo de Testagem Universal.

Palavras-chave: Testagem Universal; Desenho Universal; Diversidade Cognitiva; Modelo Integrativo.

Abstract

In the present article are described the unfolding achieved by the universal design, emphasizing the development of instruments in psychology - universal testing. It outlines a roadmap for the preparation of tests according to the principles of universal testing covering the concept of testing, the choice of number and type of items, analyzes and corrections, field application, qualitative and quantitative analyzes and final review. In universal testing, people with and without disabilities

can carry out the instruments in an accessible way, thus seeking to embrace human diversity. One of the main contributions of universal testing is the reduction of the need for post-instrument adaptation and consequent reduction of costs with new studies. Technology becomes a strong ally in the application of the principles of universal testing as well as, specifically, assistive technologies - both are powerful features of accessibility in universal testing projects. In addition, international studies and a Brazilian study of the application of universal testing are presented with very promising results. Still, some developments of the Universal Design models are presented followed by the perspective of cognitive diversity in universal testing and the possibility of mediating the interference of these demands in the testing process. It ends with the proposal of an Integrative Model of universal testing.

Keywords: Universal Testing; Universal Design; Cognitive Diversity; Integrative Model.

Introdução

O Desenho Universal surgiu no campo da arquitetura na busca de espaços que fossem pensados desde o início para todas as pessoas, sejam elas pessoas com deficiência ou sem deficiência, idosos e crianças. Mais recentemente (Thurlow, Lazarus, Christensen & Shyyan, 2016; Shinn & Ofiesh, 2012) foi incluído o pensar na diversidade cognitiva humana que abarcou indivíduos com TDAH (Transtorno de Déficit de Atenção e Hiperatividade), déficit cognitivo, dificuldades de aprendizagem, depressão, entre outros. Em sua expansão, o Desenho Universal alcançou produtos, projetos de espaços escolares até chegar ao ensino em si e concomitantemente aos processos avaliativos e seu instrumental. Assim, aporta-se na avaliação psicológica e nos testes psicológicos, foco principal deste artigo.

A aplicação do Desenho Universal à avaliação escolar, à avaliação em larga escala e à avaliação psicológica possui uma ampla literatura científica internacional (Burgstahler, 2001; Dolan, 2001; Thompson, Johnstone & Thurlow, 2002; Johnstone, 2003; Dolan, Hall, Barnerjee, Chun & Stangman, 2005; Hanna, 2005; Ketterlin-Geller, 2005; Almond, Winter, Cameto, Russell, Sato, Clarke-Midura, Torres, Haertel, Beddow & Lazarus, 2010; Albus & Thurlow, 2013; Thurlow,

Lazarus, Christensen & Shyyan, 2016, entre outros). Os desdobramentos teóricos que serão abordados neste artigo contribuem para o reconhecimento e atendimento tanto da variação corporal quanto da diversidade cognitiva humana.

O conceito de Desenho Universal reflete a busca pela acessibilidade plena a espaços, produtos, materiais de ensino e ao currículo de ensino incluindo pessoas com e sem deficiência e tenta contemplar a diversidade corporal e cognitiva humana em um mesmo projeto/produto/material. As *Diretrizes para Avaliação e Intervenção com indivíduos com Deficiência*³ da APA (*American Psychology Association*) do ano de 2009 pressupõem o uso do Desenho Universal na busca de soluções de acessibilidade para todos visando proporcionar aos indivíduos com deficiência a possibilidade de tomar suas próprias decisões, sem intervenções externas.

Tal conceito desdobrou-se em sete princípios os quais traduzem sua aplicação no campo da arquitetura e áreas afins (Burgstahler, 2001; Governo do Estado de São Paulo, 2010; Story, 2001):

1) o *uso equitativo*, o qual propõe espaços, objetos e produtos que possam ser utilizados por usuários com capacidades diferentes, tenta fornecer uso idêntico ou equivalente para uma ampla gama de usuários;

2) o *uso flexível*, o qual pretende criar ambientes ou sistemas construtivos que permitam atender as diferentes demandas dos usuários considerando as habilidades e preferências e possibilitando a adaptabilidade, e acomoda uma ampla variedade de preferências, habilidades individuais e variações corporais;

3) o *uso simples e intuitivo*, que visa a eliminar a complexidade desnecessária e permitir a fácil compreensão e apreensão do espaço independente da experiência do usuário, de seu grau de conhecimento, habilidade de linguagem ou nível de concentração;

4) a *informação de fácil percepção*, que procura utilizar diferentes meios de comunicação, como símbolos, informações sonoras, táteis, entre outras, para comunicar eficazmente a informação necessária ao usuário;

5) a *tolerância ao erro*, a qual busca a segurança minimizando perigos de ações acidentais ou não intencionais, procura desencorajar ações acidentais em tarefas que requerem vigilância;

³ Guidelines for Assessment of and Intervention with Individuals Who Have Disabilities.

6) o *esforço físico mínimo*, o qual procura dimensionar elementos e equipamentos para que sejam utilizados de maneira eficiente e segura, confortável e com o mínimo de fadiga; e,

7) o *dimensionamento de espaços para acesso e uso abrangente*, que visa a permitir o acesso e uso confortáveis para usuários, tanto sentados quanto em pé independentemente do tamanho do corpo do usuário, possibilitando o alcance visual dos ambientes e produtos a todos, acomodando variações ergonômicas e oferecendo condições de manuseio e contato para usuários com as mais variadas dificuldades de manipulação, toque e pegada, entre outras.

Quando o Desenho Universal por meio do seu conceito e princípios é aplicado à Avaliação Psicológica e em especial aos testes psicológicos obtém-se a Testagem Universal (*Universal Assessment*) (Thompson, Johnstone & Thurlow, 2002). Assim, na busca pela acessibilidade plena ou máxima possível aos processos avaliativos, os sete princípios do Desenho Universal foram transformados em sete princípios da testagem. Ressalta-se, entretanto, que não há uma correspondência de um para um, isto é, em um princípio da Testagem Universal pode ser contemplado um ou mais princípios do Desenho Universal. Os sete princípios da Testagem Universal em sua definição constitutiva são (Oliveira, 2013; Johnstone, 2003; Thompson, Johnstone & Thurlow, 2002):

1) *População de avaliação ampla e inclusiva*

Os testes psicológicos são desenvolvidos ou adaptados objetivando uma população ampla, geralmente pessoas com e sem deficiências. Atende a uma demanda crescente por maior inclusão da diversidade humana. Princípios do Desenho Universal refletidos: *Uso equitativo* e *Uso flexível*.

2) *Definição precisa do construto*

Os testes psicológicos são desenvolvidos ou adaptados visando à máxima definição dos construtos a serem medidos, evitando que possam ser afetados por outros construtos alheios ao que se pretende medir. Para tanto, faz-se imprescindível que os instrumentos empreguem teorias fortemente embasadas. Evita-se, assim, a construção de barreiras proporcionadas por conteúdo irrelevante semelhantemente às barreiras físicas que são evitadas nos projetos do Desenho Universal aplicados à arquitetura. Princípios do Desenho Universal refletidos: *Uso equitativo* e *Tolerância ao erro*.

3) *Itens acessíveis e não tendenciosos*

Os testes psicológicos são desenvolvidos ou adaptados para que indivíduos pertencentes a diferentes grupos (deficientes, estrangeiros,

sem deficiência, entre outros) com a mesma habilidade no construto avaliado tenham a mesma chance de acertar os itens do teste psicológico (em testes de desempenho) ou a apresentarem certos padrões de respostas (em testes de personalidade, interesse, entre outros), e que todos estes grupos compreendam as instruções dos procedimentos envolvidos na testagem. A construção dos itens e as instruções procuram evitar o viés cultural e alcançar a magnitude perceptiva do usuário. Princípios do Desenho Universal refletidos: *Uso equitativo* e *Uso flexível*.

4) *Testes flexíveis a acomodações*

Os testes psicológicos são desenvolvidos ou adaptados de tal forma que apliquem os princípios do Desenho Universal para os tornarem flexíveis a acomodações. Em alguns casos o testando, mesmo que o teste tenha sido construído segundo os princípios do Desenho Universal, necessita de adaptações extras. A vantagem do teste que foi construído segundo a Testagem Universal é permitir o uso destas acomodações com o mínimo de erro de validade e de modo que admita a comparabilidade dos resultados entre as aplicações com e sem adaptações extras. Princípios do Desenho Universal refletidos: *Uso equitativo*, *Uso flexível*, *Informação de fácil percepção* e *Dimensionamento de espaços para acesso e uso abrangente*.

5) *Instruções e procedimentos simples, claros e intuitivos*

Os testes psicológicos são desenvolvidos ou adaptados evitando-se que instruções e procedimentos em linguagem complexa ou confusa prejudiquem a compreensão do que é questionado, evitando que o testando incorra em erros por não compreender o que lhe é pedido, mesmo possuindo habilidade no construto medido. Princípios do Desenho Universal refletidos: *Uso equitativo*, *Uso simples e intuitivo*, *Informação de fácil percepção* e *Tolerância ao erro*.

6) *Leitura agradável e de máxima inteligibilidade*

Os testes psicológicos são desenvolvidos ou adaptados para reduzir a complexidade verbal e de organização textual dos itens e instruções preservando seu conteúdo essencial. Princípios do Desenho Universal refletidos: *Uso equitativo*, *Uso flexível* e *Uso simples e intuitivo*.

7) *Máxima legibilidade*

Os testes psicológicos são desenvolvidos ou adaptados de forma que a aparência física do texto, os gráficos, tabelas e ilustrações, bem como o formato das respostas possam ser percebidos e decifrados com facilidade. Princípios do Desenho Universal refletidos: *Uso equitativo*,

Uso simples e intuitivo, Informação de fácil percepção, Esforço físico mínimo e Dimensionamento de espaços para acesso e uso abrangente.

A Testagem Universal preocupa-se com o formato dos testes e estuda a sua influência na realização do mesmo e foca as adaptações como proporcionadoras de acessibilidade, analisando sua repercussão na qualidade dos instrumentos. Uma diferença básica entre Testagem Psicológica e Testagem Universal é que a Testagem Psicológica pretende tornar o teste claro e conciso para grupos populacionais específicos já a Testagem Universal busca um *design*, um formato de teste, que possa acomodar-se a uma população ampla com o máximo de acessibilidade possível. Na primeira, o testando é que tem de acomodar-se ao teste; na segunda, é o teste que se acomoda ao testando. Enfim, a Testagem Universal é mais flexível e abrangente (Oliveira, 2013).

Método

Foram realizadas pesquisas em bases de dados nacionais e internacionais, sobretudo nas segundas, uma vez que a literatura nacional é escassa nesta temática. A verificação das publicações conforme o ano, 1998 até 2016, utilizou, principalmente, as seguintes palavras-chave: *accessibility, universal design, universal assessment, cognitive diversity e disability*. Os textos que compõem o presente trabalho foram selecionados conforme a proximidade com a temática avaliação/testagem e Desenho Universal, abrangendo as relações e desdobramentos da interlocução entre ambas.

Resultados e Discussões

A Construção e Adaptação de instrumentos de Testagem Universal

Quando a Testagem Universal é aplicada ao desenvolvimento de instrumentos de avaliação desde o início de seu projeto, há a redução da necessidade de adaptações pós-elaboração e, conseqüente, redução de custos com novos estudos. Além disso, possibilita-se, o que é mais importante, uma avaliação de qualidade para uma população ampla incluindo pessoas com deficiência.

Os princípios da Testagem Universal podem ser aplicados aos testes de várias formas e em várias etapas da sua construção ou pós-construção. Quando os testes não foram construídos segundo os princípios da Testagem Universal um recurso que pode ser utilizado para potencializar sua acessibilidade é a adaptação do instrumento a estes

princípios. Para tanto, são utilizados como recursos as adaptações buscando a máxima eficiência do teste, utilizando-se de mais de uma adaptação em um mesmo teste ou do desenvolvimento de vários formatos para que o testando escolha segundo sua preferência. Contudo, a adaptação aos princípios da Testagem Universal não busca criar vantagens para um grupo ou outro, mas permitir através do uso de modificações nos itens, estrutura, formato e formas de aplicação que um número amplo de indivíduos acesse as informações requeridas pelo teste, livre de prejudicar-se pelas características estruturais destes instrumentos (Ketterlin-Geller, 2005).

A aplicação dos princípios da Testagem Universal à adaptação e construção de testes apresentam vantagens frente à utilização de acomodações isoladas. Os testes desenvolvidos ou adaptados com base na Testagem Universal são mais flexíveis às acomodações e ao uso de tecnologias assistivas e seus resultados apresentam maior confiabilidade em populações de alta diversidade e demandam menores custos na realização de estudos posteriores a construção (Ketterlin-Geller, 2005; Dolan, Hall, Banerjee, Chun & Strangman, 2005; Johnstone, 2003; Thompson, Johnstone & Thurlow, 2002). Contudo, como aplicar os princípios da Testagem Universal com vistas a alcançar um maior número de recursos auxiliares e potencializadores da capacidade de proporcionar acessibilidade?

O uso do computador pode ser um grande aliado para os desenvolvedores de testes que procuram aplicar os princípios da Testagem Universal na construção de instrumentos mais acessíveis, uma vez que permite criar um ambiente que proporciona maiores possibilidades de variação, isto é, maior flexibilidade no uso, escolha de formatos, tamanhos de fontes, além do fato de que admite o acoplamento de tecnologias assistivas que venham a complementar e potencializar a Testagem Universal. Porém, a Testagem Universal pode também ser aplicada satisfatoriamente ao formato tradicional – lápis e papel (Ketterlin-Geller, 2005; Dolan & Hall, 2001). O ponto chave está nas possibilidades viabilizadas pelo computador que são superiores às do formato lápis e papel. Ressalta-se, entretanto, que se o testando não estiver familiarizado com o computador este poderá se transformar em barreira ao invés de um recurso vantajoso.

No *A State Guide to the Development of Universally Designed Assessments* (2006) é descrito um roteiro para elaboração de testes segundo os princípios da Testagem Universal, o qual será exposto de forma resumida nos parágrafos subsequentes.

Na construção de testes de Testagem Universal cada detalhe de formatação e escolha de design é observado procurando a máxima acessibilidade possível. Contudo, a aplicação dos princípios de acessibilidade plena aos instrumentos não se resumem aos aspectos da estética sendo uma tarefa complexa que parte de uma definição precisa do construto. As etapas envolvidas na elaboração de instrumentos de Testagem Universal estão apresentadas na Figura 1 (Oliveira, 2013; Johnstone, Altman & Thurlow, 2006).

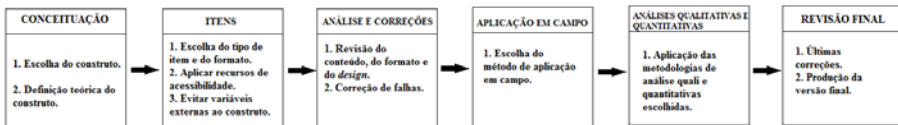


Figura 1. Elaboração de Instrumentos de TU

1) *A conceituação do teste*: com a escolha do construto a ser medido e a sua definição explícita e precisa. Alguns questionamentos norteadores para construção de instrumentos de Testagem Universal: Este construto é realmente relevante? Aplica-se para esse público? Possui literatura que o embasa?

2) *Escolha do número e tipo de itens, minimização do efeito de fatores externos ao construto*: no que se refere ao formato, cada item deve ser desenvolvido de forma a ser compatível com recursos de acessibilidade. Já no tocante ao conteúdo, deve-se evitar qualquer conteúdo que possa oferecer vantagem ou desvantagem para qualquer subgrupo da amostra ao qual se destina o teste. Isto implica em possuir na equipe de desenvolvimento do instrumento pessoas especializadas na construção de itens para grupos variados. A fim de minimizar o efeito de fatores externos (por exemplo: evitar o uso de gráficos ou figuras que não sejam transponíveis para o Braille, ou evitar complexidade linguística desnecessária), deve-se prever uma gama de desempenhos para prevenir os efeitos de teto ou piso e desenvolver um conjunto de itens em número suficiente para permitir a eliminação caso sejam inapropriados. Ainda, faz-se importante aplicar a técnica do “*Think Aloud*”, a qual é uma técnica de pensar em voz alta o conteúdo do item para verificar como o mesmo é compreendido pela população alvo e, assim, se ele mede realmente o construto desejado. Alguns questionamentos norteadores: O conteúdo de algum item oferece vantagem ou desvantagem a um grupo específico? Os itens são acessíveis ou possuem recursos proporcionadores de acessibilidade?

3) *Análise e correções*: consiste em submeter o instrumento a uma equipe de revisão de testes. A equipe deve possuir indivíduos peritos com características semelhantes à população a que se destina o teste, pessoas com as diferentes deficiências alvo, os quais sejam pesquisadores e profissionais especialistas no construto a ser medido. Nesta etapa é realizada a revisão de conteúdo, formato e *design* do instrumento, e, deste modo, a análise e correção de possíveis falhas até o momento (por exemplo: verificação do funcionamento de *softwares* leitores de tela ou outro recurso de leitura do instrumento, como voz gravada ou vídeo). Além da revisão pela equipe de pesquisa, é importante a avaliação do instrumento por uma equipe de juízes externos, o que implica na escolha da metodologia da Análise de Juízes. Alguns questionamentos norteadores: Os recursos de acessibilidade estão funcionando corretamente? O formato está adequado? O conteúdo se coaduna com o objetivo desejado? O design do instrumento está em consonância com os princípios do desenho universal/Testagem Universal?

A seguir, apresenta-se um *checklist* dos principais elementos de Desenho Universal/Testagem Universal com vistas a auxiliar o processo de revisão do teste desenvolvido pelos autores deste artigo.

CHECKLIST DE TESTAGEM UNIVERSAL		
Público ao qual o teste se destina: _____		
Avaliador: _____	Data: _____	
Conteúdo dos Itens	Atende	Não atende
<ol style="list-style-type: none"> 1. O texto do item é de fácil entendimento. 2. As palavras utilizadas são de fácil entendimento por grupos variados. 3. O texto é curto. 4. O conteúdo é equitativo para pessoas com e sem deficiência. 5. O texto do item está livre de ambiguidade. 6. O texto é lido com qualidade pelo software leitor de tela. 7. O texto se adequa aos recursos de Tecnologia Assistiva que serão utilizados. 		
Instruções		
<ol style="list-style-type: none"> 1. O texto das instruções é de fácil entendimento. 2. As palavras utilizadas são de fácil entendimento por grupos variados. 3. O texto é curto. 4. As instruções possuem conteúdo equitativo para pessoas com e sem deficiência. 		

<ol style="list-style-type: none"> 5. O texto das instruções está livre de ambiguidade. 6. O texto das instruções é lido com qualidade pelo software leitor de tela. 7. O texto se adequa aos recursos de tecnologia assistiva que serão utilizados. 8. As instruções são compreensíveis para os mais variados tipos de leitores (Ex. cegos, videntes, surdos, etc.). 		
<p>Design do Instrumento</p> <ol style="list-style-type: none"> 1. O <i>design</i> é compatível com os recursos de Tecnologia Assistiva planejados. 2. O <i>design</i> do instrumento favorece a tolerância ao erro. 3. O <i>design</i> proporciona a fácil percepção do conteúdo do instrumento. 		
<p>Formato dos itens</p> <ol style="list-style-type: none"> 1. O formato é acessível. 2. O formato dos itens é flexível aos recursos de Tecnologia Assistiva. 3. O formato é equitativo para pessoas com e sem deficiência. 4. O formato dos itens proporciona tolerância ao erro. 		

4) *A aplicação em campo*: pode assumir vários modelos, como por exemplo, ser realizada para uma população de pessoas com ou sem deficiência ou para grupos específicos. A escolha do modelo de estudo de campo ou estudo-piloto e, conseqüentemente, da amostra dependerá dos objetivos do teste, do construto medido e da amplitude do público alvo, como por exemplo, pessoas com deficiências variadas e pessoas sem deficiência – ambas de vários grupos populacionais. Perguntas Norteadoras: Como será o estudo-piloto? Qual a amplitude da amostra?

5) *Análises qualitativas e quantitativas*: As análises qualitativas estão voltadas para a acessibilidade e usabilidade do instrumento e como estas são percebidas por seus usuários. Já as análises quantitativas permeiam a utilização de recursos estatísticos variados incluindo estatísticas descritivas e multivariadas e, ainda, a Teoria de Resposta ao Item (TRI), a qual é reconhecida como uma técnica de análise estatística valiosa nos estudos da aplicação do Desenho Universal a instrumentos de avaliação, sendo também possível utilizar-se de técnicas da Teoria Clássica dos Testes (TCT). Perguntas Norteadoras: Quais análises qualitativas serão úteis para este instrumento? Quais análises quantitativas se aplicam a esse instrumento?

6) *Revisão final*: são realizados os últimos ajustes do instrumento e produz-se a sua versão final. Perguntas norteadoras: A versão final alcançou o objetivo desejado? Será necessário retornar a alguma etapa anterior para alcançar a versão final?

O roteiro para elaboração de testes segundo os princípios da Testagem Universal apresentado nos parágrafos anteriores apresenta de um plano preliminar e deve ser aplicado e avaliado após a utilização na elaboração de instrumentos. O roteiro dá uma direção, um parâmetro, ao mesmo tempo em que é um auxiliar do método científico, pois permite a replicação em novos estudos. Tal roteiro objetiva, deste modo, o alcance da aplicação dos princípios da Testagem Universal com a máxima qualidade.

O papel da tecnologia como viabilizadora da Testagem Universal

A tecnologia contribui enormemente para tornar os testes acessíveis. Ferramentas informáticas simples, tais como programas de acessibilidade e sintetizadores de voz, podem ser utilizados para viabilizar o atendimento aos princípios da Testagem Universal. Na interface entre tecnologias e deficiências têm-se a tecnologia assistiva, que abrange um conjunto de instrumentos e técnicas desenvolvidos especificamente para pessoas com deficiências com o objetivo de auxiliá-los a desempenhar suas atividades diárias com a maior independência possível.

O destaque da tecnologia assistiva é em grande parte reflexo dos avanços tecnológicos gerais e das Tecnologias de Informação e Comunicação (TIC). Entretanto, a tecnologia assistiva é tão antiga quanto à fabricação dos primeiros instrumentos pelo homem. São considerados recursos de tecnologia assistiva tanto uma bengala quanto um aparelho de amplificação utilizado por uma pessoa com surdez moderada. Soluções simples e artesanais são consideradas pela sua utilidade para as pessoas com deficiência como tecnologia assistiva sem necessariamente possuírem um grande rebuscamento tecnológico (ABNT, 2004; Galvão Filho, 2009).

No caso específico das pessoas com deficiência auditiva, na utilização do computador e da internet é comum aliar-se vídeos em Libras aos recursos apresentados em Língua Portuguesa ou, até mesmo, programas tradutores (Língua Portuguesa – Língua Brasileira de Sinais) com vistas a alcançar a máxima acessibilidade (Rosa & Cruz, 2001). Ainda diante do apresentado nos parágrafos anteriores, outros exemplos de tecnologias assistivas para os diversos tipos de deficiência são:

brinquedos e roupas adaptadas, computadores, *softwares* e *hardwares* com recursos de acessibilidade, dispositivos para adequação da postura sentada, recursos para mobilidade manual e elétrica, equipamentos de comunicação alternativa, equipamentos de comandos computacionais por voz, chaves e acionadores especiais, aparelhos de escuta assistida, auxílios visuais, materiais protéticos, entre outros. Portanto, as tecnologias assistivas são úteis como mediadoras da acessibilidade favorecendo o Desenho Universal/Testagem Universal.

Testagem Universal: pesquisas e instrumentos

Nos Estados Unidos o Desenho Universal é aplicado desde 2004 às avaliações em larga escala no contexto educacional com o intuito de possibilitar que tanto os estudantes com deficiência quanto os sem deficiência as realizem de forma equivalente (Johnstone, Thompson, Moen, Bolt & Kato, 2005; Johnstone, Altman & Thurlow, 2006). Tais avaliações possuem certa semelhança ao ENEM (Exame Nacional do Ensino Médio), o qual é realizado por uma grande parcela dos estudantes brasileiros concluintes do Ensino Médio. Entretanto, no Brasil a acessibilidade da prova é reduzida, restringindo-se a um pequeno número de adaptações, como, por exemplo, a adaptação para o Braille, a utilização de uma versão ampliada da prova ou a disponibilização de um interprete de Libras (Brasil, 2013).

Em concordância com o parágrafo anterior, os instrumentos encontrados nos estudos internacionais que abordam a aplicação do conceito de Desenho Universal ao desenvolvimento ou adaptação de testes, estão, em sua maioria, relacionados ao contexto educacional e voltados principalmente para avaliação em larga escala, como: o *Stanford 10*, o *Test of 3rd grade mathematics ability* e o CBT-TTS (*National Assessment of Educational Progress United States history and civics testcomputer-based system*).

Os três instrumentos citados têm como aspectos comuns o fato de que na sua adaptação (CBT-TTS) ou desenvolvimento (*Test of 3rd grade mathematics ability* e *Stanford 10*) utilizou-se o computador como plataforma acrescida de recursos em áudio. É importante destacar que as adaptações realizadas nos mesmos não se restringiram à forma de aplicação, mas que os próprios itens dos testes foram desenvolvidos ou adaptados com base nos princípios da Testagem Universal (Case, 2003; Ketterlin-Geller, 2005; Dolan, Hall, Banerjee, Chun & Strangman, 2005).

Johnstone (2003) realizou uma pesquisa que teve como objetivo verificar se a eliminação de material irrelevante, isto é, a aplicação de alguns princípios da Testagem Universal à construção de instrumentos de avaliação, torna-os medidas mais precisas do desempenho dos alunos. Um teste foi construído no formato tradicional e um segundo teste foi criado a partir do primeiro, mas incluindo aspectos da Testagem Universal. Um especialista avaliou a correspondência do construto medido nos instrumentos item a item. Os resultados da pesquisa assinalaram que os alunos, independente da sua variação corporal ou diversidade cognitiva, pontuaram significativamente de forma mais elevada no teste de Testagem Universal dando indícios do aumento da qualidade do instrumento com a aplicação de tais princípios.

Ketterling-Geller (2005) desenvolveu um estudo no qual foram respeitadas todas as etapas para a construção de instrumentos de Testagem Universal, e pode-se destacar dentre as contribuições da Testagem Universal utilizadas pelos autores as seguintes:

a) A plataforma do teste computadorizada foi desenvolvida por programadores e especialistas com vistas a permitir que alunos com poucas habilidades no manuseio do computador pudessem utilizá-lo sem dificuldade;

b) O aluno poderia retornar ao item anterior para revê-lo ou corrigi-lo;

c) O teste possuía acomodações de suporte geral sendo flexível, caso fosse necessário, a acomodações adicionais para grupos específicos;

d) O teste agregava tecnologia relacionada à Testagem Adaptativa (*Computerized adaptive testing* – CAT) com o objetivo de maximizar a eficiência e aumentar a precisão das estimativas de habilidade;

e) Foram utilizados itens de múltipla escolha com vistas a evitar complexidade desnecessária. Tal recurso justifica-se porque o construto medido pelo teste era *matemática*, assim, na presença de excesso de texto o indivíduo necessitará de habilidades adicionais que não fazem parte do construto em questão, como por exemplo, compreensão textual; e,

f) O formato do teste viabilizava redundância de material, o qual está em consonância com a Testagem Universal, pois é possível acessar a informação tanto auditivamente quanto por texto escrito.

Ainda no mesmo estudo do parágrafo anterior, a validade foi verificada por meio da validade de conteúdo, de critério e de construto. Professores peritos analisaram o conteúdo de cada item seguindo-se à

exclusão ou correção de falhas. Os resultados do estudo foram analisados utilizando o modelo de 2-parâmetros da Teoria de Resposta ao Item (TRI). Complementarmente, realizou-se um estudo qualitativo com membros da comunidade educacional sobre se o formato do teste era realmente acessível. Os resultados encontrados neste estudo apontaram para a criação de um instrumento realmente mais acessível para uma população ampla com e sem deficiência (Ketterling-Geller, 2005).

O quarto estudo e último estudo desta seção, perpetrado por Dolan, Hall, Banerjee, Chun e Strangman (2005), consistiu na adaptação do teste *National Assessment of Educational Progress United States history and civics* de uma versão lápis e papel para uma informatizada. A principal adaptação analisada no estudo foi o recurso de áudio (leitura do texto escrito). Alguns aspectos da Testagem Universal aplicados ao instrumento adaptado (CBT-TTS) foram: a) Apresentação de uma pergunta de cada vez; b) A possibilidade de marcar itens para serem revistos posteriormente através de uma célula à direita do item (*Review later*); c) A adaptação para o *read-aloud* (leitura em voz alta); d) A possibilidade de ler e rever, ouvir e reouvir o áudio; e, e) Uma barra de navegação, a qual permitia que o aluno acompanhe o seu progresso no teste. Os resultados indicaram aumento significativo nos escores do teste na versão computadorizada adaptada a Testagem Universal quando comparada com a ferramenta tradicional para perguntas com mais de 100 palavras.

Os estudos acima descritos refletem avanços no campo do instrumental para avaliação das pessoas com deficiência. Todavia, as conclusões dos estudos apresentam limitações tais como os construtos estudados, os quais foram restritos ao contexto educacional não abarcando construtos próprios da Psicologia. Outra limitação clara é a dificuldade de proporcionar a acessibilidade plena almejada diante da variabilidade de habilidades e necessidades humanas carecendo de aprofundamento e incremento no número de estudos sobre o tema. Com vistas a propor possibilidades para algumas destas limitações foi perpetrado um estudo de aplicação da Testagem Universal no Brasil.

Aplicação da Testagem Universal no Brasil

Oliveira (2013), seguindo a perspectiva da Testagem Universal, realizou o estudo da adaptação de um teste psicológico para avaliação da personalidade informatizado à Testagem Universal. O teste de personalidade informatizado adaptado procurou atender aos 7 princípios

do Desenho Universal/Testagem Universal. Tal instrumento foi submetido a uma amostra de 146 indivíduos com deficiência visual e 150 indivíduos sem deficiência. Complementarmente, foi desenvolvido um instrumento para verificar a adequação do referido instrumento às adaptações realizadas.

A Bateria Informatizada para Avaliação da Personalidade, instrumento que teve o seu formato adaptado aos princípios do Desenho Universal/Testagem Universal, baseia-se no modelo dos cinco grandes fatores da personalidade (Extroversão, Socialização, Neuroticismo, Realização e Abertura). Seus itens são de autorrelato e fazem parte do banco de itens do Laboratório de Pesquisa em Avaliação Psicológica – LPAP da Universidade Federal de Santa Catarina - UFSC.

Complementarmente foi desenvolvido um instrumento para avaliação da aplicação do Desenho Universal primando por avaliar os princípios do Desenho Universal/Testagem Universal. As adaptações foram realizadas na estrutura do formato de aplicação do teste: 1. Orientações dadas para execução do teste; 2. Formato dos itens; 3. Tamanho da fonte e tipo de letra; 4. Cores e espaços; e, 5. Ordem de apresentação - objetivando atender ao público com deficiência visual. Apesar das adaptações priorizarem o atendimento ao público com deficiência visual, foi objetivo do estudo desenvolver um formato do teste para o Desenho Universal/Testagem Universal e, desta forma, os recursos foram desenvolvidos para atender também às pessoas sem deficiência.

As pessoas com deficiência visual frequentemente empregam os leitores de tela quando da utilização do computador e da *Internet*. Os leitores de tela são um tipo de tecnologia assistiva proporcionada por sistemas computacionais. Utilizando-se deste recurso o indivíduo com deficiência visual consegue ter acesso aos conteúdos escritos da *Internet*, pois uma voz sintetizada os lê transformando-os de texto escrito para áudio-texto. Os leitores possuem diversos recursos de mediação desta navegação, os quais são realizados por meio de comandos digitados no teclado.

A forma adaptada e o instrumento para avaliação do Desenho Universal/Testagem Universal foram testados pelos pesquisadores com cinco tipos de leitores de tela/sintetizadores de voz (*Jaws*, *Dosvox*, *NVDA*, *Virtual Vision* e *Voice Over*) com o objetivo de verificar se os itens se adequavam a estes recursos de Tecnologia Assistiva, ou seja, se o instrumento atendia ao quarto princípio da Testagem Universal - *Testes flexíveis a acomodações* (e aos princípios do Desenho Universal: *Uso equitativo*, *Uso flexível*, *Informação de fácil percepção* e

Dimensionamento de espaços para acesso e uso abrangente). Durante este processo, palavras e formatos que não eram flexíveis aos leitores foram modificados.

Os instrumentos e seus vários formatos possíveis (itens, cores e fontes) foram submetidos primeiramente à equipe de pesquisa composta por dois pesquisadores sem deficiência visual e dois especialistas cegos, sendo uma psicóloga e um professor de informática para deficientes visuais. Foram tomadas decisões, coletadas sugestões e realizadas correções preliminares pela equipe. Formatos variados de itens foram testados e, posteriormente, foi definida uma forma adaptada para o estudo-piloto.

O estudo-piloto seguiu a lógica da Análise de Juízes em dois momentos. No primeiro momento foi aplicado a cinco indivíduos com alta escolaridade, e, no segundo momento, a cinco indivíduos com baixa escolaridade. Novas correções foram realizadas sendo obtida a forma final do instrumento, o qual foi aplicado na amostra de pesquisa. Ressalta-se ainda, que a escolha nesta pesquisa pelo modo de apresentação do teste adaptado como uma ferramenta informatizada *online* perpassou por vários princípios do Desenho Universal, pois permitiu o acesso ao teste por pessoas com e sem deficiência utilizando seu computador pessoal independentemente da distância que os indivíduos estavam do local de origem da pesquisa (Universidade Federal de Santa Catarina - UFSC) viabilizando um uso flexível e equitativo.

Os resultados do instrumento que avaliou a adaptação do instrumento para adaptação da Testagem Universal apontaram para o atendimento de todos os sete princípios empregados e concomitantemente a acessibilidade do instrumento para indivíduos com e sem deficiência visual. Alcançou-se a acessibilidade máxima possível para uma população ampla e inclusiva. No entanto, o teste adaptado à Testagem Universal possuiu como limitação o fato de não aplicar-se a outros tipos de deficiência, o qual, para tanto, necessitará de novos estudos.

Vários autores (Amiralian, 1997; Kastrop, Sampaio, Almeida & Carijó, 2009; Case, Zucker & Jeffries, 2005) apontam para a ineficiência das adaptações realizadas em testes psicológicos para os indivíduos com deficiência visual uma vez que os mesmos geralmente partem da perspectiva dos videntes. A pesquisa de Oliveira (2013) relatada buscou partir da perspectiva do deficiente visual, de sua forma de utilização do computador, dos leitores de tela e como o uso destas tecnologias repercutem nos formatos dos itens e na forma de apresentação das

instruções. Considera-se que tal ponto de partida contribuiu para o alcance dos resultados positivos reportados, ou seja, a acessibilidade segundo os princípios do Desenho Universal/Testagem Universal.

Desdobramentos dos modelos de Desenho Universal

O Desenho Universal possui caráter multidisciplinar alcançando várias áreas, dentre elas a educacional (Desenho Universal na Instrução, Desenho Universal na Aprendizagem e Testagem Universal). O Desenho Universal na Instrução (DUI) tem como objetivo principal potencializar a aprendizagem para uma ampla gama de alunos com as mais variadas características (Burgstahler, 2012; Burgstahler, 2001; Silver, Bourke & Strehorn, 1998). Aplica os princípios do DU a todos os aspectos da instrução, desde o espaço físico até os métodos educacionais (Burgstahler, 2012; Burgstahler, 2001).

Na aplicação dos princípios do Desenho Universal à instrução tem-se duas classificações distintas. A primeira é a de Burgstahler (2007): 1) Clima da Classe: Optar por práticas que reflitam a diversidade e a inclusão dos alunos; 2) Interação: Proporcionar interações entre os alunos e entre estes e os professores e garantir a acessibilidade da comunicação; 3) Ambientes e produtos físicos: garantir a utilização e a acessibilidade de ambientes e todos os produtos utilizados pelos alunos; 4) Métodos de ensino: Utilizar vários métodos de ensino acessíveis permitindo que os alunos escolham entre várias opções que favoreçam a aprendizagem; 5) Recursos de informação e tecnologia: Utilizar materiais envolventes, flexíveis e acessíveis para todos os alunos; 6) *Feedback*: fornecer *feedback* regularmente durante as atividades escolares; 7) Avaliação: avaliar periodicamente o progresso dos alunos utilizando vários métodos e ferramentas acessíveis. 8) Acomodação: plano de novas acomodações para os alunos que não forem abarcados pelo design instrucional.




Semelhantemente à classificação de Burgstahler (2007), Valle e Connor (2014) apresentam a seguinte: 1) No uso equitativo: “A instrução é projetada para ser útil e acessível às pessoas com capacidades diversas”. Por exemplo: áudio-livros e *softwares* leitores de tela para alunos cegos ou disléxicos.; 2) Flexibilidade de uso: “A instrução é projetada para acomodar uma ampla variedade de capacidades individuais”. Por exemplo: os conteúdos das aulas podem ser estudados por meio de livros, documentos, sites, entrevistas, filmes, etc.; 3) Simples e intuitivo: “A instrução é fácil de entender, independentemente do conhecimento, da experiência, das habilidades

linguísticas e do nível atual de concentração do estudante”. Por exemplo: instruções claras para todas as tarefas, simples de acompanhar sejam elas apresentadas na forma oral e escrita, ou em ambas.; 4) Informações perceptíveis: “*A instrução é projetada de modo que as informações necessárias sejam comunicadas de forma bem-sucedida aos estudantes, independentemente de onde eles estejam ou quais sejam as suas capacidades sensoriais*”. Por exemplo: Cópias digitais, impressas e ampliadas podem ser disponibilizadas. 5) Tolerância ao erro: “*O instrutor prevê variações no ritmo de aprendizagem individual dos alunos e nas habilidades necessárias*”. Por exemplo: os alunos entregam os trabalhos em partes obtendo *feedback* após cada entrega. 6. Baixo esforço físico: “*A instrução é projetada para minimizar o esforço físico desnecessário, a fim de permitir o máximo de atenção voltada à aprendizagem*”. Por exemplo: utilização de um processador de palavras nas provas. 7) Tamanho e espaço para abordagem e uso: “*A instrução é projetada de modo a considerar o tamanho e o espaço apropriados para a abordagem, alcance, manipulação e uso, independentemente do tamanho, da postura ou da mobilidade do corpo do aluno*”. Por exemplo: utilização de equipamentos ajustáveis.; 8) Uma comunidade de alunos: “*O ambiente instrucional promove interação e comunicação entre os estudantes e entre estudantes e docentes*”. Por exemplo: todos sabem os nomes de todos e as habilidades de cada um são reconhecidas; 9) Clima instrucional: “*A instrução é projetada para ser acolhedora e inclusiva*”. Por exemplo: o *feedback* é constante e específico a cada indivíduo.

No que se refere ao Desenho Universal para aprendizagem (Universal Design for Learning – UDL), este faz parte do DUI, porém corresponde à sua aplicação específica ao currículo. Preocupa-se, sobretudo, com currículos que permitam a todas as pessoas, com entusiasmo, adquirirem conhecimentos e habilidades: “*almeja manter altos padrões de desempenho para todos*” (Burgstahler, 2012; Burgstahler, 2007). Ainda, o UDL exige a integração de vários meios de representação, ação, expressão e engajamento no currículo do curso. Assim, o currículo deve: (a) fornecer múltiplas representações de conteúdos, (b) fornecer várias opções de expressão, e (c) oferecer várias opções para o engajamento (Rose, Meyer & Hitchcock, 2005; Wehmeyer, 2006).

O estudo de como a aprendizagem ocorre é o que sedimenta o UDL e partiu do uso de tecnologias de imageamento cerebral, como a tomografia por emissão de pósitrons (PET), ressonância magnética funcional (fMRI) e eletroencefalografia quantitativa (EEGq), os quais

revelam como o cérebro funciona durante a aprendizagem (Meyer & Rose, 2005). Por meio destas técnicas de imagem cerebral foi observada a atividade em três conjuntos de redes nervosas que desempenham um papel primordial na aprendizagem, que são (vide Figura 2): as redes de reconhecimento, de estratégias e as afetivas. As primeiras, redes de reconhecimento, são especializadas em receber e analisar informação (o "o quê" da aprendizagem), enquanto as redes de estratégias são especializadas em planejar e executar ações (o "como" da aprendizagem) e as redes afetivas especializam-se em avaliar e definir prioridades (o "porquê" da aprendizagem) (Dolan & Hall, 2001; Rose & Meyer, 2002; Meyer & Rose, 2005).

Redes de Reconhecimento	Redes de Estratégia	Redes Afetivas
Aprender o QUÊ	Aprender COMO	Aprender POR QUÊ
		
Como reunimos factos e categorizamos o que vemos, ouvimos e lemos. A identificação de letras, palavras ou o estilo de um autor são tarefas de reconhecimento.	Planejar e desempenhar tarefas. Como organizamos e expressamos as nossas ideias. Escrever um texto ou resolver um problema de matemática são tarefas estratégicas.	Como os alunos se empenham e se mantêm motivados. Como reagem aos desafios, se estimulam e interessam. Estas são dimensões afetivas.
➡ Apresente a informação e os conteúdos em diferentes formatos	➡ Diversifique os modos como os alunos podem expressar o que sabem	➡ Estimule o interesse e a motivação por aprender
Mais formas de promover Múltiplos Meios de Representação	Mais formas de promover Múltiplos Meios de Ação e Expressão	Mais formas de promover Múltiplos Meios de Envolvimento

Fonte: CAST: What is UDL? (<http://cast.org/research/udl>)

Figura 2. Papel das redes nervosas na aprendizagem

No âmbito das três grandes redes, Meyer e Rose (2005) descrevem que existe um modo geral de funcionamento: “*uma assinatura da atividade no cérebro*”. Entretanto, as áreas cerebrais específicas dentro de cada uma das três redes dependem da frequência de exposição e da extensão da atividade cerebral, compondo um mapa único que, portanto, varia de indivíduo para indivíduo. O uso destas novas tecnologias de imagem cerebral também tem mostrado que o cérebro se modifica quando submetido a experiências de aprendizagem. Mesmo em adultos, as áreas de processamento individual podem crescer ou encolher conforme a experiência. Deste modo, o cérebro é muito plástico tendo na repetição e na prática mudanças que não estão restritas ao nível comportamental, sob a forma de um melhor desempenho

observável, mas que ocorrem também ao nível neuronal (Meyer & Rose, 2005).

Como síntese das aplicações do Desenho Universal e consequentes modelos e novos modelos baseados neste, apresenta-se o Quadro 1.

Quadro 1.

Desdobramentos dos modelos de Desenho Universal

	Desenho Universal (Universal Design)	Desenho Universal na Instrução (Universal Design of Instruction)	Desenho Universal na Instrução (Universal Design of Instruction)	Desenho Universal na Aprendizagem (Universal Design for Learning)	Testagem Universal (Universal Assessment)
1	<i>Uso equitativo.</i>	<i>Uso equitativo aplicado à instrução.</i>	<i>Clima da Classe.</i>	<i>Fornecer Múltiplas representações de conteúdo.</i>	<i>População de avaliação ampla e inclusiva.</i>
2	<i>Uso flexível.</i>	<i>Uso flexível aplicado à instrução.</i>	<i>Interação.</i>	<i>Fornecer várias opções de expressão.</i>	<i>Definição precisa do construto.</i>
3	<i>Uso simples e intuitivo.</i>	<i>Uso simples e intuitivo aplicado à instrução.</i>	<i>Ambientes e produtos.</i>	<i>Oferecer várias opções de engajamento.</i>	<i>Itens acessíveis e não tendenciosos.</i>
4	<i>Informação de fácil percepção.</i>	<i>Informação de fácil percepção aplicada à instrução.</i>	<i>Métodos de ensino.</i>	----	<i>Testes flexíveis a acomodações.</i>
5	<i>Tolerância ao erro.</i>	<i>Tolerância ao erro aplicada à instrução.</i>	<i>Recursos de informação e tecnologia.</i>	----	<i>Instruções e procedimentos simples, claros e intuitivos.</i>
6	<i>Esforço físico mínimo.</i>	<i>Esforço físico mínimo aplicado à instrução.</i>	<i>Feedback.</i>	----	<i>Leitura agradável e de máxima inteligibilidade.</i>
7	<i>Dimensionamento de espaços para acesso e uso abrangente.</i>	<i>Dimensionamento de espaços para acesso e uso abrangente aplicado à instrução.</i>	<i>Avaliação.</i>	----	<i>Máxima legibilidade.</i>
8	----	<i>Uma comunidade de alunos.</i>	<i>Acomodação.</i>	----	----
9	----	<i>Clima instrucional.</i>	----	----	----

Neurocognição e Avaliações de Desenho Universal

A diversidade cognitiva segundo Shinn e Ofiesh (2012) consiste nos diferentes traços cognitivos (por exemplo: memória, atenção e velocidade de processamento) de cada indivíduo. Os referidos autores preocuparam-se com a variabilidade cognitiva no contexto da aprendizagem assumindo que tal complexidade cognitiva inclui pessoas com e sem deficiências e ainda indivíduos com dislexia, TDAH, depressão, déficits cognitivos, etc. Observa-se que a ideia de uma diversidade cognitiva possui ampla consonância com a perspectiva do Desenho Universal uma vez que toca à diversidade humana e procura incluí-la em vários contextos, sobretudo o educacional. A partir deste último, alcança os processos avaliativos e seus instrumentais.

Shinn e Ofiesh (2012) dividem as demandas cognitivas relacionadas ao processo de testagem em três categorias: 1) Acessar o teste - o conteúdo e formato (*test access*): a compreensão da linguagem, o raciocínio visuoespacial e a fluência no conteúdo do teste são as características cognitivas mais exigidas; 2) O processo de responder ao teste (*test output*): Integração visuo-motora e a memória de longo prazo; e, ainda, 3) Que participam das duas categorias anteriores, ou seja, demandas cognitivas que estão presentes no processo de acessar e responder ao teste: Memória de trabalho, atenção e velocidade de processamento. Observa-se que tal categorização expressa as áreas que estão primariamente relacionadas aos processos cognitivos. Contudo, tanto no acessar como no responder serão necessárias a interações de todas estas demandas cognitivas.

As demandas cognitivas parecem variar em um espectro no qual está presente tanto o processo de acessar o teste como de respondê-lo (Figura 3). O desempenho do indivíduo será baixo caso o indivíduo consiga acessar o teste, mas não consiga respondê-lo e também se ele não conseguir nem mesmo acessar o instrumento. O acréscimo das demandas cognitivas à execução de instrumentos decorre, entre outras variáveis, do aumento da dificuldade do teste. Espera-se, e até se deseja, que os testes ou seus itens/questões demandem mais cognitivamente na medida em que as questões se tornem mais difíceis. Contudo, questões fáceis ou testes fáceis podem se tornar difíceis devido a outros fatores, como o formato implementado, o vocabulário utilizado, o número de palavras, etc. Quanto menos acessíveis forem os testes, mais eles irão demandar cognitivamente de modo indesejado (Figura 3). Assim, testes

acessíveis possibilitam que indivíduos com os mais variados níveis de proficiência no construto medido o realizem, medindo realmente o que se pretende medir sem a interferência de variáveis estranhas; este o intuito da aplicação do Desenho Universal aos instrumentos avaliativos, dentre eles, os testes psicológicos.



Figura 3. Tendência da relação entre acessibilidade e demandas cognitivas.

Principais demandas cognitivas e a Testagem Universal

A *compreensão da linguagem* abarca a linguagem falada e escrita e processos como a identificação, a decodificação e/ou a leitura de palavras (Cardoso-Martins, 2008). Para acessar o instrumento é necessário que o testando entenda o que é pedido nos enunciados das questões e as possibilidades de resposta. Para isso, ele poderá ler os enunciados ou utilizar recursos como os *softwares* leitores de telas ou processadores de palavras. Vislumbrando alcançar tal demanda cognitiva com menor interferência da mesma no construto medido, aplica-se a Testagem Universal, por exemplo, no planejamento do texto que será apresentado, tornando-o descritivo e consequentemente de possível compreensão quando convertido de texto em fala.

Na interação da demanda cognitiva da *compreensão da linguagem* com os princípios da Testagem Universal buscar-se-á: 1) *População de avaliação ampla e inclusiva*: elaborar um modelo de texto falado e/ou escrito que inclua indivíduos com os mais variados graus de compreensão da linguagem desde indivíduos com déficits nesta demanda até os proficientes; 2) *Definição precisa do construto*: evitar textos que não tenham uma relação clara com o construto que será medido, bem como palavras rebuscadas, técnicas ou regionalismos que não façam parte dos objetivos do processo avaliativo; 3) *Itens acessíveis e não tendenciosos*: instruções claras para os itens; 4) *Testes flexíveis a acomodações*: desenvolver o instrumento desde o início objetivando à utilização de recursos de acessibilidade que permitam a compreensão do

texto; 5) *Instruções e procedimentos simples, claros e intuitivos*; 6) *Leitura agradável e de máxima inteligibilidade*; e, 7) *Máxima legibilidade*: Estes três princípios irão se refletir na escolha das palavras (evitando palavras de duplo sentido ou que tenham um sentido coloquial diferente para algum dos grupos avaliados), no número de palavras utilizadas, (ou seja, evitando textos desnecessários), e, ainda, no tipo de letras utilizadas e seus tamanhos. Pode-se apontar como exemplos de transtornos ou comprometimentos associados a esta demanda cognitiva os seguintes: Dificuldades de Aprendizagem, Transtornos de Linguagem e Déficits Cognitivos (vide Tabela 1).

O *raciocínio visuoespacial* “*refere-se à capacidade de perceber, analisar, sintetizar e pensar com padrões visuais, incluindo a capacidade de armazenar e recuperar representações visuais*” (Mather & Jaffe, 2002). O mesmo envolve a identificação do formato do instrumento, como ele é organizado e os tipos de questões (múltipla escolha, questões abertas, etc.). Tal demanda requer inúmeros desafios por parte dos desenvolvedores de instrumentos de Testagem Universal, os quais refletem principalmente os seguintes princípios (Valle & Connor, 2014; Thurlow, Lazarus, Abus & Hodgson, 2010; Anderson-Inman & Horney, 2007; Ofiesh, Rojas & Ward, 2006): 1) *População de avaliação ampla e inclusiva*: vislumbrar os diferentes graus de raciocínio visuoespacial na população que será submetida ao instrumento; 4) *Testes flexíveis a acomodações*: permitir adaptações, como o aumento das letras ou mudança da fonte em testes informatizados (Fuchs, Fuchs, Eaton, Hamlett, Binkley e Crouch (2000) recomendam o tamanho 14 ao invés do 12) e disponibilizar vários formatos de apresentação do instrumento (ex. impresso, em Braille, informatizado e em áudio) e, ainda, que os itens, em seus formatos, devem adequar-se a recursos de tecnologia assistiva; 5) *Instruções e procedimentos simples, claros e intuitivos*: tornar os procedimentos intuitivos requer, entre outros aspectos, colocar instruções e legendas diretamente sobre o texto ou questão a qual se aplicar, e, ainda prover uma navegação intuitiva (em testes informatizados) nos menus, links, conteúdos e questões; 6) *Leitura agradável e de máxima inteligibilidade*; e, 7) *Máxima legibilidade*: evitar desordem no *design*, organizar os elementos textuais deixando espaço apropriado entre questões, textos e imagens e, ainda, escolher o tipo de letras, tamanhos e cores primando pela legibilidade. Pode-se apontar como exemplos de transtornos ou comprometimentos associados a esta demanda cognitiva os seguintes: Déficits cognitivos (vide Tabela 1).

A *Fluência no conteúdo* se refere ao domínio de assuntos/conteúdos/habilidades/construtos partindo-se do simples para o complexo. Abarca o quanto o indivíduo consegue acessar dos conteúdos para realizar o teste, por exemplo, em um teste de raciocínio não verbal (R1), no qual as questões iniciais são mais fáceis espera-se que quanto maior for a habilidade do indivíduo, mais questões ele responda (por dominar o construto). Contudo, outras demandas cognitivas podem diminuir a probabilidade do indivíduo responder ao teste, mesmo que ele saiba a resposta. Esse último caso ocorre, por exemplo, quando o indivíduo tem uma velocidade de processamento muito lenta necessitando de um tempo mais prolongado para responder a um mesmo número de questões, ou quando ocorre um déficit atencional e o indivíduo incorre em muitos erros, não porque não sabia do conteúdo, mas porque se distraiu de detalhes importantes do instrumento. A fluência no construto/conteúdo relaciona-se aos sete princípios da Testagem Universal, na medida em que, quando forem empregados, o indivíduo expressará por meio do instrumento sua real fluência no construto – seja ela baixa ou alta. Tal demanda está associada a vários transtornos, mas não tem uma relação de especificidade que caracterize um em particular.

A *velocidade de processamento* refere-se à capacidade de processar ou fazer sentido das informações de entrada e, em seguida, produzir uma resposta (Shinn & Ofiesh, 2012). Em se tratando de velocidades de processamentos mais baixas, faz-se necessário que seja resguardado o quarto princípio da TU (*Testes flexíveis a acomodações*). Aos instrumentos, é possível acoplar-se tecnologias assistivas que auxiliem no processo de percepção e compreensão das palavras, digitação de texto, sobretudo em se tratando de testes informatizados. Em casos nos quais o indivíduo possua uma velocidade de processamento da informação baixa será necessário, ainda, aumentar o tempo de realização do instrumento ou mesmo fragmentá-lo em partes. Os princípios 5) *Instruções e procedimentos simples, claros e intuitivos*; 6) *Leitura agradável e de máxima inteligibilidade*; e, 7) *Máxima legibilidade* irão contribuir para que o indivíduo realize o instrumento, mesmo possuindo uma velocidade de processamento relativamente baixa, porém não vão mudar o fato do indivíduo possuir tal velocidade. Pode-se apontar como exemplos de transtornos ou comprometimentos associados a esta demanda cognitiva os seguintes: TDAH, Depressão, dificuldade de aprendizagem, Ansiedade e déficits cognitivos (vide Tabela 1).

A *memória de trabalho* é a recordação enquanto o trabalho ainda está sendo feito, e é uma das demandas cognitivas mais importantes possuindo uma ampla relação com as demais. Nos casos em que o teste seja demasiado complexo para execução, exija um grande número de ferramentas, demande muitos passos durante a execução, etc., os indivíduos com problemas na memória de trabalho terão dificuldade em executá-los e poderão ainda incorrer em erros não pelo desconhecimento do conteúdo, mas por não conseguirem se apropriar do formato do instrumento. Segundo Thurlow, Lazarus e Hodgson (2010) avaliações computadorizadas podem exigir muito da memória de trabalho. Assim, simplicidade e clareza são fundamentais tanto no formato informatizado quanto no impresso para que o teste seja realmente acessível. Para tanto, faz-se útil a aplicação dos seguintes princípios da Testagem Universal: 5) *Instruções e procedimentos simples, claros e intuitivos*; 6) *Leitura agradável e de máxima inteligibilidade*; e, 7) *Máxima legibilidade*. O objetivo principal da aplicação dos referidos princípios da Testagem Universal é diminuir as demandas cognitivas relacionadas à memória de trabalho no processo de realização dos instrumentos. A menos que o construto medido seja a própria memória de trabalho, não se deseja a interferência deste construto em outros. O aumento do tempo de execução também pode ser uma medida importante em alguns casos para diminuir o efeito da memória de trabalho em outros construtos. Pode-se apontar como exemplos de transtornos ou comprometimentos associados a esta demanda cognitiva os seguintes: TDAH, depressão, dificuldade de aprendizagem, ansiedade e déficits cognitivos (vide Tabela 1).

A *atenção* e, sobretudo, o processo de mantê-la é crítico para execução dos testes. Conforme McCabe, Roediger, McDaniel, Balota, & Hambrick (2010), a atenção “*é importante à manutenção de foco por um longo período de tempo, a filtragem da informação relevante e o processo de ignorar o que não é necessário*”. Com vistas a alcançar uma 6) *Leitura agradável e de máxima inteligibilidade* e 7) *Máxima legibilidade* nos aspectos específicos ao processo atencional, aponta-se que informações importantes deverão estar em destaque e serem de fácil localização, que marcadores nos textos e questões facilitam a percepção, informações redundantes quando bem empregadas são úteis (por exemplo, informar a página e o número de questões em todas as telas de testes informatizados), que deve-se evitar informações e textos desnecessários e, ainda, que destaques variados do que não pode ser esquecido são fundamentais. Ressalta-se a importância da tolerância ao erro, por exemplo, programando-se os testes informatizados para

destacarem questões que não foram respondidas permitindo que os indivíduos retornem a elas quando desejarem. Pode-se apontar como exemplos de transtornos ou comprometimentos associados a esta demanda cognitiva os seguintes: TDAH, depressão e déficits cognitivos (vide Tabela 1).

Tabela 1.
Transtornos ou comprometimentos associados.

	Atenção	Memória de Trabalho	Velocidade de Processamento	Raciocínio visuo-espacial	Compreensão da Linguagem.
TDAH	X	X	X		
Depressão	X	X	X		
Dificuldade de Aprendizagem: Dislexia, disgrafia, etc.		X	X		X
Transtornos de Linguagem: afasias					X
Ansiedade		X	X		
Déficits cognitivos: Deficiência Intelectual	X	X	X	X	X

Modelo Integrativo de Testagem Universal

Objetivando sintetizar as principais ideias que permeiam este artigo com vistas a alcançar um modelo integrativo de Testagem Universal, faz-se imprescindível o conhecimento das variáveis que compõem o processo de testagem. Assim, foram selecionadas algumas das variáveis para que o testando acesse e responda a um instrumento agrupando-as nas seguintes categorias (vide Figura 4).



Figura 4. Variáveis do processo de testagem.

1) Demandas Cognitivas: envolvem todos os aspectos descritos no tópico anterior deste artigo (memória de trabalho, atenção, velocidade de processamento, etc.). Tais demandas causam interferência no processo de testagem por serem ao mesmo tempo mediadoras (o processamento da informação depende de cada uma delas) e executoras (a realização da tarefa também é dependente de cada uma delas, bem como o resultado final) na execução de instrumentos avaliativos.

2) Habilidades motoras: possibilitam a expressão da cognição em sua amplitude, quando o meio não é acessível. Indivíduos com deficiência que afetam suas habilidades motoras, mesmo possuindo alto nível de proficiência no construto medido ficam impossibilitados de demonstrar suas habilidades. Por exemplo, um grande cientista com deficiência, como Stephen Hawking¹, sem acessibilidade ficaria

impossibilitado de demonstrar suas reais habilidades em um teste de aritmética simples, no formato lápis e papel.

3) Órgãos dos sentidos e percepções: os órgãos dos sentidos tem um papel crucial na percepção do que é apresentado. Caso o teste não vislumbre as diferentes possibilidades de percepção do material (por exemplo: visual, auditiva, tátil, em Braille e em língua de sinais), ele não será acessível para a diversidade cognitiva humana.

4) Aprendizagem: as formas como os indivíduos aprendem irão influenciar o modo como eles respondem aos testes, por exemplo: aprendizagem guiada por imagens, aprendizagem guiada por textos, percepção das partes primeiro ou percepção do todo primeiro, etc. Ademais, o fato de já ter passado por tarefas semelhantes irá influenciar os seus resultados em testes futuros, ou seja, em atividades similares. Meyer e Rose (2005) apontam que *“o cérebro do indivíduo é capaz de generalizar, gastando menos esforço para processar as exigências de uma tarefa com a qual ela já tenha lidado muitas vezes”*.

5) Ambiente: um meio estimulante, desencadeante de novos desafios, aumenta o sentimento de autoeficácia dos indivíduos sejam eles com ou sem deficiência. Assim, o estímulo à realização de atividades variadas e das próprias atividades da vida diária (AVDs) podem contribuir para o engajamento dos indivíduos em atividades futuras. Para tais atividades, sejam novas ou repetições de tarefas já realizadas, espera-se que um ambiente estimulante de testagem favoreça a realização destes instrumentais. Ainda, experiências em meios acessíveis podem estimular o sentimento de realização do indivíduo, mudando o paradigma de compreensão do espaço e de si mesmo, ou seja, a inacessibilidade está no meio externo e não no indivíduo por tratar-se de pessoa com deficiência ou com uma demanda cognitiva específica.

6) Meio Familiar: que promove a independência e o protagonismo e possui qualidades que o aproximam do desenho universal. Assim, o meio familiar favorece e media processos promotores de acessibilidade por encararem o indivíduo como capaz impedindo a estagnação. Espera-se que o meio familiar seja um promotor de acessibilidade por excelência uma vez que é nele que o indivíduo se desenvolve. Como tal, o engajamento do indivíduo e sua familiaridade com recursos de acessibilidade irá favorecer a experiência e o engajamento na busca do Desenho Universal e concomitantemente da Testagem Universal.

7) Tecnologias Assistivas: responsáveis pela mediação do processo de acessibilidade. Coadunam-se com a acessibilidade plena

proposta pelo Desenho Universal/Testagem Universal uma vez que, quando inteligentemente utilizadas, permitem que os indivíduos realizem as mais diversas atividades sem a necessidade de auxílio externo, isto é, independentemente. São recursos potentes na elaboração de projetos de DU/TU tanto no que se refere a espaços e produtos quanto aos testes psicológicos.

As categorias descritas contribuem para a compreensão do modelo de Testagem Universal, o qual não depende de duas variáveis apenas, quais sejam: o teste e o indivíduo, de forma alguma. O processo de testagem é bem mais complexo e a criação de ferramentas que primam pela qualidade para grupos amplos requer o desdobrar de um olhar ampliado para o fenômeno da testagem e da avaliação psicológica.

Conclusões

O presente artigo detalhou o surpreendente alcance dos princípios do Desenho Universal, inicialmente elaborados com o foco em projetos de arquitetura. Os mesmos evoluíram teoricamente e mostraram a utilidade de sua aplicação em outras áreas. Na Educação, compreendeu um campo vasto com aplicações que vão desde o espaço físico até o currículo e o processo de aprendizagem. As aplicações do Desenho Universal aos testes psicológicos e ao processo de testagem foram os enfoques principais deste trabalho, os quais se tratam de um campo amplo de possibilidades, criador de uma forma diferenciada de vislumbrar a testagem para além do momento ou fração temporal em que ele ocorre, visando, assim, abarcar as inúmeras variáveis que podem interferir neste processo. Faz-se, para tanto, de fundamental importância uma metodologia de construção de testes de Testagem Universal que permitam a segurança técnica necessária a esta tarefa.

Conforme apresentado, as etapas envolvidas na construção de testes de Testagem Universal são semelhantes às etapas já estandardizadas na construção de instrumentos psicológicos. Acresce-se que o conteúdo de cada etapa ou suas sub etapas assumem maior riqueza e detalhamento de aspectos estruturais, sobretudo em relação ao formato e a sua adequação à utilização de tecnologias assistivas. Em se tratando da Testagem Universal, torna-se uma preocupação permanente a acessibilidade do instrumento tanto em relação ao formato quanto ao seu conteúdo, bem como a operacionalização dos princípios da Testagem Universal em cada etapa do processo de construção. Consequentemente, a Testagem Universal permite a ampliação da qualidade dos instrumentos para grupos amplos.

A Testagem Universal ultrapassou a visão do testando estático e em seus desdobramentos mais atuais considerou as demandas cognitivas dos indivíduos. Tais demandas possibilitaram a ampliação do grau de detalhamento do processo de realização de testes, dividindo-o no processo de acessar e de responder ao teste, uma vez que apesar destes dois processos possuírem algumas demandas cognitivas que se assemelham, possuem inúmeras outras demandas bastante diversas. Logo, o pensar em uma população ampla e inclusiva de Testagem Universal ultrapassa o pensar apenas em nível de pessoas com e sem deficiência, mas abarca a diversidade cognitiva humana e reforça o compromisso ético dos profissionais de Psicologia com a qualidade destes materiais. Espera-se, por fim, que este trabalho contribua para ampliar o olhar em relação à elaboração de instrumentos em Psicologia e que a Testagem Universal alcance não só os instrumentos, mas as atitudes e o pensamento dos profissionais de Psicologia, os quais incluam estudos desta natureza quando do processo de desenvolvimento de seus instrumentais.

Referências

- Albus, D., & Thurlow, M. L., (2013). *Accommodation policies for states' alternate assessments based on alternate achievement standards (AA-AAS)* (Synthesis Report 90). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2010). Technology-Enabled and Universally Designed Assessment: Considering Access in Measuring the Achievement of Students with Disabilities—A Foundation for Research. *Journal of Technology, Learning, and Assessment*, 10(5). Retrieved [date] from <http://www.jtla.org>.
- Anderson-Inman, L., & Horney, M.A., (2007). Supported e-text: Assistive technology through text transformations, *Reading Research Quarterly*, 42, 153-160. Recuperado em julho de 2016 de, http://ncset.uoregon.edu/files/pdf/supported_etext.pdf.
- Armstrong, T. (2012). Students Who Challenge Us: First, Discover Their Strengths.. *Educational Leadership*. Vol. 70, nº. 2. p. 10-16. Recuperado em julho de 2016 de,

<http://www.ascd.org/publications/educational-leadership/oct12/vol70/num02/First,-Discover-Their-Strengths.aspx>.

- Associação Brasileira de Normas Técnicas, ABNT (2004). *Acessibilidade a edificações, mobiliário, espaços e equipamentos urbanos*. NBR 9050/2004.
- Brasil (2013). *Edital nº 01, de 08 de maio de 2013*. Exame nacional do ensino médio. Brasília: Instituto nacional de estudos e pesquisas educacionais Anísio Teixeira. Ministério da educação. Recuperado em dezembro de 2013, de <http://goo.gl/XDbQqw>.
- Brasil (2013). *Edital nº 01, de 08 de maio de 2013*. Exame nacional do ensino médio. Brasília: Instituto nacional de estudos e pesquisas educacionais Anísio Teixeira. Ministério da educação. Recuperado em dezembro de 2013, de <http://goo.gl/XDbQqw>.
- Burgstahler, S. (2001). *Equal Access: Universal Design of Instruction*. University of Washington. DO-IT (Disabilities, Opportunities, Internetworking, and Technology). Recuperado em Agosto de 2012 de, http://www.washington.edu/doit/Brochures/Academics/equal_access_udi.html.
- Burgstahler, S. (2007). *Universal Design of Instruction (UDI): Definition, Principles, Guidelines, and Examples*. University of Washington. DO-IT (Disabilities, Opportunities, Internetworking, and Technology). Recuperado em junho de 2016 de, <http://www.washington.edu/doit/universal-design-instruction-udi-definition-principles-guidelines-and-examples>.
- Burgstahler, S. (2012). *Equal access: Universal design of instruction*. University of Washington. DO-IT (Disabilities, Opportunities, Internetworking, and Technology). Recuperado em Agosto de 2016 de, <https://students.case.edu/academic/disability/faculty/doc/UniversalDesign.pdf>.
- Cardoso-Martins, C. (2008). Desenvolvimento das habilidades de leitura e escrita. In Fuentes, D., Malloy-Diniz, L. F., Camargo, C. H. P. & col. *Neuropsicologia: Teoria e prática*. Porto Alegre: Artmed.

- Case, B. J. (2003). *Universal Design*. Pearson: policy report. Recuperado em março de 2011, de www.pearsonassessments.com/UniversalDesign.
- Dolan, R. P. & Hall, T. E. (2001). Universal Design for Learning: Implications for Large-Scale Assessment. *IDA Perspectives* 27(4): 22-25. Recuperado em maio de 2011, de <http://www.cast.org/system/galleries/download/byCAST/udlassesment.pdf>.
- Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment*, 3(7). Recuperado em maio de 2011, de <http://www.jtla.org>.
- Fuchs, L., Fuchs, D., Eaton, S. Hamlett, C., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children*, 67, 67-81.
- Galvão Filho, T. A. A. (2009). Tecnologia Assistiva: de que se trata? In: MACHADO, G. J. C.; SOBRAL, M. N. (Orgs.). *Conexões: educação, comunicação, inclusão e interculturalidade*. 1 ed. Porto Alegre: Redes Editora, p. 207-235, 2009.
- Hanna, E. L. (2005). *Inclusive Design for Maximum Accessibility: a practical approach to Universal Design*. Pearson Educational Measurement. Recuperado em Março de 2011, de http://www.pearsonassessments.com/NR/rdonlyres/BB1BC770-BCC1-4F06-9DA5-8D19A81E6C49/0/RR_05_04.pdf.
- Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved. Recuperado em Agosto de 2011, de <http://goo.gl/jTwZLk>.
- Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved. Recuperado em Agosto de

- 2011, <http://education.umn.edu/NCEO/OnlinePubs/Technical37.htm>.
- Johnstone, C. J., Thompson, S. J., Moen, R. E., Bolt, S., & Kato, K. (2005). *Analyzing results of large-scale assessments to ensure universal design*. Technical Report 41. Minneapolis, MN: University of Minnesota National Center on Educational Outcomes.
- Johnstone, C., Altman, J. & Thurlow, M. (2006). A state guide to the development of universally designed assessments. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retirado de http://www.niusileadscape.org/docs/FINAL_PRODUCTS/LearningCarousel/UDmanual.pdf.
- Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning and Assessment*, 4(2). Recuperado em agosto de 2011, de <http://www.jtla.org>.
- Korbin, J., & Young, J. (2003). The cognitive equivalence of reading comprehension test items via computerized and pencil-and-paper administration. *Applied Measurement in Education*, 12, 11-14.
- Mather, N., & Jaffe, L. (2002). *Woodcock-Johnson III: Reports, recommendations, and strategies*. New York: Wiley & Sons.
- McCabe, D., Roediger, H., McDaniel, M., Balota, D., & Hambrick, D. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24, 222–243.
- Meyer, A., & Rose, D. H. (2005). The future is in the margins: The role of technology and disability in educational reform. In D. H. Rose, A. Meyer & C. Hitchcock (Eds.), *The universally designed classroom: Accessible curriculum and digital technologies* (pp. 13-35). Cambridge, MA: Harvard Education Press. Recuperado em junho de 2016 de, http://www.udlcenter.org/sites/udlcenter.org/files/Meyer-Rose_FutureisintheMargins.pdf.
- Ofiesh, N., Rojas, C., & Ward, R. (2006). Universal design and the assessment of student learning in higher education: promoting

thoughtful assessment. *Journal of Postsecondary Education and Disability*, 9, 173- 181.

- Oliveira, C. M. ; Nuernberg, A. H. & Nunes, C. H. S. S. (2013). Desenho universal e avaliação psicológica na perspectiva dos direitos humanos. *Avaliação Psicológica*. Vol. 3. <http://goo.gl/qtkhx3>.
- Rosa, A., C., & Cruz, C., C., (2001). Internet: Fator de Inclusão da Pessoa Surda. *Rev. Online da Bibl. Prof. Joel Martins*, Campinas, v.2, n.3, p.38-54 , jun.2001. Recuperado em junho de 2013 de, <http://periodicos.sbu.unicamp.br/ojs/index.php/etd/article/view/580/595>.
- Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal Design for Learning*. Alexandria, VA: ASCD.
- Rose, D., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Shinn, E. & Ofiesh, N. S. (2012). Cognitive Diversity and the Design of Classroom Tests for All Learners. *Journal of Postsecondary Education & Disability*. Vol. 25. Issue 3, p.227. Recuperado em outubro de 2015, de <http://files.eric.ed.gov/fulltext/EJ994288.pdf>.
- Silver, P., Bourke, A., & Strehorn, K. C. (1998). Universal instructional design in higher education: An approach for inclusion. *Equity & Excellence in Education*,31(2), 47-51. Recuperado em agosto de 2016 <http://www.tandfonline.com/doi/pdf/10.1080/1066568980310206?needAccess=true>.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [today's date], from the World Wide Web: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>.
- Thurlow, M. L., Lazarus, S. S., Christensen, L. L., & Shyyan, V. (2016). *Principles and characteristics of inclusive assessment systems in a changing assessment landscape* (NCEO Report 400). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

- Thurlow, M., Lazarus, S. S., Albus, D., & Hodgson, J. (2010). *Computer-based testing: Practices and considerations* (Synthesis Report 78). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Valle, J. W., & Connor, D.J. (2014). *Ressignificando a Deficiência: da abordagem social às práticas inclusivas na escola*. Porto Alegre: AMGH. 240p.
- Wehmeyer, L., Garner, N., Yeager, D., Lawrence, M., & Davis, A., K., (2006). Infusing Self-Determination into 18 - 21 Services for Students with Intellectual or Developmental Disabilities: A Multi-Stage, Multiple Component Model. *Education and Training in Developmental Disabilities*, 41(1), 3-13. Recuperado em julho de 2016 de http://www.beachcenter.org/Research/FullArticles/PDF/Wehmeyer_et_al_2006.pdf.

Artigo 2. Construção de um Banco de Itens de Personalidade e Testagem Universal em Testagem Adaptativa.

Resumo

A elaboração de um banco de itens em testagem adaptativa é uma tarefa de alta complexidade que demanda amplo conhecimento técnico e teórico. Soma-se a tal fato a elaboração do banco de itens segundo os princípios da Testagem Universal do construto personalidade, cinco grandes fatores. Assim, o objetivo geral deste estudo consistiu em desenvolver um banco de itens de personalidade e Testagem Universal e os específicos foram construir itens segundo os 7 princípios da Testagem Universal, propor um delineamento para análise de juízes e construir um instrumento para calibração dos itens. Foram desenvolvidos itens para cada uma das 3 facetas dos 5 Grandes Fatores. Aplicação dos 7 princípios da Testagem Universal. Realizou-se as análises referentes a Testagem Universal e a análise de juízes. A organização da análise de juízes deu-se da seguinte forma: 600 itens em blocos de 150, cada dupla de juízes analisou 300 itens. Resultados das análises de juízes: reescritura, exclusão de itens, mudança de faceta. Os itens foram implementados no programa Concerto seguindo o Modelo de BIB escolhido para organização do banco. Ao final deste processo foi desenvolvido um instrumento informatizado para coleta de calibração dos itens.

Palavras-chave: Banco de Itens; Testagem Universal; Testagem Adaptativa; Deficiência; Personalidade.

Abstract

The elaboration of a bank of items in adaptive testing is a task of high complexity that demands ample technical and theoretical knowledge. Add to that fact the elaboration of the bank of items according to the principles of the universal test of the personality construct, five major factors. Thus, the general objective of this study was to develop a Bank of Personality and Universal Test items and the specific ones were to construct items according to the 7 principles of Universal Testing, to propose a design for the analysis of judges and to construct an instrument for calibration of the items. Items were developed for each of the 3 facets of the 5 Major Factors. Application of the 7 principles of Universal Testing. Analyzes were carried out regarding Universal Testing and the analysis of judges. The organization of the analysis of judges was as follows: 600 items in blocks of 150, each pair of judges

analyzed 300 items. Results of the judges' analysis: rewriting, exclusion of items, change of facet. The items were implemented in the Concerto program following the BIB Model chosen for the organization of the bank. At the end of this process a computerized instrument was developed to collect items calibration.

Keywords: Bank of Items; Universal Testing; Adaptive Testing; Deficiency; Personality.

Introdução

Testagem Adaptativa Computadorizada, banco de itens, Blocos Balanceados Incompletos, Testagem Universal, acessibilidade plena, deficiência e variabilidade humana, são alguns dos temas abordados neste estudo. Tal artigo é desafiador, uma vez que pretende agregar ao estudo das medidas em Psicologia questões profundas e atuais calcadas na acessibilidade plena aplicada a instrumentos avaliativos: a Testagem Universal, a qual teve suas origens nos princípios do Desenho Universal. Deste modo, a preocupação principal é: como unir na construção de um banco de itens para testagem adaptativa os princípios da Testagem Universal?

Na construção de testes psicológicos um desafio proeminente é o desenvolvimento de itens que sejam capazes de avaliar o público alvo do instrumento. Caso os itens sejam muito fáceis para os testandos, em testes de desempenho, ou pouco representativos, em testes de personalidade ou interesse, os resultados agregam pouca informação e podem fomentar comportamentos indesejados no testando como erros por descuido ou até deliberadamente fazer a escolha incorreta. Tais comportamentos são influenciados, sobretudo, pelo fato de a tarefa ser pouco desafiadora, ou seja, desmotivadora. O inverso, itens muito difíceis ou muito ostensivos, também são pouco informativos, já que os testandos podem recorrer a tentar adivinhar a resposta ou simplesmente, dar qualquer resposta por achá-lo sem sentido (Linacre, 2000). A Testagem Adaptativa Computadorizada objetiva resolver ou minimizar estas questões.

A Testagem Adaptativa Computadorizada (*Computerized Adaptive Testing* – CAT) consiste na aplicação de recursos computacionais para que cada testando responda a um teste personalizado de acordo com a sua proficiência no construto do teste, minimizando os efeitos apontados no parágrafo anterior – itens muito fáceis ou muito difíceis para uma dada proficiência. Na CAT elaborase um algoritmo computacional segundo regras de escolha de itens e

encerramento do teste (regra de parada). Assim, conforme as respostas são dadas aos itens apresentados, novos itens são selecionados em um banco de itens (BI). Para iniciar o teste, um ou mais itens são apresentados ao testando, calcula-se uma estimativa provisória da proficiência, em seguida, na etapa adaptativa, um item mais informativo é veiculado e a proficiência é reestimada. Este ciclo irá se repetir até que a regra de encerramento programada seja alcançada (Magis & Barrada, 2014; Veldkamp & Mariagiulia, 2013; Luecht & Sireci, 2011).

Na execução de um CAT, de modo geral, ao responder corretamente ao item apresentado, o próximo item oferecido será mais difícil e subsequentemente outro mais difícil será apresentado ou caso a resposta seja incorreta, um item mais fácil será apresentado, e assim até atingir-se a regra de parada. Algumas das vantagens da CAT são que os indivíduos respondem a um número menor de itens ao passo que têm uma medida mais precisa da sua proficiência e evita-se a fadiga e exposição a itens pouco informativos. Também, têm-se um banco de itens com vários níveis de dificuldade e não apenas um pequeno conjunto de itens específicos, assim é mais difícil que o testando apenas memorize as respostas corretas em testes de desempenho. Os itens podem, ainda, agregar mais tecnologia e inovação em seus formatos, possuir maior flexibilidade no tempo de testagem, ou seja, cada indivíduo pode trabalhar em seu próprio ritmo e velocidade, torna-se possível o feedback dos resultados imediatamente após a testagem e itens antigos podem ser excluídos enquanto novos itens podem ser agregados (Wainer, 2000).

Os componentes de testes adaptativos computadorizados são: a) conjunto de itens calibrados (banco de itens); b) método para iniciar o teste; c) método de seleção dos itens; d) método de estimativa da proficiência; e e) regra de parada do teste. Um dos aspectos centrais da CAT é a construção do banco de itens, pois se os itens não tiverem qualidade psicométricas e/ou não atenderem aos objetivos da medição de forma geral, todos os demais produtos serão de baixa qualidade (Nunes, Spenassato, Bornia, Oliveira & Primi, 2015; Moreira Júnior; Tezza; Andrade & Bornia, 2013; Pasquali, 2007).

Os bancos de itens são desenvolvidos tanto visando ao objetivo de avaliar toda a extensão do traço latente quanto para tomarem-se decisões dicotômicas. No primeiro caso, o desafio é elaborar itens suficientemente discriminativos e em quantidade para toda a amplitude de níveis do traço latente. Em se tratando das decisões dicotômicas, deve-se buscar itens que possuam dificuldade em torno de um nível desejado e que sirvam de ponto de corte para a decisão, por exemplo,

proficiência ou não proficiência em determinado construto (Pasquali, 2007; Embretson & Reise, 2000).

A criação e a manutenção de um banco de itens demandam alguns cuidados para garantir a máxima eficiência e informação da CAT. O tamanho do banco, segundo Embretson e Reise (2000), precisa ser de cerca de 100 itens por dimensão. Conseqüentemente, estudos de análise fatorial necessitam ser implementados para o conhecimento das dimensões envolvidas em cada construto. Acresce-se que o banco deve ser representativo de apenas um único construto. Na formulação do banco, diferentes modelos da TRI e de escalas podem ser aplicados, conquanto seja respeitada a qualidade psicométrica dos itens. Ainda, é indicada a remoção de itens que se mostrem inadequados ao longo do tempo, com a evolução ou modificação do construto, e, assim, novos itens precisam ser acrescidos utilizando-se os procedimentos de equiparação dos mesmos conforme a TRI. Por fim, ressalta-se a necessidade de se efetuar revisões subsequentes da consistência da escala (Pasquali, 2007; Embretson & Reise, 2000). A CAT é um tipo de testagem dinâmica que está sempre em construção e reconstrução em busca de eficiência e qualidade, o que a aproxima dos pressupostos da Testagem Universal.

Na Testagem Universal almejam-se sistemas realmente inclusivos desde a sua forma de acesso, seus formatos, seus itens, a utilização de tecnologias assistivas variadas, até as estimativas das habilidades dos indivíduos - nestas últimas a CAT contribui enormemente. A união da CAT com a Testagem Universal permite somar-se a todas as vantagens dos testes informatizados um número menor de respostas necessárias para estimativas mais eficientes das habilidades dos indivíduos. O fato de utilizar um número menor de itens evita a fadiga e, portanto, facilita a realização pelos indivíduos com deficiências, sobretudo quando da presença de várias demandas ou comprometimentos físicos e/ou cognitivos (Almond; Winter; Cameto; Sato; Clarke-Midura; Torres; Haertel; Dolan & Lazarus, 2010; Ketterlin-Geller, 2005).

A CAT é utilizada nos instrumentos de Testagem Universal para o incremento da precisão e da eficiência. Medir os resultados dos testandos que se encontram na extremidade inferior ou superior do espectro de proficiência, tanto de indivíduos com deficiência quanto sem deficiência, é bastante relevante em certos construtos no campo da saúde mental (ex. depressão, ansiedade). Tais estimativas potencializam o objetivo da Testagem Universal de maior inclusão (Almond; Winter;

Cameto; Sato; Clarke-Midura; Torres; Haertel; Dolan & Lazarus, 2010; Ketterlin-Geller, 2005).

Do mesmo modo que aos demais testes informatizados, todos os princípios da Testagem Universal são aplicáveis aos testes de CAT. Os princípios da Testagem Universal, são: População de avaliação ampla e inclusiva; 2. Definição precisa do construto; 3. Itens acessíveis e não tendenciosos; 4. Testes flexíveis a acomodações; 5. Instruções e procedimentos simples, claros e intuitivos; 6. Leitura agradável e de máxima inteligibilidade; e, 7. Máxima legibilidade. Apesar de não serem imprescindíveis à elaboração de instrumentos de Testagem Universal, os recursos da CAT se tornam ideais para criar instrumentos de Testagem Universal (Almond; Winter; Cameto; Sato; Clarke-Midura; Torres; Haertel; Dolan & Lazarus, 2010; Ketterlin-Geller, 2005). Assim, a CAT torna-se um dos recursos mais potentes da TRI aplicada a Testagem Universal. A aplicação da Testagem Universal em CAT deverá partir desde o projeto inicial de construção do instrumento, tendo no banco de itens aplicações que serão determinantes para que o teste seja realmente considerado de Testagem Universal.

Na construção de um BI os itens que o compõe passam por várias etapas: 1. Desenvolvimento; 2. Análise de juízes; 3. Busca por evidências de validade; e, 3. Estimativa dos parâmetros pela TRI (Nunes, Spenassato, Bornia, Oliveira & Primi, 2015). Comumente, os relatos de pesquisas no âmbito da CAT preocupam-se com a calibração do banco de itens ou com a busca de evidências de validade, neste artigo ressaltar-se-á os passos que antecedem esta calibração com vistas à análise qualitativa dos itens segundo os pressupostos da Testagem Universal, quais sejam: 1. Desenvolvimento dos itens; e, 2. Análise de Juízes e a preparação do banco para a calibração. Com a aplicação da Testagem Universal se busca alcançar itens mais acessíveis para os mais diversos públicos, sejam pessoas com deficiência ou sem deficiência, mantendo a invariância dos mesmos.

Foram desenvolvidos itens do construto personalidade, cinco grandes fatores, quais sejam: Abertura, Realização, Extroversão, Amabilidade e Estabilidade Emocional (Neuroticismo), considerando cada qual com três facetas (A: Imaginação, Intelecto e Sensibilidade à estética; R: Disciplina, Organização e Segurança; E: Assertividade, Atividade e Sociabilidade; A: Compaixão, Confiança e Polidez; EE: Autoconfiança, Resiliência ao estresse e Estabilidade emocional) (Primi, Santos, John & De Fruyt, 2016). Na elaboração dos itens primou-se pela fidelidade ao construto e aplicou-se os princípios da Testagem Universal. Acresce-se que no tocante à nomenclatura utilizada em

relação às pessoas com deficiência aplicou-se as referendadas pelo modelo social da deficiência.

O modelo social da deficiência teve seu marco durante a luta pelo fim das barreiras arquitetônicas para as pessoas com deficiência, ampliando seus conceitos ao logo dos últimos anos. Atualmente aborda questões como o empoderamento e interdependência (Barnes, 2009; Lang, 2009; Santos, 2010). Devido à sua importância, destaca-se a convenção pelos direitos da pessoa com deficiência (2007), a qual aborda, dentre outros aspectos, questões que tangem à autonomia individual e à liberdade de fazer suas próprias escolhas pela pessoa com deficiência, à inclusão e a não discriminação, à deficiência como parte da diversidade humana e à acessibilidade. Tais aspectos foram amplamente considerados durante este estudo desde a elaboração dos itens e construção do BI, estando tal modelo em consonância com a Testagem Universal.

No tocante à análise de juízes, seguiu-se os pressupostos estabelecidos na literatura científica da área, segundo a qual, na análise do conteúdo dos itens, os juízes são peritos na área do construto e sua tarefa consiste em ajuizar a correspondência dos itens com seus fatores e facetas (Pasquali, 2010). Contudo, devido ao grande número de itens, fatores e facetas, bem como a aplicação da Testagem Universal, a tarefa se tornou deveras complexa. Deste modo, foi necessário um delineamento de análise de juízes que permitisse a qualidade do processo, além de acrescentar-se a etapa de avaliação da Testagem Universal pela equipe de pesquisa, como será descrito em tópicos subsequentes deste artigo.

Com vistas à organização do banco de itens para o processo de calibração, a estratégia utilizada foi a construção dos Blocos Incompletos Balanceados (BIB), os quais são uma solução estatística que permite a coleta de um grande número de informações de uma quantidade extensa de itens, sendo que um mesmo testando responde a apenas uma fração destes. Busca-se evitar, deste modo, a fadiga e a desmotivação seguidas de desistências e omissões que podem ocorrer diante de um grande número de itens. Os itens são separados em blocos e uma quantidade k de blocos constituirá um caderno ou prova. Assim, cada caderno possui uma quantidade incompleta dos blocos de itens (Bekman, 2001). Neste estudo, este recurso estatístico foi utilizado na preparação para a coleta de dados de um banco de itens de personalidade e Testagem Universal construído para testagem adaptativa.

No BIB cada caderno possui o mesmo número de blocos e cada bloco é utilizado um mesmo número de vezes em um mesmo conjunto

total de cadernos e, ainda, cada par de blocos é utilizado o mesmo número de vezes no conjunto total de cadernos (Beckman, 2001). Além dos aspectos já mencionados, a organização da coleta em BIB permite que as análises obtenham uma matriz de correlações completa entre os itens (Nunes, Spenassato, Bornia, Oliveira & Primi, 2015). Ronald e Fisher (1963) desenvolveram as tabelas das combinações possíveis para o BIB, entretanto, ressalta-se que em alguns casos não é possível realizar-se a combinação desejada (Beckman, 2001). A etapa, portanto, exige planejamento prévio do banco de itens para que o número de itens elaborados adeque-se ao BIB esperado.

O presente texto científico abarca a aplicação da Testagem Universal na construção de um banco de itens para testagem adaptativa que se utiliza de recursos estatísticos como os Blocos Incompletos Balanceados visando a alcançar os objetivos esperados.

Método

A presente pesquisa enquadra-se no âmbito dos estudos de desenvolvimento de instrumentos em Avaliação Psicológica. O objetivo geral consistiu em desenvolver um banco de itens de personalidade de Testagem Universal e os específicos foram construir itens segundo os 7 princípios da Testagem Universal, propor um delineamento para análise de juízes e construir um instrumento para calibração dos itens. Foram desenvolvidos e selecionados os itens que formaram o banco de itens. Os itens desenvolvidos neste estudo passaram por análise de juízes e por avaliação dos critérios oriundos da Testagem Universal. Os itens foram desenvolvidos para indivíduos maiores de idade sem e com deficiência (visual e motora). Os produtos deste estudo foram: o banco de itens de Testagem Universal e um instrumento para calibração dos itens, a Bateria de Personalidade e Testagem Universal (BPTU). Tal bateria foi alocada na Plataforma *Concerto (Open-source Online Rbased Adaptive Testing Platform)*, a qual é uma plataforma para criação de instrumentos de código aberto que permite que os usuários criem avaliações *on-line*.

Resultados e Discussão

Banco de Itens

A construção de um Banco de Itens para testagem adaptativa informatizada segundo os princípios da Testagem Universal assume alto

grau de complexidade. Nesta pesquisa, tal processo conteve 5 passos ou etapas:

1º Passo: Desenvolvimento de itens.

Foram desenvolvidos itens para cada uma das 15 facetas, sendo 3 facetas de cada um dos 5 Grandes Fatores (Figura 1). A meta era o desenvolvimento de 40 itens por faceta, ou seja, 120 itens por fator, totalizando 600 itens, respeitando-se, portanto, o que aponta a literatura da área (Pasquali, 2007; Embretson & Reise, 2000) que aponta um número de 100 itens por dimensão com a previsão de perdas de itens de até 20% do total. No estudo, optou-se pela elaboração de 20 itens excedentes para minimizar possíveis perdas futuras.

ABERTURA	REALIZAÇÃO	EXTROVERSÃO	AMABILIDADE	ESTABILIDADE EMOCIONAL
Imaginação Intelecto Sensibilidade	Disciplina Organização Segurança	Assertividade Energia Sociabilidade	Compaixão Confiança Polidez	Autoconfiança Resiliência ao estresse Estabilidade emocional

Figura 1. Cinco Grandes Fatores e Facetas.

Os itens além de serem desenvolvidos segundo o construto proposto – personalidade no modelo dos Cinco Grandes Fatores – foram elaborados aplicando-se os 7 princípios da Testagem Universal (Figura 2):

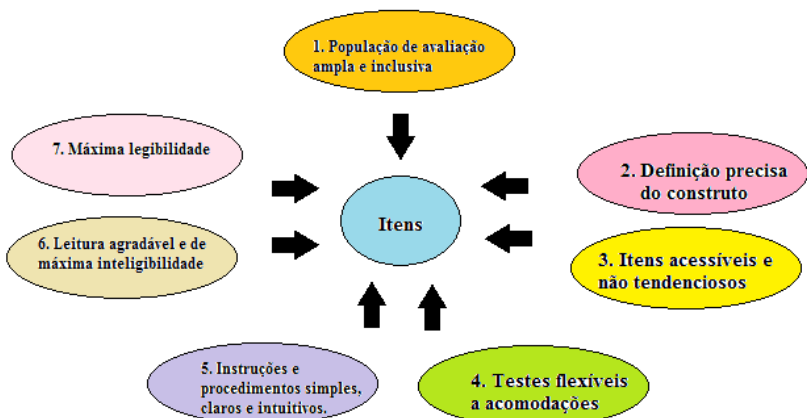


Figura 2. Os 7 princípios da Testagem Universal.

Na aplicação dos princípios da Testagem Universal ao desenvolvimento dos itens, procurou-se adequar o conteúdo dos itens para que fossem de fácil entendimento pelos grupos de pessoas com e sem deficiência e observou-se as palavras utilizadas para evitar linguagem técnico ou mais rebuscado a fim de que cada item contivesse apenas palavras que fossem de fácil entendimento pelo público alvo do estudo, primando-se, assim, pela simplicidade do item. Observou-se ainda o número de palavras para que o texto do item fosse curto objetivando evitar erros acidentais de interpretação e fadiga. O conteúdo de cada item foi escolhido com vistas a assegurar a equidade para pessoas com e sem deficiência e para evitar ambiguidades de qualquer natureza (Gaster & Clark, 1995; Shinn & Ofiesh, 2012). A Tabela 1 resume a relação entre cada princípio da Testagem Universal e sua aplicação no desenvolvimento dos itens do teste:

Tabela 1.

Aplicação dos princípios da Testagem Universal ao desenvolvimento dos itens do teste.

PRINCÍPIOS DA TESTAGEM UNIVERSAL	APLICAÇÃO NOS ITENS DO TESTE
1. População de avaliação ampla e inclusiva.	Conteúdo do item adequado para pessoas com e sem deficiência.
2. Definição precisa do construto.	Simplicidade do conteúdo do item e evitar ambiguidade.
3. Itens acessíveis e não tendenciosos.	Equidade para pessoas com e sem deficiência.
4. Testes flexíveis a acomodações.	Texto curto (preocupação com o número de palavras). Evitou-se linguagem técnico ou rebuscado e ambiguidade.
5. Instruções e procedimentos simples, claros e intuitivos.	Simplicidade e clareza nos textos dos itens.
6. Leitura agradável e de máxima inteligibilidade.	Texto curto. Simplicidade do conteúdo dos itens. Evitar ambiguidade, linguagem técnico ou rebuscado.
7. Máxima legibilidade.	Texto curto.

O desenvolvimento dos itens deste trabalho também possuiu uma forte influência do modelo social da deficiência (Diniz, 2003; Barnes, 2009; Barton, 2009; Lang, 2009; Santos, 2010), pois foi permeado por toda uma preocupação no que se refere às nomenclaturas utilizadas para que as mesmas estivessem em consonância com esse modelo. Assim, em consonância com o modelo social da deficiência, procurou-se transpor barreiras atitudinais e físicas durante a construção de cada item que compõe o BI gerado neste estudo. Assim sendo, considerou-se a deficiência do ponto de vista da diversidade humana, portanto, como uma questão de direitos humanos (Gesser, Filgueiras Toneli & Nuernberg, 2012).

2º Passo: Revisão dos itens.

A grafia dos itens foi revisada bem como o equilíbrio entre o número de itens afirmativos e os invertidos. Ainda foram realizadas análises da Testagem Universal quanto ao conteúdo dos itens, sobretudo no tocante a equidade, clareza e a simplicidade.

Nesta etapa foi constatado um desequilíbrio entre o número de itens afirmativos e invertidos pela equipe de pesquisa. Com vistas a solucionar tal questão foram elaborados mais itens invertidos para cada faceta. Deste modo, chegou-se a um número de itens superior ao inicialmente esperado por faceta (total de 1.111 itens). Acresce-se que, mesmo sendo desenvolvidos primando-se pelos princípios da Testagem Universal, alguns poucos itens não atenderam completamente a estes princípios, tendo sido, então, excluídos ou reformulados.

A Figura 3 apresenta o fluxograma de resumo das atividades dos Passos 1 e 2.

3º Passo: Seleção dos itens para Análise de Juízes.

Este passo consistiu em selecionar no BI os 40 itens por faceta que fizeram parte da análise de juízes objetivando manter a meta inicial de 600 itens.

Os critérios adotados para permanência dos itens foram: Realizar uma previsão qualitativa de ostensividade dos itens mantendo no banco certo equilíbrio numérico dos itens mais ostensivos, medianos e menos ostensivos na tentativa de alcançar uma maior amplitude de valores de *theta* nas futuras análises; e, ainda, manter o maior equilíbrio possível no número de itens afirmativos e invertidos.

A Figura 4 resume as atividades do 3º Passo.

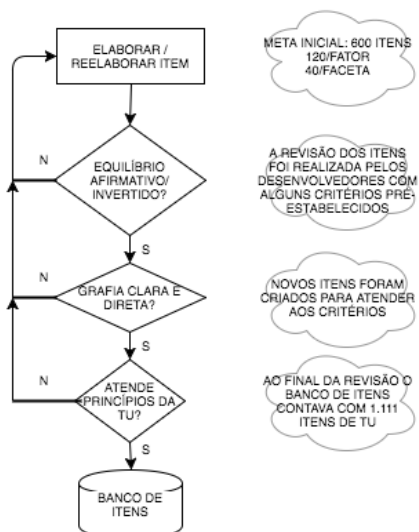


Figura 3. Fluxograma dos 1º e 2º Passos.

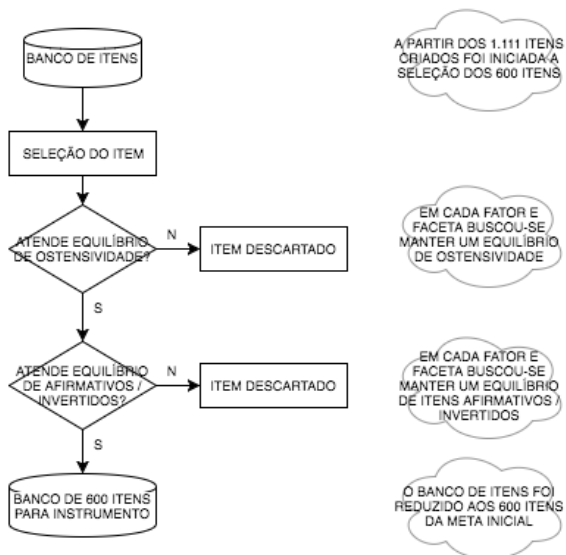


Figura 4. Fluxograma do 3º Passo.

4º Passo: Preparação dos blocos e cadernos para a Análise de Juízes. Aleatoriedade dos itens que farão parte dos blocos e cadernos da análise de juízes.

Foi realizado o desenho da análise de juízes para 4 juízes (Figura 5), com vistas a evitar fadiga e conseqüentemente queda da qualidade da análise dos juízes devido a extensão do banco (600 itens). O delineamento foi elaborado para que cada juiz analisasse um caderno de 300 itens (Figura UU). Os itens foram divididos em 4 blocos com 150 itens. Do agrupamento de dois a dois destes blocos foram compostos os 4 cadernos (Caderno 1 – cinza; Caderno 2 – marrom; Caderno 3 – azul; e, Caderno 4 – areia).

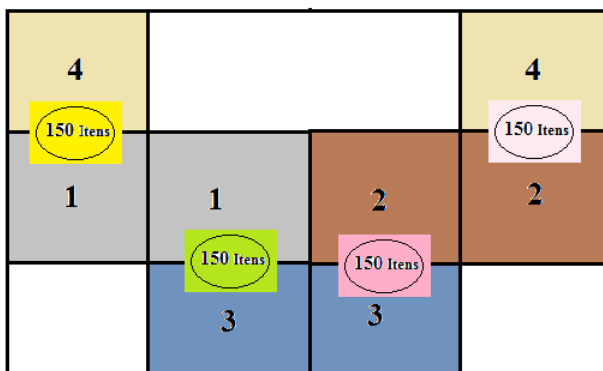


Figura 5. Desenho da Análise de Juízes.

Acresce-se que na formação dos blocos e cadernos, os 40 itens de cada faceta foram mudados de ordem aleatoriamente. Os 10 primeiros itens de cada faceta formaram o bloco 1, os 10 seguintes o bloco 2, e assim sucessivamente para o bloco 3 e o bloco 4 – totalizando os 150 itens de cada bloco – os quais foram distribuídos para os 4 juízes conforme apresentado na Tabela 2. Este procedimento de ordem aleatória foi realizado utilizando a função *Random* do programa “Excel”. Em seguida, foram agrupados os itens de cada bloco, criou-se códigos referentes a fator (1f,2f,3f,4f e 5f) e facetas (1fc,2fc,3fc) e foi gerada uma nova sequência aleatória dos 150 itens para evitar o agrupamento de facetas. Os itens seguiram para os juízes com seus respectivos códigos referentes a fator e faceta. Os cadernos foram montados de forma que cada bloco fosse analisado por dois juízes diferentes (Tabela 2).

Tabela 2.

Distribuição dos blocos por juiz.

JUIZES	BLOCOS	
JUIZ 1	BLOCO 1	BLOCO 2
JUIZ 2	BLOCO 2	BLOCO 3
JUIZ 3	BLOCO 3	BLOCO 4
JUIZ 4	BLOCO 1	BLOCO 4

A Figura 6 apresenta um resumo das etapas cumpridas no 4º Passo.

5º *Passo*: Resultados da Análise de Juízes.

Os 4 juízes foram escolhidos pelo expertise no tema (Pasquali, 2010), sendo todos pesquisadores da área da personalidade e Cinco Grandes Fatores. Cada qual recebeu um caderno com dois blocos de itens (Figura II) e uma breve descrição dos fatores e facetas. Na análise, os juízes deveriam indicar o fator e apenas após indicar o fator, indicar a faceta correspondente, se consideravam o item invertido (Sim/Não), indicar sugestões de alterações nos itens e ainda alguma observação adicional. A escolha dos fatores, facetas e inversão de itens foram feitas com a ferramenta de validação de dados do Excel (Figura 7).

Após os juízes analisarem os itens, julgaram-se os resultados segundo os critérios da análise de juízes para permanência, exclusão ou modificação do item e também quanto às mudanças em fator e faceta. A permanência ocorreu, principalmente, nos itens em que houve concordância entre fator e faceta (entre juízes e pesquisadores, entre 1 dos juízes e pesquisador e entre os juízes) sem a sugestão de alteração na escrita do item. A concordância apenas entre fatores também foi critério de permanência, ao passo que se juízes e pesquisadores discordassem quanto ao fator, o item foi excluído – o que raramente ocorreu. Em relação às facetas, foi-se menos exigente, sendo critério apenas de modificação, sobretudo nos casos em que os dois juízes concordaram e apresentaram resultados diferentes dos pesquisadores, e não de exclusão do item.

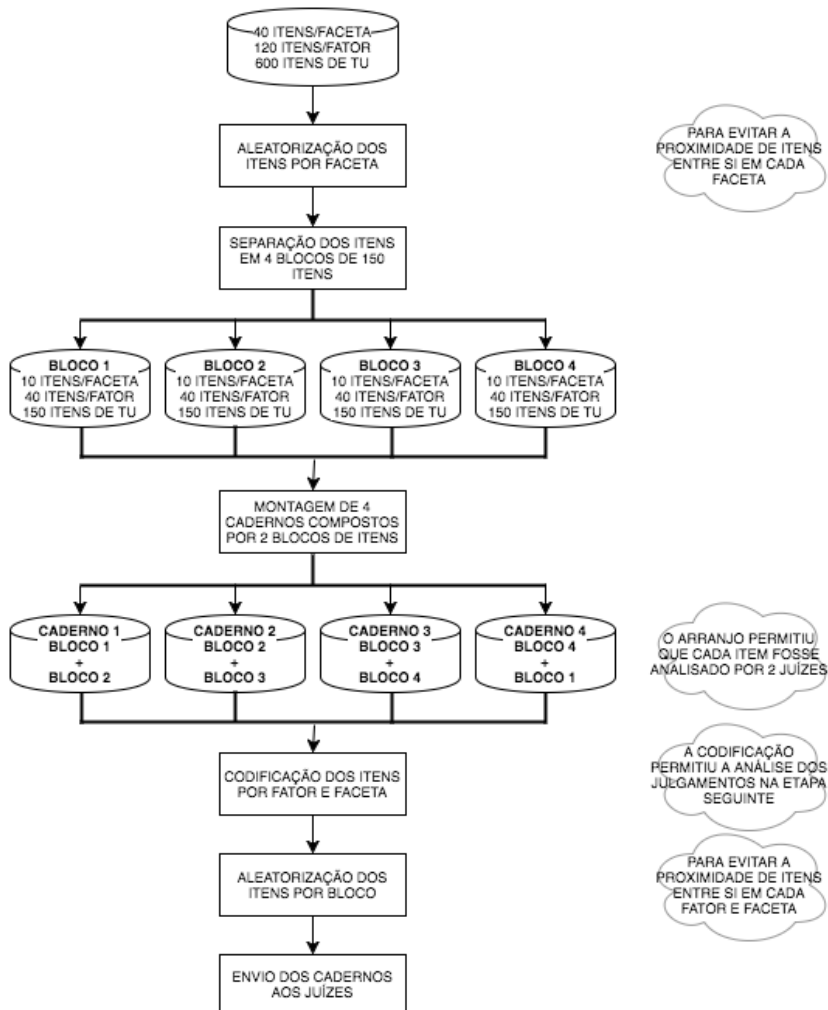


Figura 6. Fluxograma do 4º Passo.

	A	B	C	D	E	F
1						
2	Item	Fator	Faceta	Invertido	Sugestões de alteração do Item	Observações
3	Costumo me destacar por ser bom com palavras / Sou bom com as palavras.					
4	Tenho esperança no meu futuro.					
5	Meus amigos dizem que eu sou entusiasmado.					
6	Procuro ser discreto no meu modo de falar / Sou discreto no meu modo de falar					
7	Fico chateado quando ocorrem mudanças no meu dia-a-dia / Mudanças no ambiente					
8	Divirto-me com jogos de adivinhação.					
4	BLOCO 1	BLOCO 2				

Figura 7. Caderno para análise de juízes.

Os juízes sugeriram a reescrita ou a retirada de itens, as quais foram em grande parte atendidas. Ainda, os itens que estavam agrupados (por exemplo: “*Costumo me destacar por ser bom com palavras./ Sou bom com as palavras*”) foram desmembrados e contabilizados, ou um dos itens permaneceu no banco enquanto os demais foram excluídos. Itens que, apesar das análises anteriores, ainda possuíam palavras rebuscadas ou ambiguidades foram retirados (por exemplo: no item “*Costumo tomar minhas decisões com prudência*”, a palavra “prudência” é muito rebuscada, o que ocasionou a exclusão do item). Ao final desta etapa perduraram no banco mais de 600 itens, pois, apesar das exclusões, ocorreram desmembramentos, o que provocou a superação da meta inicial.

A Figura 8 sintetiza as decisões tomadas no 5º Passo.

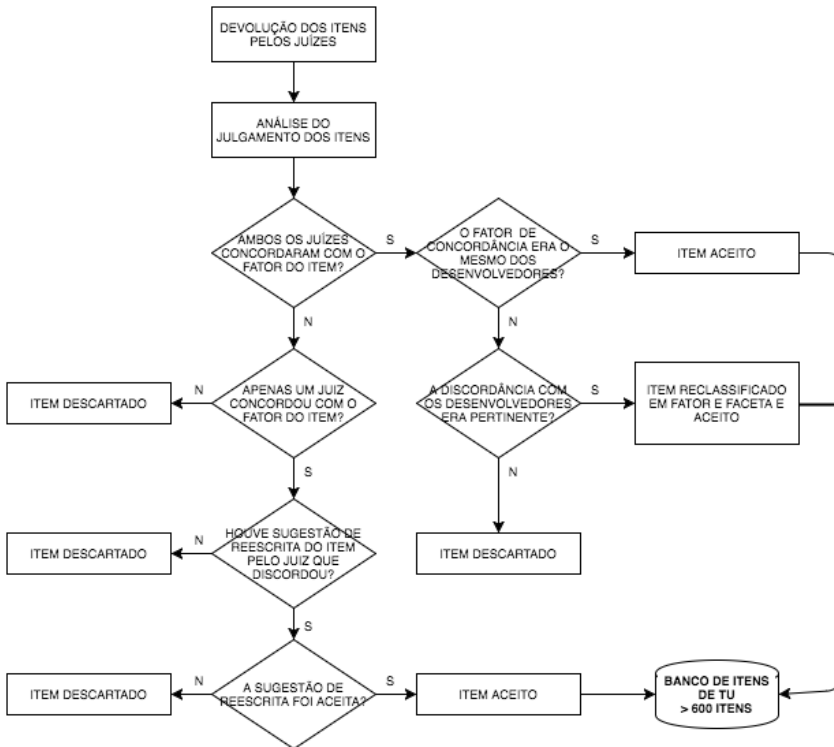


Figura 8. Fluxograma do 5º Passo.

6º Passo: Organização do BIB.

No modelo de BIB adotado (15, 35, 3) (Figura 9) os itens foram separados em 35 cadernos (c), 15 blocos (b), com um número 3 (k) blocos por caderno e o número de 7 (r) repetições de cada bloco. Ainda, o fator de utilização (fu) foi de 0,2, ou seja, 20% dos testandos foi submetido a cada bloco (Ronald & Fisher, 1963; Andrade, Tavares & Valle, 2000; Beckman, 2001; Moreira Junior, 2011). Do total de itens disponíveis no banco foram selecionados 525, quantidade que atende aos pressupostos do BIB e a extensão do banco de itens. Assim sendo, os 525 itens foram distribuídos no banco em 35 itens por bloco e 105 itens em cada caderno ($525/15=35$; $3*35=105$).

		CADERNOS (C)																																										
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35								
B L O C O S (B)	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0						
	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0					
	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0			
	4	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0			
	5	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0			
	6	1	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
	7	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
	8	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
	9	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	
	10	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	
	11	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	
	12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1
	13	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	
	14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	
	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1		

Figura 9. Modelo BIB utilizado no banco de itens (STATA 14).

Na seleção dos itens que permaneceram no banco, adotou-se a mesma quantidade por fator e faceta, ou seja, 105 por fator ($105*5=525$) e 35 por faceta ($35*3=105$). Na montagem do BIB cada caderno tinha 35 itens, subdivididos em 7 itens de cada fator, sendo 2 ou 3 itens de cada faceta para cada fator. Essa diferença numérica numa das facetas em um dado caderno foi compensada no caderno subsequente. A seleção dos itens que compuseram cada caderno foi realizada novamente por sorteio aleatório com a utilização da função *Random* do Excel (Excel 2010). Depois de escolhidos os itens de cada caderno foi realizada uma nova revisão item a item, a fim de equilibrar os cadernos com itens mais e menos ostensivos. Esta etapa de revisão resultou na mudança de itens entre alguns cadernos, ou simplesmente na mudança da posição do item dentro do caderno.

A Figura 10 apresenta as etapas realizadas no 6º Passo.



Figura 10. Fluxograma do 6º Passo.

7º Passo: Desenvolvimento do instrumento informatizado para coleta de calibração dos itens. Implementação no *Concerto*.

Para o processo de coleta de dados foi desenvolvido um instrumento informatizado na plataforma *Concerto*. O *Concerto* (*Open-source Online Rbased Adaptive Testing Platform*) é uma plataforma para criação de instrumentos de código aberto que permite que os usuários criem avaliações *on-line*. O formato de apresentação do teste – Bateria de Personalidade e Testagem Universal (BPTU) - no *Concerto* foi desenvolvido para atender aos pressupostos da Testagem Universal.

A elaboração dos *templates* implementados no instrumento foi realizada pela equipe de pesquisa (Figura 11). O *template* é um modelo de documento sem conteúdo, apenas com a apresentação visual, um modelo. Os conteúdos dos *templates* foram acrescidos na medida em que os mesmos foram elaborados. Durante este processo foi verificado se os *templates* eram acessíveis aos leitores de tela e os formatos que não se adequaram foram alterados ou excluídos. A formatação dos *templates* repercutiu na utilização de tecnologias assistivas, como os leitores de tela, e tem relação com questões como espaçamento, tipo e tamanho de letras, marcações e numerações, que precisam ser testadas com as tecnologias assistivas após a implementação.


 <p>UNIVERSIDADE FEDERAL DE SANTA CATARINA</p> <p>LPAP - Laboratório de Pesquisas em Avaliação Psicológica Bateria de Personalidade e Testagem Universal</p> <p>Apresentação</p> <p>Esta pesquisa tem como objetivo o desenvolvimento de um teste informatizado para - você deverá indicar o quanto você se identifica com frases que descrevem situações d</p> <p>Este projeto segue as exigências e cuidados éticos para a realização de pesquisas UFSC.</p>	<p>Na sequência quatro frases serão apresentadas como treino. Após cada frase são apresentadas as:</p> <p>Escolha 1 se a frase o descreve muito mal. Escolha 2 se a frase o descreve mal. Escolha 3 se a frase o descreve mais ou menos. Escolha 4 se a frase o descreve bem. Escolha 5 se a frase o descreve muito bem.</p> <p>Apenas uma opção entre as cinco deve ser escolhida.</p> <p>Sugestão: O teclado numérico acelera a escolha das opções.</p> <p>A seguir, como um treino, responda aos seguintes exemplos:</p> <p>Responda aos quatro itens:</p> <p>Item 1. Seu afável <input type="text"/></p> <p>Item 2. Seu comunicativo <input type="text"/></p> <p>Item 3. Gosto de pessoas <input type="text"/></p>
<p>Bateria de Personalidade e Testagem Universal</p> <p>Você responderá agora a 13 questões importantes sobre você.</p> <p>Observação: Algumas questões estão em caixas de testagem e outras deverão ser escritas nas caixas de texto.</p> <p>1. Sexo: <input type="text"/></p> <p>2. Idade (em anos): <input type="text"/></p> <p>3. Desejo sexo: <input type="text"/></p> <p>4. Caso você queira obter seus resultados informe seu e-mail: <input type="text"/></p> <p>5. Estado onde reside atualmente: <input type="text"/></p> <p>6. Escolaridade: <input type="text"/></p> <p>7. Ocupação atual: <input type="text"/></p>	<p>Bateria de Personalidade e Testagem Universal</p> <p>Página 1</p> <p>Responda aos 35 itens a seguir:</p> <p>1. Esforço-me para vencer obstáculos quando estou realizando um trabalho. <input type="text"/></p> <p>2. Detesto seguir instruções. <input type="text"/></p> <p>3. Gosto de desafios que me façam pensar. <input type="text"/></p>

Figura 11. Exemplos de *templates* implementados.

No processo de elaboração dos *templates* com vistas à garantia dos princípios da Testagem Universal foram observados os textos de apresentação dos itens e instruções garantindo que os mesmos fossem descritivos, simples e claros (Figura 11). O formato dos itens em caixas de seleção ou combinação foram implementados com o intuito de serem acessíveis aos leitores de tela (Figura 11). As páginas do teste foram organizadas de modo que a informação visual relevante sempre possuísse um correspondente textual. Cuidados com o uso da numeração e informações adicionais de quantas páginas ainda faltam para realizar o instrumento em cada página do teste foram implementados. O tamanho das letras seguiu a recomendação de Fuchs, Fuchs, Eaton, Hamlett, Binkley e Crouch (2000) de estarem em tamanho maior que 12, sendo escolhido o tamanho 16.

Para considerar-se a BPTU como um instrumento de Testagem Universal fez-se imprescindível sua adequação à tecnologias assistivas (TA). As tecnologias assistivas agregam um conjunto de materiais, desde instrumentos já amplamente conhecidos e utilizados como as bengalas até potentes *softwares* leitores de telas (Oliveira, 2013; Oliveira & Nunes, 2015). Na BPTU seguiu-se as orientações de Shinn e Ofiesh (2012) quanto à utilização desta TA (*softwares* leitores de telas) as quais se coadunam com os já referidos princípios da Testagem Universal: texto claro, palavras simples, evitação de termos técnicos, frases compostas separadas em frases simples, prioridade para ideias importantes, introdução de uma ideia de cada vez e apresentação de instruções na sequência em que devem ser executadas. Complementarmente, para primar pela acessibilidade, foram disponibilizados sugestões de comandos utilizados por programas leitores de telas (Figura 12).

Bateria de Personalidade e Testagem Universal.

Sugestão de comandos em alguns programas leitores de tela.

Caso você esteja utilizando um programa leitor de tela segue abaixo alguns comandos que provavelmente você utilizará durar quaisquer destes programas passe para próxima página):

Nos leitores de tela JAWS, NVDA e Virtual Vision alguns comandos que podem ser úteis para realização do teste são:

1. Control+Home para ir para o início da tela;
2. As setas para navegar entre as linhas – por exemplo: No texto que está no topo de cada tela;
3. Tab para caminhar nas questões do teste e links;
4. Tab+enter+setas para escolher nas questões a opção que desejar; e,
5. Shift+Tab para voltar para a questão anterior.

Observação: No Leitor Virtual Vision acrescenta-se, apenas que se utiliza o Esc para sair de uma questão e posteriormente o

*As instruções apresentadas estão sujeitas a variações dependendo da versão do leitor utilizado.

[Próxima página](#)

Figura 12. Alguns comandos utilizados nos programas leitores de telas.

No desenvolvimento deste instrumento de Testagem Universal utilizaram-se os pressupostos já estabelecidos em Oliveira (2013) (Tabela 3).

Tabela 3.

Aplicação dos princípios da Testagem Universal na BPTU.

Princípios da Testagem Universal	Elementos	Aplicações
<i>1. População de avaliação ampla e inclusiva</i>	Pessoas com e sem deficiência	Texto descritivo; Flexível à utilização de leitores de tela.
<i>2. Definição precisa do construto</i>	Construto	Conteúdo dos itens.
<i>3. Itens acessíveis e não tendenciosos</i>	Itens	O Formato dos itens em caixas de seleção ou combinação. <i>Template</i> dos itens acessível aos leitores de tela.
<i>4. Testes flexíveis a acomodações</i>	Tecnologia Assistiva (agente do usuário)	O teste pode ser respondido com ou sem os leitores de tela.
<i>5. Instruções e procedimentos simples</i>	Instruções e demais textos	Texto descritivo, claro e simples.
<i>6. Leitura agradável e de máxima inteligibilidade</i>	Texto descritivo	Informou-se no início de cada tela que contém os itens a sua extensão (número de itens); Botões tela anterior/ próxima tela possuem correspondente textual.
<i>7. Máxima legibilidade</i>	Formato	Letra grande (Arial 16). Cor das letras (predominantemente pretas).

Bateria de Personalidade e Testagem Universal

Instruções e Treino — **Título resumindo o conteúdo da página.**

Você vai responder a um teste que contém frases que descrevem sentimentos, opiniões e atitudes. Por favor, leia atentamente com elas.

Na sequência quatro frases serão apresentadas como treino. Após cada frase são apresentadas as opções numeradas de 1 a 5.

Escolha 1 se a frase o descreve muito mal.
 Escolha 2 se a frase o descreve mal.
 Escolha 3 se a frase o descreve mais ou menos.
 Escolha 4 se a frase o descreve bem.
 Escolha 5 se a frase o descreve muito bem.

Apenas uma opção entre as cinco deve ser escolhida.

Sugestão: O teclado numérico acelera a escolha das opções.

A seguir, como um treino, responda aos seguintes exemplos:

Responda aos quatro itens: **Formato dos itens.**

Item 1. Sou afável

Item 2. Sou comunicativo

Item 3. Gosto de pessoas

Figura 13. Exemplos de aplicações da Testagem Universal nos *Templates*.

Ao final da implementação do teste no *Concerto* realizou-se um checklist dos principais elementos de Testagem Universal com vistas a auxiliar no processo de revisão do teste - desenvolvido pelos autores desta pesquisa.

CHECKLIST DE TESTAGEM UNIVERSAL		
Público ao qual o teste se destina: _____		
Avaliador: _____		Data: _____
Conteúdo dos Itens	Atende	Não atende
8. O texto do item é de fácil entendimento. 9. As palavras utilizadas são de fácil entendimento por grupos variados. 10. O texto é curto. 11. O conteúdo é equitativo para pessoas com e sem deficiência. 12. O texto do item está livre de ambiguidade. 13. O texto é lido com qualidade pelo software leitor de tela. 14. O texto se adequa aos recursos de Tecnologia Assistiva que serão utilizados.		
Instruções		
9. O texto das instruções é de fácil entendimento. 10. As palavras utilizadas são de fácil entendimento por grupos variados.		

<ol style="list-style-type: none"> 11. O texto é curto. 12. As instruções possuem conteúdo equitativo para pessoas com e sem deficiência. 13. O texto das instruções está livre de ambiguidade. 14. O texto das instruções é lido com qualidade pelo software leitor de tela. 15. O texto se adequa aos recursos de tecnologia assistiva que serão utilizados. 16. As instruções são compreensíveis para os mais variados tipos de leitores (Ex. cegos, videntes, surdos, etc.). 		
<p>Design do Instrumento</p> <ol style="list-style-type: none"> 4. O <i>design</i> é compatível com os recursos de Tecnologia Assistiva planejados. 5. O <i>design</i> do instrumento favorece a tolerância ao erro. 6. O <i>design</i> proporciona a fácil percepção do conteúdo do instrumento. 		
<p>Formato dos itens</p> <ol style="list-style-type: none"> 5. O formato é acessível. 6. O formato dos itens é flexível aos recursos de Tecnologia Assistiva. 7. O formato é equitativo para pessoas com e sem deficiência. 8. O formato dos itens proporciona tolerância ao erro. 		

Após a verificação dos pontos do *checklist* pela equipe de pesquisa, a qual possui membros com experiência em Testagem Universal, foram realizadas as mudanças apontadas. Assim, finalizaram-se as etapas da elaboração do instrumento de Testagem Universal.

Conclusão

A Testagem Universal busca agregar as melhores práticas no que se refere aos formatos, as estruturas, aos tipos de mensuração e a avaliação dos instrumentos. Coaduna-se, portanto, plenamente com a testagem adaptativa informatizada, primando pela qualidade do banco de itens e de toda a tecnologia utilizada durante a construção do mesmo. Realizou-se, assim, a construção de um banco de itens para CAT segundo os princípios da Testagem Universal e medindo o construto personalidade.

Na presente pesquisa foram tomadas decisões qualitativas embasadas na literatura científica no que tange à construção de instrumentos de forma ampla e de instrumentos de testagem adaptativa e Testagem Universal mais especificamente. Acresce-se que na construção de testes psicológicos de Testagem Universal, faz-se imprescindível o atendimento aos pressupostos teóricos e técnicos inerentes ao processo de construção de instrumentos sedimentados na literatura científica da área.

Os princípios da Testagem Universal foram aplicados na elaboração dos itens, em seus conteúdos, preocupando-se com a simplicidade, evitando ambiguidade e primando pela equidade. Igualmente, aplicou-se tais princípios ao instrumento para calibração do banco de itens (Bateria de Personalidade e Testagem Universal) no formato dos itens, nos *Templates* elaborados, cores, tamanho de fonte e no texto das instruções. Tais medidas visaram a ampliar a qualidade do teste para a utilização por pessoas com e sem deficiência abarcando a diversidade cognitiva humana – embora, no presente caso, ainda parcialmente restrito para pessoas sem deficiência e com deficiência visual e motora.

Algumas limitações encontradas na aplicação dos princípios da Testagem Universal foram que em se tratando do primeiro princípio (população ampla e inclusiva) os itens foram desenvolvidos apenas para indivíduos com deficiência visual (baixa visão e cegueira) e deficiência motora, não abarcando um espectro maior de deficiências e suas comorbidades. No tocante ao referencial da diversidade cognitiva humana, não foi abarcado no planejamento dos itens indivíduos com TDAH (Transtorno de Déficit de Atenção e Hiperatividade), déficit cognitivo, dificuldades de aprendizagem, depressão, entre outros.

Ainda, no âmbito das limitações, uma dificuldade encontrada foi a de não possuir na equipe de pesquisa pelo menos um especialista em informática e acessibilidade. Tal ausência foi parcialmente suprida pela equipe, porém culminou por limitar as possibilidades de recursos de tecnologias assistivas testados, restringindo-se apenas aos leitores de tela. Complementarmente, os recursos de acessibilidade que a equipe de pesquisa foi capaz de implementar na plataforma *Concerto* foram reduzidos. Sugere-se para novos estudos a presença de um profissional da informática que seja especialista em acessibilidade.

Como aspectos bastante positivos destaca-se o delineamento feito para a análise de juízes que permitiu, apesar do extenso número de itens, que os juízes realizassem a tarefa com qualidade, e a aplicação do BIB, como um recurso estatístico útil em bancos de itens para coleta de

um grande número de informações durante o processo de calibração dos itens. Conclui-se, portanto que a construção de um banco de itens de personalidade e Testagem Universal em testagem adaptativa é uma tarefa de alta complexidade que demanda várias etapas com delineamentos diversos e amplos recursos técnicos, tecnológicos e, sobretudo, metodológicos.

Referências

- Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2010). Technology-Enabled and Universally Designed Assessment: Considering Access in Measuring the Achievement of Students with Disabilities—A Foundation for Research. *Journal of Technology, Learning, and Assessment*, 10(5). Retrieved [date] from <http://www.jtla.org>.
- Andrade, D. F. Tavares, H. R. & Valle, (2000). Teoria da Resposta ao Item: Conceitos e Aplicações. São Paulo: SINAPE 2000. 154 p.
- Barnes, C. (2009). Un chiste “malo”: ¿rehabilitar a las personas con discapacidad en una sociedad que discapacita? In P. Brogna (Org.), *Visiones e revisiones de la discapacidad*. (pp. 101-122). México: FCE.
- Barton, L. (2009). La posición de las personas con discapacidad. ¿Qué celebrar y por qué celebrarlo? ¿Cuáles son las consecuencias para los participantes? In P. Brogna (Org.), *Visiones e revisiones de la discapacidad* (pp. 123-136). México: FCE.
- Beckman, R. M. (2001). Aplicação dos Blocos Incompletos Balanceados na teoria de resposta ao item. *Estudos em Avaliação Educacional*, nº 24. Recuperado de <http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/990/990.pdf>, em Agosto de 2016.
- Brasil (2007). *Convenção sobre os Direitos das Pessoas com Deficiência: Protocolo Facultativo à Convenção sobre os Direitos das Pessoas com Deficiência*. Brasília. Recuperado em janeiro de 2012, de www.acessibilidadeweb.com/luso/Convencao.pdf.
- Diniz, D. (2003). *Modelo social da deficiência: a crítica feminista*. Série Anís, Brasília, v. 28, p. 1-10.

- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Psychology Press.
- Fisher, S. R. A.; Yates, F. (1963). *Statistical Tables for biological, agricultural and medical research*. Sixth Edition. Recuperado de https://hekyll.services.adelaide.edu.au/dspace/bitstream/2440/10701/1/stat_tab.pdf, junho de 2015.
- Fuchs, L., Fuchs, D., Eaton, S. Hamlett, C., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children*, 67, 67-81. Recuperado de <http://ecx.sagepub.com/content/67/1/67.short>, em junho de 2015.
- Fuchs, L., Fuchs, D., Eaton, S. Hamlett, C., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children*, 67, 67-81.
- Gaster, L., & Clark, C. (1995). *A guide to providing alternate formats*. West Columbia, SC: Center for Rehabilitation Technology Services. ERIC Document; No. ED 405689. Recuperado de <http://eric.ed.gov/?id=ED405689>, em Junho de 2016.
- Gesser, M; Filgueiras Toneli, M J; Nuernberg, A H; (2012). A Contribuição do Modelo Social da Deficiência à Psicologia Social. *Psicologia & Sociedade*, 24(3): 557-566. Recuperado de <http://www.redalyc.org/articulo.oa?id=309326585026>.
- Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning and Assessment*, 4(2). Recuperado em agosto de 2011, de <http://www.jtla.org>.
- Lang, R. (2009). The United Nation Convention on the right and dignities For persons with disability: a panacea For ending disability discrimination? *European Journal of Disability*. 3, 266-285.
- Linacre, J. M. (2000). Computer-Adaptive Testing: A Methodology Whose Time Has Come. Published in Sunhee Chae, Unson Kang, Eunhwa Jeon, and J. M. Linacre. (2000) *Development of Computerized Middle School Achievement Test [in Korean]*. Seoul, South Korea: Komesa Press. Recuperado de <http://www.rasch.org/memo69.pdf>, em dezembro de 2015.

- Luecht, R. M. & Sireci, S. G. (2011). A Review of models for computer-based testing. *Measurement*. Recuperado em <http://files.eric.ed.gov/fulltext/ED562580.pdf>, em agosto de 2016.
- Magis, D., & Barrada, J. R. (2014). Computerized adaptive testing with R: Recent updates of the package catR. *Journal of Statistical Software*. Recuperado de <https://cran.r-project.org/web/packages/catR/catR.pdf>, em Junho de 2016.
- Moreira Junior, F. de J. (2011) Sistemática para Implantação de Testes Adaptativos Informatizados Baseados na Teoria da Resposta ao Item. Tese. Recuperado de <https://repositorio.ufsc.br/handle/123456789/95506>, em junho de 2015.
- Moreira Junior, F. de J.; Tezza, R.; Andrade, D. F. de; Bornia, A. C. (2013). Algoritmo de um teste adaptativo informatizado com base na teoria da resposta ao item para a estimativa da usabilidade de sites de e-commerce. *Produção*, v. 23, n. 3, p. 525-536. Recuperado de <http://dx.doi.org/10.1590/S0103-65132012005000095>, em agosto de 2016.
- Nunes, C. H. S. S.; Spenassato, D.; Bornia, A. C.; Oliveira, C. M. ; Primi, Ricardo (2015). Testes Adaptativos Computadorizados - CAT. In: Rocha da Silva, M. C.; Bartholomeu, D.; Vendramini, C. M. M.; Montiel, J. M. (Org.). *Aplicações de Métodos Estatísticos Avançados à Avaliação Psicológica e Educacional*. 1 ed. São Paulo: Vetor Editora Psico-Pedagógica LTDA, v. 1, p. 37-76.
- Oliveira, C. M. (2013). Evidências de Validade de uma Bateria Informatizada para Avaliação da Personalidade adaptada ao Desenho Universal. *Dissertação de Mestrado*, Programa de Pós-graduação em Psicologia, Universidade Federal de Santa Catarina, Florianópolis, SC. Recuperado de <https://repositorio.ufsc.br/bitstream/handle/123456789/107335/317793.pdf?sequence=1>, em junho de 2013.
- Oliveira, C. M. ; Nuernberg, A. H. & Nunes, C. H. S. S. (2013). Desenho Universal e avaliação psicológica na perspectiva dos direitos humanos. *Avaliação Psicológica*. Vol. 3. <http://goo.gl/qtkhx3>.

- Oliveira, C. M. & Nunes, C. H. S. S. (2015). Instrumentos para Avaliação Psicológica de Pessoas com Deficiência Visual: Tecnologias para Desenvolvimento e Adaptação. *Psicologia: Ciência e Profissão*, 2015, 35(3), 886-899. <http://dx.doi.org/10.1590/1982-3703001902013>.
- Pasquali, L. (2007). *Teoria de Resposta ao Item*. Brasília: LabPAM/UnB. p. 236.
- Pasquali, L. (2010). *Instrumentação Psicológica*. Porto Alegre: Artmed.
- Primi, R.; Santos, D.; John, O. P. & De Fruyt, F. (2016). Development of an Inventory Assessing Social and Emotional Skills in Brazilian Youth. *European Journal of Psychological Assessment*, 32, pp. 5-16. Hogrefe Publishing. <https://doi.org/10.1027/1015-5759/a000343>.
- Santos, W. (2010). Deficiência, desigualdade e assistência social: o Brasil e o debate internacional. In D. Diniz, *Deficiência e discriminação* (pp. 117-141), Brasília: Letras Livres: EdUnB.
- Shinn, E. & Ofiesh, N. (2012). Cognitive Diversity and Implications for Test Design. *Journal of Postsecondary Education and Disability*, 25 (3), pp. 227-245.
- The Psychometrics Centre (2014). *Concerto platform for the development of online adaptive tests*. Cambridge. Acessado em junho de 2015 de <http://www.psychometrics.cam.ac.uk/newconcerto>.
- STATA 14. Data Analysis and Statistical Software. Acessado em <http://www.stata.com/>.
- Veldkamp, B.P. and Matteucci, M. (2013). *Bayesian computerized adaptive testing*. *Ensaio: Aval. Pol. Públ. Educ.*, Rio de Janeiro, 78 (21). 57 - 82. ISSN 0104-4036. Recuperado de http://doc.utwente.nl/83906/1/aop_0313.pdf.
- Wainer, H. (2000). CATs: Whither and whence. *Psicológica*, 21: 121-133. Recuperado de <http://www.uv.es/revispsi/articulos1y2.00/wainer.pdf>, em junho de 2016.

Artigo 3. Dimensionalidade de um banco de itens de Personalidade e Testagem Universal.

Resumo

A análise fatorial é um recurso metodológico e instrumental muito utilizado no estudo da personalidade, sobretudo, no que se refere ao desenvolvimento de medidas. Estudos dessa natureza buscam verificar se os fatores identificados através da análise fatorial apresentam uma adequada correspondência com os traços e/ou as dimensões da personalidade postuladas a partir de modelos teórico-conceituais. O objetivo da pesquisa consistiu em verificar a dimensionalidade de um banco de itens de Personalidade e Testagem Universal. Os 525 itens deste estudo foram resumidos em 5 fatores por meio da análise fatorial e apresentaram relação evidente com o modelo adotado na construção dos itens, o dos Cinco Grandes Fatores da Personalidade: Abertura, Realização, Extroversão, Socialização e Neuroticismo. Os resultados obtidos foram compatíveis com os reportados na literatura científica da área.

Palavras-chave: Personalidade, Análise Fatorial; Cinco Grandes Fatores da Personalidade.

Abstract

Factorial analysis is a methodological and instrumental resource that is widely used in the study of personality, especially when it comes to its measurement. The factors measured through the factorial analysis in this case are the personality traits and / or dimensions. The objective of the research was to verify the dimensionality of a bank of Personality and Universal Test items. The 525 items of this study were summarized in 5 factors by factorial analysis. The Big Five: Openness, Realization, Extroversion, Socialization and Neuroticism, presented similar results to the settled scientific literature of the area.

Keywords: Personality; Factorial Analysis; Big Five.

Introdução

“What we seem to have overlooked is that a science, much like a bird, must have two strong wings if it is to fly . . . theory and research. . .” (Wallis, 2010, p. 74).

Na Psicologia, desde os seus primórdios, existem diversas teorias que procuram definir o que seria a *Personalidade* e cada qual expõe a sua visão de acordo com seus pressupostos teóricos e epistemológicos (Baptista, 2008; Fadiman & Frager, 2002; Schultz & Schultz, 2002). Alguns aspectos que repetidamente têm gerado controvérsias no estudo do tema são se a personalidade possui aspectos que são permanentes ou se as variáveis da personalidade são situacionais, isto é, dependem do momento ou situação vivida pelo indivíduo.

Outra questão frequente é como a personalidade deve ser abordada; se de forma nomotética ou idiográfica. Comumente surgem defensores de que a personalidade deve ser abordada apenas através das semelhanças populacionais enquanto outros preferem defender o ponto de vista das diferenças e das idiossincrasias. Por fim, há a dúvida sobre se a personalidade é predominantemente condicionada a fatores hereditários ou a fatores ambientais (Mayer, 2015; Baptista, 2008; Fadiman & Frager, 2002; Schultz & Schultz, 2002). Estas controvérsias fazem parte da teoria psicológica da personalidade e sua evolução e não é objetivo deste estudo respondê-las, mas apenas o conhecimento da sua existência como auxílio à compreensão e definição da personalidade.

No editorial do *Journal of Research in Personality*, Failowska e DeYoung (2015) afirmam que a ideia de integrar os modelos de personalidade em um único modelo que fosse prevalente aos demais tem gerado persistentes tentativas ao longo dos anos, não sendo algo novo, e citam como exemplo os trabalhos de Allport (1937), Bandura (1977), Cattell (1943), Eysenck (1970) e Murray (1938). Contudo, ainda para Failowska e DeYoung (2015) apesar de reconhecerem as contribuições destes esforços, questionam a viabilidade de uma só teoria da personalidade completa e coesa e argumentam que a diversidade teórica é uma das riquezas da ciência, já que estas várias teorias irão operar em diferentes níveis na busca de aspectos bastante distintos da realidade.

Neste mesmo número da referida revista científica, as contribuições para o crescimento e desenvolvimento da teoria da personalidade foram divididos em três categorias temáticas (Failowska & DeYoung, 2015): 1) Metateorias (análise sistemática da própria teoria da personalidade identificando uma gama de fenômenos que o termo personalidade engloba, bem como os processos em estudo e os tipos de abordagens que devem ser aplicados para sua compreensão); 2) Teorias explicativas e integrativas (teorias da personalidade sintéticas, abrangentes e mecanicistas) e 3) Métodos de integração (pesquisas que contribuem com inovações metodológicas que facilitam tanto a pesquisa

quanto os desdobramentos em teorias integrativas da personalidade). Tais eixos temáticos tentam abarcar a amplitude dos estudos de Personalidade.

Revelle e Condon (2015) descrevem a Personalidade como *“...uma abstração usada para descrever e explicar a padronização coerente ao longo do tempo e espaço dos afetos, cognições, desejos e comportamentos resultantes do que o indivíduo experiência e expressa”*. Já para Mayer (2015; 2005) a *“personalidade é o sistema organizado, desenvolvido dentro do indivíduo que representa a ação coletiva motivacional, emocional, cognitiva, de planejamento social e de outros subsistemas psicológicos”*. Em uma tentativa de síntese em relação aos questionamentos frequentes que envolvem a temática, uma definição possível de personalidade seria um conjunto de características permanentes relativamente estáveis e previsíveis influenciadas por fatores hereditários (genes), da aprendizagem e situacionais. Para esta concepção as características ou atributos de uma pessoa que compõem a sua personalidade vão além do que a define fisicamente ou superficialmente, envolvendo uma série de aspectos sociais e emocionais (Baptista, 2008; Fadiman & Frager, 2002; Schultz & Schultz, 2002).

Neste estudo, a personalidade será abordada sob a perspectiva da teoria dos traços, optando-se pela definição de Allport (1961) *apud* Nunes (2005) *“a personalidade é a organização dinâmica, no indivíduo, dos sistemas psicofísicos que determinam seu comportamento e seu pensamento característicos”*. Para este mesmo autor, sistemas psicofísicos são um conjunto de elementos em integração mútua representando potenciais para atividade, cada sistema representaria um conceito, um traço (Nunes, 2005). Os principais representantes da teoria dos traços são Allport, Cattell e Eysenck (Bartholomeu, 2008; Paim, Rossi & Richter, 2008). Os traços não são medidos diretamente e sim por meio de comportamentos ou de expressões destes comportamentos na linguagem.

Historicamente, os primeiros pesquisadores dos traços, baseados na hipótese léxica, buscaram avaliá-los a partir de seus descritores linguísticos (descritores de traços). Na linguagem falada os descritores são expressos usualmente por meio de adjetivos empregados para descrever ou caracterizar as diferenças individuais. A hipótese léxica parte da concepção de que, quando utilizam a linguagem falada, os indivíduos procuram nomear características importantes para a sua cultura, características que definem comportamentos e características que diferenciam comportamentos, formas de ser e agir socialmente, ou seja, traços de personalidade. Os estudiosos da área consideram que caso

uma característica seja importante ela será nomeada e que, portanto, através do estudo das palavras é possível chegar às dimensões da personalidade. No processo de medida de tais dimensões utiliza-se o concurso da análise fatorial (Guzzo, Pinho & Carvalho, 2002; Nunes, 2005; John, Naumann & Soto, 2007).

A análise fatorial é um recurso metodológico e instrumental muito utilizado no estudo da personalidade, sobretudo, no que se refere à mensuração da mesma. Os fatores medidos através da análise fatorial neste caso são os traços e/ou as dimensões da personalidade. Cattell chegou a um modelo explicativo da personalidade de 16 fatores, mas pesa a este fato os poucos recursos operacionais de que dispunha na época, e suas análises foram realizadas sem o auxílio do computador, utilizando régua de cálculo. Soluções com menos fatores começaram a ser propostas de forma recorrente após o advento de recursos computacionais mais potentes. Atualmente, o modelo que defende a existência de Cinco Grandes Fatores da Personalidade (*Big Five*) tem se sedimentado por meio do acúmulo de relevantes evidências científicas (Nunes, Hutz & Nunes, 2010; Hutz, Silveira, Serra, Anton & Wiczorek, 1998).

O primeiro a apresentar uma explicação teórica da personalidade a partir de cinco fatores foi McDougall, mas sem grande repercussão. Estudos empíricos foram desenvolvidos de forma isolada e sem continuidade por Thurstone (1934) e Fiske (1949) (Nunes, 2005). Ainda segundo Nunes (2005), apenas recentemente os estudos ressurgiram alcançando resultados consistentes e de reconhecida relevância no meio científico com Norman (1963), Borgatta (1964), Norman e Goldberg (1966) e Tupes e Christal (1992). No Brasil Hutz *et al.* (1998), Nunes (2005), Silva, Schlottfeldt, Rozenberg, Santos e Lelé (2007), Vasconcellos (2007), Nunes, Hutz e Nunes (2010), Andrade, J. (2008), Nunes, Hutz e Giacomoni (2009), Araújo (2010) e Primi, Santos, John e De Fruyt (2016) representam alguns dos estudos que vem sendo desenvolvidos nos últimos anos e que tem contribuído para o acúmulo de evidências de validade de instrumentos psicológicos baseados no modelo dos Cinco Grandes Fatores construídos ou adaptados à realidade brasileira, bem como para fortalecer teoricamente e empiricamente este modelo.

Acresce-se ainda que internacionalmente os desdobramentos do modelo dos Cinco Grandes Fatores alcançaram grandes proporções, abarcando temas como a sedimentação do seu modelo, a comparação deste com outros modelos estruturais para personalidade e a construção de instrumentos (John, Angleitner, & Ostendorf, 1988; McCrae & John,

1992; Zuckerman, Kuhlman, Joireman, Teta & Kraft, 1993; Goldberg & Rosolack, 1994; John & Srivastava, 1999; Srivastava, John, Gosling, & Potter, 2003; McAdams & Pals, 2006; De Young, Quilty & Perterson, 2007; Soto, John, Gosling, & Potter, 2008; Soto, John, Gosling, & Potter, 2011; John & De Fruyt, 2014; DeYoung, 2015).

Neste artigo, o modelo teórico para o desenvolvimento dos itens foi o dos Cinco Grandes Fatores (Abertura, Realização, Extroversão, Socialização e Neuroticismo). Contudo, além do CGF, aplicou-se também os princípios da Testagem Universal a qual busca a acessibilidade plena ou máxima possível aos itens e ao conteúdo do teste como um todo, como sua apresentação, organização estrutural e textos de explicação da aplicação, deste modo, permeando tanto os elementos de padronização do teste como seu conteúdo teórico (Oliveira, 2013; Oliveira; Nuernberg & Nunes, 2013). Ainda, na Testagem Universal almeja-se uma população ampla e inclusiva que abranja tanto indivíduos com quanto sem deficiência. Assim, procurou-se adequar o conteúdo dos itens para que fossem de fácil entendimento evitando-se palavras rebuscadas e em grande número, primando-se, assim, simplicidade do item (Gaster & Clark, 1995; Shinn & Ofiesh, 2012).

Método

O presente estudo enquadra-se no âmbito dos estudos de desenvolvimento de instrumentos em Avaliação Psicológica. O objetivo da pesquisa consistiu em verificar a dimensionalidade de um banco de itens de Personalidade e Testagem Universal. Este compõe um estudo mais amplo de elaboração de um banco de itens (BI) de Personalidade e Testagem Universal para Testagem Adaptativa.

Participantes

A amostra compôs-se de indivíduos maiores de idade de ambos os sexos, com e sem deficiência (n total = 1.228; n dos Itens Âncoras= 638), sendo destes 76% do sexo feminino. Os participantes foram de 24 estados brasileiros e do Distrito Federal. Os estados com o maior número de participantes foram Santa Catarina (33%) e Paraíba (27%). Os estados com menor número de participantes foram Amapá, Acre e Amazonas, cada qual com 1 participante somente – Rondônia e Roraima foram os únicos estados sem participantes. A média de idade foi de 33,4 (DP=11), com valor mínimo de 18 anos e máximo de 74 anos. Quanto à escolaridade, apenas 13% possui até o Ensino Médio Completo, 70% da

amostra possui Ensino Superior Incompleto e Completo ou Especialização, enquanto 16,9% mestrado ou doutorado.

Instrumentos

Construção dos Instrumentos

Os itens são de autorrelato com cinco opções de respostas (escala tipo Likert) e foram desenvolvidos segundo o modelo dos Cinco Grandes Fatores e os pressupostos da Testagem Universal, totalizando 525 itens. A amostra tratou-se de indivíduos maiores de idade sem e com deficiência (visual e motora). Organizaram-se os itens em uma Bateria de Personalidade e Testagem Universal (BPTU) na qual foram acrescentados 40 itens com parâmetros de qualidade já avaliada em outros estudos (itens âncora) do Laboratório de Pesquisa em Avaliação Psicológica (LPAP/UFSC). Tal bateria foi alocada na Plataforma *Concerto (Open-source Online Rbased Adaptive Testing Platform)*, a qual é uma plataforma para criação de instrumentos de código aberto que permite que os usuários criem avaliações *on-line*. Acresce-se que tais itens passaram por todas as etapas de construção de itens já sedimentadas na literatura científica da área (Pasquali, 2010), como a análise de juízes e ainda por avaliação dos critérios oriundos da Testagem Universal (Oliveira & Nunes, 2015; Oliveira; Nuernberg & Nunes, 2013).

A BPTU foi desenvolvida utilizando a estratégia de construção por Blocos Incompletos Balanceados (BIB), a qual é uma solução estatística que permite a coleta de um grande número de informações de uma quantidade extensa de itens, sendo que um mesmo testando responde a apenas uma fração destes (Fisher & Yates, 1963; Andrade, Tavares & Valle, 2000; Beckman, 2001; Moreira Junior, 2011). No modelo de BIB adotado os itens foram separados em 35 cadernos (c), 15 blocos (b), com um número de 3 (k) blocos por caderno e o número de 7 (r) repetições de cada bloco. Ainda, o fator de utilização (fu) foi de 0,2, ou seja, 20% dos testandos foi submetido a cada bloco (Fisher & Yates, 1963; Andrade, Tavares & Valle, 2000; Beckman, 2001; Moreira Junior, 2011). Assim, os 525 itens foram distribuídos no banco em 35 itens por bloco e 105 itens em cada caderno ($525/15=35$; $3*35=105$). Ao final da verificação de dimensionalidade, relatada neste artigo, foi formado o Banco de Itens (preliminar) de Personalidade e Testagem Universal.

Procedimentos

As coletas foram realizadas *on-line* e o *link* para tal coleta foi amplamente divulgado por *e-mails* e redes sociais (institucionais e individuais). Ainda, contactou-se a Associação Catarinense para Integração do Cego (ACIC) e divulgou-se o *link* em listas de *e-mails* desta associação. Complementa-se que este estudo foi submetido ao comitê de ética em pesquisas com seres humanos (CEPSH/UFSC), por meio da Plataforma Brasil/DATASUS, estando aprovado com número CAAE 51788015.6.0000.012.

Análise de Dados

Foram realizados testes de suposição de análise fatorial (testes de pré-estimativa) para averiguar a possibilidade de realização das mesmas e testes de pós-estimativa para verificar a utilização do modelo de análise fatorial. Deste modo, as etapas da análise fatorial realizadas foram: 1) Verificação das condições para análise fatorial; 2) Seleção do tipo e delineamento da análise - determinação do número de fatores; 3) Seleção e eliminação de itens - definição dos fatores. Tais etapas serão detalhadas mais adiante neste artigo. Os programas utilizados nas análises de dados foram o STATA 14 e o Microsoft Excel 2010.

Resultados

Verificação das condições para análise fatorial

A análise fatorial possui certa sensibilidade aos *outliers* e, portanto, procedeu-se à identificação dos *outliers* multivariados pela técnica Hadi (Hadi, 1992) em cada um dos 15 cadernos ($p=0,05$), tendo sido contabilizados apenas 43 casos dentre as 1.228 observações. Destes, 36 casos estavam presentes em um dos Blocos Incompletos Balanceados (BIB) e os demais 7 em pares de Blocos (Tabela 1). Praticamente em todos os casos, a identificação como *outliers* deveu-se à pouca variabilidade nas respostas com acentuado número de respostas extremas 1 ou 5 (Figueira, 1998; Hadi, 1992). Os casos foram julgados e retirados os que se fizeram necessários (apenas dois), os demais foram mantidos por considerar-se que poderiam possuir dados relevantes para compreensão da amostra, ou seja, indivíduos que, da perspectiva dos cinco fatores, realmente estão nos extremos medidos pelos itens.

Tabela 1.
Outliers por caderno.

Caderno	Número de Observações	Número de <i>Outliers</i> por cadernos
Caderno 1	182	2
Caderno 2	194	2
Caderno 3	194	0
Caderno 4	192	4
Caderno 5	213	7
Caderno 6	248	5
Caderno 7	176	2
Caderno 8	211	6
Caderno 9	206	4
Caderno 10	171	1
Caderno 11	230	5
Caderno 12	184	0
Caderno 13	187	5
Caderno 14	196	2
Caderno 15	217	5
Total	1228	36+7 pares = 43

Ainda, utilizou-se o critério de Kaiser-Meyer-Olkin (KMO) para verificar se a matriz de dados é passível de fatoração, cujo processo será descrito no próximo tópico.

Seleção do tipo e delimitação da análise fatorial – extração dos fatores

Primeiramente, foi realizada a análise fatorial exploratória, a qual explora e resume a estrutura correlacional de um conjunto de dados (Damásio, 2012; Pasquali, 1999; Hair, Black, Babin, Anderson & Tatham, 2009). Tal procedimento foi realizado nos itens com parâmetros já conhecidos (itens âncoras) e, assim, diferentes métodos de extração foram adotados, demonstrando que uma solução com cinco fatores era apropriada (Eigenvalue >1 segundo o critério de Kaiser-Guttman (Damásio, 2012)). A solução mais simples adotada por Análise de Fatores Comuns sem rotação encontra-se, como exemplo, no Gráfico 1 – a mesma já apontou para uma solução de cinco fatores.

O Gráfico 1 corresponde a um *screen plot*. Neste gráfico os *eigenvalues* dos fatores são colocados no eixo Y e no eixo X está o número de fatores. O último fator a ser extraído é o localizado no ponto de inflexão da curva representada graficamente (Hair, Black, Babin, Anderson & Tatham, 2009). Os modelos de análise fatorial (Análise de Fatores Comuns e Análise de Componentes Principais) foram testados seguindo-se os testes de rotações, tanto oblíqua quanto ortogonal. O modelo fatorial que melhor ajustou os dados foi o de Análise de Componentes Principais (CP) com rotação ortogonal do tipo Parsimax,

acresce-se que se primou por cargas fatoriais maiores que 0,30 (vide Gráfico 2). Para os itens âncoras os *engeinvalues* por fatores foram: Neuroticismo (4,13); Extroversão (3,46); Abertura (3,13); Realização (3,00); e, Socialização (2,65) e a variância explicada por esta solução de cinco fatores foi de 40,99% .

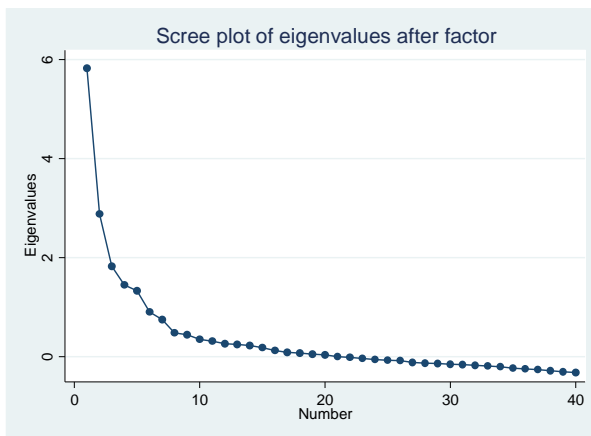


Gráfico 1. Cinco Fatores: análise de Fatores Comuns (STATA 14).

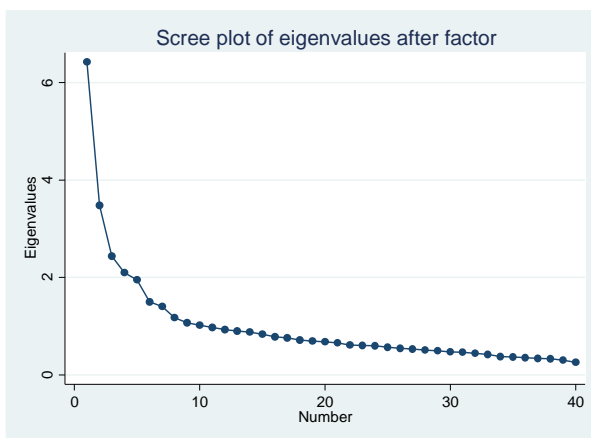


Gráfico 2. Cinco Fatores: CP Parsimax (STATA 14).

Na análise fatorial dos 15 cadernos utilizou-se o mesmo modelo implementado nos Itens Âncoras – Análise de Componentes Principais (CP) para 5 fatores com rotação ortogonal do tipo Parsimax. Com vistas

a potencializar o poder amostral criou-se um escore para cada fator a partir da análise fatorial realizada com os Itens Âncoras (Os escores foram gerados pela média dos itens âncoras selecionados considerando o seu sentido). Tais escores dos fatores foram acrescidos nas análises fatoriais realizadas em cada caderno separadamente. Assim, realizou-se análise fatorial exploratória por caderno.

Complementarmente, objetivando verificar a adequação dos resultados da amostra à realização das análises fatoriais, foram calculados os KMOs (Kaiser-Meyer-Olkin) para cada um dos cadernos após a realização da já referida Análise de Componentes Principais, também por caderno. Resultados muito baixos de KMO indicam que o conjunto de variáveis possui pouco em comum para permitir a realização de uma AF. Assim, valores de KMO de 0,00 a 0,49 são inaceitáveis, entre 0,50 e 0,59 são muito baixos, já de 0,60 a 0,69 são medianos e de 0,80 a 0,89 são bons e de 0,90 a 1,00 são excelentes.

Na Tabela 2 são apresentados os resultados dos KMOs por caderno e sua interpretação. Dentre os cadernos, 3 apresentaram KMOs medianos, 8 foram baixos e 4 muito baixos. Na tentativa de ampliar os valores de tal estimativa, os cadernos que apresentaram valores muito baixos foram analisados pela observação da matriz de covariância dos itens dos cadernos e alguns itens com valores inferiores a 0.5 foram retirados. Os 4 cadernos passaram de valores muito baixos de KMO para valores baixos (vide Tabela 2).

Tabela 2.

Valores de KMO por caderno.

Caderno	KMO	Interpretação	KMO	Interpretação	Retirados
Caderno 1	0.5954	muito baixa	0.6169	baixa	3
Caderno 2	0.6701	Baixa	-	-	-
Caderno 3	0.5982	muito baixa	0.6015	baixa	3
Caderno 4	0.5949	muito baixa	0.6192	baixa	3
Caderno 5	0.6443	Baixa	-	-	-
Caderno 6	0.7026	Mediano	-	-	-
Caderno 7	0.6210	Baixa	-	-	-
Caderno 8	0.6968	Baixa	-	-	-
Caderno 9	0.6672	Baixa	-	-	-
Caderno 10	0.7039	mediano	-	-	-
Caderno 11	0.6142	Baixa	-	-	-
Caderno 12	0.6234	baixa	-	-	-
Caderno 13	0.7026	mediano	-	-	-
Caderno 14	0.5713	muito baixa	0.6005	baixa	12
Caderno 15	0.6692	baixa	-	-	-

Seleção e eliminação de itens - definição dos fatores

Os itens foram nomeados conforme os fatores em: fator 1 = abertura; fator 2 = socialização; fator 3 = extroversão; fator 4 = neuroticismo; e, fator 5 = Realização. Nos 15 cadernos houve correspondência entre o fator do escore dos itens âncoras e os itens em seus fatores com ajustes variados (vide exemplo abaixo do Caderno 10). A partir dos resultados das análises fatoriais os itens permaneceram ou foram retirados em cada caderno (Tabela 3).

Os critérios de exclusão ou permanência dos itens no Banco de Itens foram: 1) itens que não carregaram em nenhum fator foram retirados; 2) itens que carregaram em mais de um fator foram julgados quanto ao pertencimento a um ou outro fator, se as cargas eram mais baixas ($>0,30$ e $<0,50$) ou semelhantes (por exemplo: 0,44 e 0,36). Julgou-se o item quanto ao seu conteúdo e alguns destes foram retirados do BI; 3) Ainda, quando o item carregou em um fator diferente do previsto durante o seu desenvolvimento e com um valor de mediano a alto ($>0,40$), observou-se se fazia sentido teórico a modificação do fator, assim o fator poderia ser alterado ou, em alguns casos, nos quais tal modificação não faziam sentido teórico, os itens foram retirados; 4) acresce-se que alguns itens com cargas próximas a 0,30 ou que pertenciam a algum outro critério de exclusão permaneceram no banco de itens. Isso ocorreu quando julgou-se a qualidade teórica do mesmo. Tal fato deu-se com itens mais intensos/ostensivos ou menos intensos/ostensivos, os quais possuem um caráter clínico e, portanto, requerem a sua aferição por amostras especificamente clínicas e diferentes da utilizada neste estudo – população em geral.

No caderno 10, que utilizaremos como exemplo, os *engenvales* foram: Fator1 = Extroversão (4,52); Fator2 = Neuroticismo (4,06); Fator3 = Realização (3,88); Fator4 = Abertura (3,14); e, Fator5 = Socialização (3,10), explicando 46,83% da variância e o KMO foi de 0.70. Exemplos de itens típicos para cada um dos fatores neste caderno e suas respectivas cargas fatoriais são: Extroversão i334c10f3fc2 (0,71) “*Meu entusiasmo é contagiante.*”; Neuroticismo i338c10f4fc1 (0,67). “*As situações da vida dão errado para mim.*”; Realização i346c10f5fc1 (0,68) “*Sou disciplinado.*”; Abertura i318c10f1fc1 (0,76) “*Costumo assistir programas sobre outras culturas.*”; Socialização i325c10f2fc1 (0,71) “*Meus familiares dizem que eu sou amável com todos.*” (Figura 1).

Na análise fatorial dos itens segundo os critérios de exclusão ou permanência no Caderno 10, o item i316c10f1fc1 (Figura 1) carregou em dois fatores e possuiu cargas fatoriais baixas em extroversão (0,36) e

realização (0,30). Em decorrência destes resultados o item foi retirado do Banco de Itens. O item i319c10f1fc2 (Figura 1) apresentou carga $<0,30$ e por isso seu valor está representado em branco na tabela o que indica a sua exclusão do BI. Já o item i320c10f1fc2 carregou em neuroticismo (0,52) e tinha sido desenvolvido para o fator abertura invertido “*Quando penso demais me confundo.*”. Analisou-se o conteúdo do item e optou-se por retirá-lo por não fazer sentido teórico a mudança de fator como também ocorreu com o item i330c10f3fc1 “*Organizo as festas na minha família.*”, o qual foi elaborado para o fator extroversão e carregou em abertura, sendo assim excluído.

O item i339c10f4fc1 “*Confio na minha capacidade de resolver os problemas da minha vida.*” que possui carga fatorial alta em extroversão (0,50) e neuroticismo (-0,49), originalmente havia sido desenvolvido para neuroticismo invertido. Pelo carregamento duplamente alto optou-se por retirar o item do banco. Já o item i324c10f2fc1 “*Dou do meu tempo para escutar os problemas de outras pessoas.*”, apesar de ter carregado em dois fatores com cargas semelhantes, foi mantido no BI na sua faceta de origem (socialização) por supô-lo como item intenso/ostensivo e que, deste modo requer novos estudos. O item i326c10f2fc2 foi desenvolvido para o fator socialização “*Sempre que precisei apareceu alguém para ajudar.*” e carregou em extroversão (0,44) e socialização (0,36), tendo maior carga fatorial em extroversão. Deste modo, optou-se pela retirada do item por não fazer sentido teórico a sua modificação para o fator extroversão. O item i327c10f2fc2 “*Geralmente as pessoas se aproveitam da fragilidade das outras.*” que originalmente havia sido elaborado para socialização carregou em neuroticismo e considerou-se haver razão teórica para tal, portanto, modificou-se o fator no BI. Por fim, o item i341c10f4fc2 “*Fico tranquilo antes de provas ou trabalhos.*” desenvolvido para neuroticismo invertido carregou em neuroticismo (-0,35) e em extroversão (0,51). Devido à superioridade do valor da carga fatorial em extroversão sem consonância com este fator optou-se pela exclusão do mesmo do BI. Os mesmos critérios de análise e interpretação implementados no caderno 10 foram aplicados aos demais 14 cadernos e os resultados estão resumidos na Tabela 3.

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
anc_a				0.7264		0.3842
anc_e	0.7107					0.3931
anc_n		0.6617				0.3322
anc_r			0.7936			0.3360
anc_s					0.7047	0.4201
i316c10f1fc1	0.3657		0.3048			0.7043
i317c10f1fc1				0.4404	-0.3350	0.6863
i318c10f1fc1				0.7585		0.3328
i319c10f1fc2						0.8799
i320c10f1fc2		0.5280				0.6424
i321c10f1fc3				0.6183		0.5015
i322c10f1fc3				0.6071		0.5534
i323c10f2fc1				0.3794	0.4167	0.6404
i324c10f2fc1				0.4159	0.4027	0.5838
i325c10f2fc1					0.7148	0.4672
i326c10f2fc2	0.4452				0.3639	0.6076
i327c10f2fc2		0.4227				0.7779
i328c10f2fc3					0.7190	0.4620
i329c10f2fc3					0.3697	0.7604
i330c10f3fc1				0.3805		0.7422
i331c10f3fc1	0.4501					0.6266
i332c10f3fc1	0.6071					0.5514
i333c10f3fc2	0.6257	0.3104				0.4583
i334c10f3fc2	0.7157					0.3969
i335c10f3fc3	0.8262					0.2800
i336c10f3fc3	-0.4165				-0.4850	0.5063
i337c10f4fc1	-0.3474	0.5714				0.5006
i338c10f4fc1		0.6716				0.4935
i339c10f4fc1	0.5099	-0.4958				0.4265
i340c10f4fc2		0.6087				0.5831
i341c10f4fc2	0.5137	-0.3558				0.5388
i342c10f4fc3		0.6692				0.4626
i343c10f4fc3		0.3883				0.7433
i344c10f5fc1		0.3651	-0.4019			0.6524
i345c10f5fc1		-0.4006	0.6111			0.4119
i346c10f5fc1			0.6848			0.4292
i347c10f5fc2			0.6513			0.5581
i348c10f5fc2		0.3142	0.6360			0.4649
i349c10f5fc3			0.5735			0.5593
i350c10f5fc3	0.3887		0.5710			0.4166

(blanks represent $\text{abs}(\text{loading}) < .3$)

Figura 1. Fatores de saturação dos itens em cada Fator/Componente do Caderno.

Nos cadernos, o número de itens retirados variou de 4, valor mínimo no caderno 2, a 16, valor máximo no caderno 14 (Tabela 3). Quanto à mudança de fator, nos cadernos 3, 10, 11 e 14, apenas 1 item teve seu fator modificado, sendo o valor máximo de 6 itens no caderno 4. Tais modificações devem-se ao grande número de itens elaborados

(525) e às similaridades entre alguns fatores, ou entre estes e o inverso do outro, como ocorreu com neuroticismo em correspondência com socialização e abertura. Estas perdas e mudanças já eram esperadas. No total apenas 47 itens dos 525 tiveram seu fator modificado e 135 foram retirados. Permaneceram no banco 390 itens. O número de itens que permanecem no BI por fator é apresentado na Tabela 4.

Tabela 3.

Itens que tiveram seus fatores modificados e Itens excluídos por Caderno.

Caderno	Número de itens	Itens retirados	Itens que mudaram de fator	Total de itens remanescentes
Caderno 1	35 Itens	6	3	29
Caderno 2	35 Itens	4	5	31
Caderno 3	35 Itens	10	1	25
Caderno 4	35 Itens	9	6	26
Caderno 5	35 Itens	7	2	28
Caderno 6	35 Itens	7	5	28
Caderno 7	35 Itens	11	3	24
Caderno 8	35 Itens	14	4	21
Caderno 9	35 Itens	5	5	30
Caderno 10	35 Itens	7	1	28
Caderno 11	35 Itens	7	1	28
Caderno 12	35 Itens	8	2	27
Caderno 13	35 Itens	10	5	25
Caderno 14	35 Itens	16	1	19
Caderno 15	35 Itens	12	3	23
Total	525 Itens	135	47	392

Tabela 4.

Número de itens que permaneceram no BI por fator.

Fator	Número de itens	Número de itens que permaneceram
Abertura	113	78
Socialização	91	69
Extroversão	94	67
Neuroticismo	112	89
Realização	115	89

Considerações Finais

O estudo da dimensionalidade dos itens a partir da análise fatorial permite examinar os padrões de relações latentes para um grande número de variáveis possibilitando a compreensão dos dados de forma mais efetiva expressando-os ou resumindo-o em fatores (Hair, Black, Babin, Anderson & Tatham, 2009). Assim, de um número amplo de itens, 525 neste estudo, foi possível resumi-los em 5 fatores. Cada um

dos cinco fatores: Abertura, Realização, Extroversão, Socialização e Neuroticismo, apresentou resultados semelhantes à literatura científica sedimentada da área (Nunes, Hutz & Nunes, 2010; John, Naumann & Soto, 2007) fato este que contribui para o desenvolvimento teórico sobre a estrutura inerente aos construtos em Psicologia (Pasquali, 1999) – um dos objetivos da realização de análises fatoriais.

No fator Abertura, para os 15 cadernos, 10 itens carregaram em extroversão, também em socialização ou realização. Ao julgarem-se os casos, percebeu-se haver relação entre os altos níveis de abertura e de extroversão, bem como a dificuldade de elaborar itens intensos/ostensivos para ambos que expressassem teoricamente apenas um dos fatores. A literatura científica da área aponta correlações entre Extroversão e Abertura (Nunes, Hutz & Nunes, 2010; John, Naumann & Soto, 2007). Assim, itens de abertura podem carregar em extroversão com valores baixos e do mesmo modo alguns itens de extroversão carregaram em abertura. Nesta pesquisa, os itens de abertura que carregaram em socialização ou realização não permaneceram no BI por não encontrar-se relações teóricas fortes entre o conteúdo expresso nos itens que justificassem tais cargas fatoriais.

Para o fator neuroticismo, alguns itens carregaram em abertura, extroversão e socialização. Os que carregaram com cargas menores e negativas foram mantidos. Tal resultado foi mais recorrente nos cadernos para neuroticismo e extroversão, porém, apenas permaneceram no BI os itens que faziam sentido teórico.

O estudo implementado acumulou indícios da replicabilidade do modelo dos Cinco Grandes Fatores para a amostra maior de idade. O Banco de Itens desenvolvido demonstrou em seus 15 cadernos a viabilidade de uma solução de cinco fatores, bem como que os itens criados para cada fator relacionaram-se à dimensão para a qual foram construídos. Tal fato demonstra a consistência teórica do modelo e a qualidade da correspondência da mesma com os itens desenvolvidos para cada um dos fatores. Em novos estudos pretende-se ampliar a amostra em cada um dos 15 cadernos e realizar o estudo da dimensionalidade em nível de facetas.

Referências

Andrade, D. F. Tavares, H. R. & Valle, (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações*. São Paulo: SINAPE 2000. 154 p.

- Andrade, J. M. de. (2008). *Evidências de Validade do Inventário dos Cinco Grandes Fatores de Personalidade para o Brasil*. Tese de Doutorado. Universidade de Brasília. Pós-graduação em Psicologia Social, do trabalho e das organizações. Recuperado em Outubro de 2011, de <http://goo.gl/ApqXo6>.
- Araújo, R. (2010). *Relação entre bem-estar subjetivo e o fator realização no modelo dos cinco grandes fatores de personalidade*. Recuperado em Outubro de 2011, de <http://goo.gl/tqIgiG>.
- Baptista, N. J. M. (2008). *Teorias da Personalidade*. Recuperado em Setembro de 2011, de www.Psicologia.com.pt.
- Bartholomeu, D. (2008). Traços de Personalidade e Comportamentos de risco no trânsito: um estudo correlacional. *Psicologia Argumento*. Curitiba: jul./set., 26(54), p.193-206. Recuperado em Outubro de 2011, de <http://goo.gl/W6Wb4k>.
- Beckman, R. M. (2001). Aplicação dos Blocos Incompletos Balanceados na teoria de resposta ao item. *Estudos em Avaliação Educacional*, nº 24. Recuperado de <http://www.fcc.org.br/pesquisa/publicacoes/ea/arquivos/990/990.pdf>, em Agosto de 2016.
- Damásio, B. F. (2012). Uso da análise fatorial exploratória em Psicologia. *Avaliação Psicológica*, 11(2), 213-228. Recuperado em 03 de novembro de 2016, de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712012000200007&lng=pt&tlng=pt.
- DeYoung, C. G. (2016) Cybernetic Big Five Theory. *Journal of Research in Personality* 56 (2015) 33–58. Recuperado em janeiro de 2016, de: <http://www.sciencedirect.com/science/article/pii/S009265661400713>.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, 93(5), 880–896. <http://dx.doi.org/10.1037/0022-3514.93.5.880>.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality*

- and Social Psychology*, 93(5), 880–896.
<http://dx.doi.org/10.1037/0022-3514.93.5.880>.
- Fadiman, J. & Frager, R. (2002). *Teorias da Personalidade*. São Paulo: Harbra. 393p.
- Failowska, M. & DeYoung, C.G. (2015). Introduction to the special issue on integrative theories of personality. Editorial. *Journal of Research in Personality*. 56 (2015) 1–3.
- Figueira, M. C. (1998). Identificação de outliers. *Millenium n° 12 Arquivos*. Recuperado em Outubro de 2016 de <http://www.ipv.pt/millenium/arq12.htm>.
- Fisher, S. R. A.; Yates, F. (1963). *Statistical Tables for biological, agricultural and medical research*. Sixth Edition. Recuperado de https://hekyll.services.adelaide.edu.au/dspace/bitstream/2440/10701/1/stat_tab.pdf, junho de 2015.
- Gaster, L., & Clark, C. (1995). *A guide to providing alternate formats*. West Columbia, SC: Center for Rehabilitation Technology Services. ERIC Document; No. ED 405689. Recuperado de <http://eric.ed.gov/?id=ED405689>, em Junho de 2016.
- Goldberg, L. R., & Rosolack, T. K. (1994). The big five factor structure as an integrative framework: An empirical comparison with eysenck's P-E-N model. In R. P. Martin (Ed.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 7–35). Hillsdale, NJ England: Lawrence Erlbaum Associates Inc.
- Goldberg, L. R., & Rosolack, T. K. (1994). The big five factor structure as an integrative framework: An empirical comparison with eysenck's P-E-N model. In R. P. Martin (Ed.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 7–35). Hillsdale, NJ England: Lawrence Erlbaum Associates Inc.
- Guzzo, R. S. L.; Pinho, C. C. M. de; Carvalho, C. F. de (2002). Construção da Taxonomia para Descritores da Personalidade. *Psicologia: Reflexão e Crítica*. Porto Alegre, 15(1), pp. 71-75. Recuperado em Outubro de 2011, de <http://goo.gl/WmCdB2>.

- Hadi, A.S. (1992). Identifying multiple outliers in multivariate data. *J. R. Statist. Soc. B*, 54, 761-771. Recuperado em Outubro de 2016 de: <http://venus.unive.it/romanaz/datamin/articoli/hadi.pdf>.
- Hair Jr, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. (2009). *Análise Multivariada de Dados*. 6. ed. Porto Alegre: Bookman, 688p.
- Hutz, C. S., Nunes, C. H. S. S., Silveira, A. D., Serra, J., Anton, M., & Wieczorek, L. S. (1998). O desenvolvimento de marcadores para a avaliação da personalidade no Modelo dos Cinco Grandes Fatores. *Psicologia: Reflexão e Crítica*, 11, 395-409. <http://dx.doi.org/10.1590/S0102-79721998000200015>.
- Hutz, C. S.; Nunes, C. H.; Silveira, A. D.; Serra, J.; Anton, M.; Wieczorek, L. S. (1998). O desenvolvimento de marcadores para a avaliação da personalidade no modelo dos cinco grandes fatores. *Psicologia Reflexão e Crítica*. Porto Alegre: Vol. 11, n.2. Recuperado em Outubro de 2011, de <http://goo.gl/xmyDtm>.
- John, O. P., & De Fruyt, F. (2014, September). An integrative framework for the assessment of social-emotional skills. Lecture presented at the Institute Ayrton Senna, Sao Paulo.
- John, O. P., & Srivastava, S. (1999). *The Big Five trait taxonomy: History, measurement, and theoretical perspectives* (Handbook of personality: Theory and research). Retirado em http://www.rc.usf.edu/~jedorio/Personality/BIG_FIVE_TRAIT_TAXONOMY.pdf
- John, O. P., & Srivastava, S. (1999). *The Big Five trait taxonomy: History, measurement, and theoretical perspectives* (Handbook of personality: Theory and research). Retirado em http://www.rc.usf.edu/~jedorio/Personality/BIG_FIVE_TRAIT_TAXONOMY.pdf
- John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, 2, 171-203. doi: 10.1002/per.2410020302
- John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality*, 2, 171-203. doi: 10.1002/per.2410020302

- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory - versions 4a and 54*. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big-Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W.
- Mayer, J. D. (2005). A tale of two visions: Can a new view of personality help integrate psychology? *American Psychologist*, 60(4), 294–307. <http://dx.doi.org/10.1037/0003-066X.60.4.294>.
- Mayer, J. D. (2015). The personality systems framework: Current theory and development. *Journal of Research in Personality* 56; 4–14. Recuperado em Janeiro de 2016, de: <http://www.sciencedirect.com/science/article/pii/S009265661500021>.
- McAdams, D. P., & Pals, J. L. (2006). A new big five: Fundamental principles for an integrative science of personality. *American Psychologist*, 61(3), 204–217. [http:// dx.doi.org/10.1037/0003-066X.61.3.204](http://dx.doi.org/10.1037/0003-066X.61.3.204).
- McAdams, D. P., & Pals, J. L. (2006). A new big five: Fundamental principles for an integrative science of personality. *American Psychologist*, 61(3), 204–217. [http:// dx.doi.org/10.1037/0003-066X.61.3.204](http://dx.doi.org/10.1037/0003-066X.61.3.204).
- McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality*, 60, 175-216. doi:10.1111/j.1467-6494.1992.tb00970.x
- McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality*, 60, 175-216. doi:10.1111/j.1467-6494.1992.tb00970.x
- Moreira Junior, F. de J. (2011) Sistemática para Implantação de Testes Adaptativos Informatizados Baseados na Teoria da Resposta ao Item. Tese. Recuperado de <https://repositorio.ufsc.br/handle/123456789/95506>, em junho de 2015.
- Nunes, C. H. S. da S. (2005). *Construção, normatização e validação das escalas de socialização e Extroversão no modelo dos cinco grandes fatores*. Tese de doutorado, Curso de Pós-graduação em

Psicologia do Desenvolvimento. Universidade Federal do Rio grande do Sul. Recuperado em Outubro de 2011, de <http://goo.gl/Uudr0l>.

- Nunes, C. H. S. da S., Hutz, C. S. & Nunes, M. F. O. (2010). *Bateria Fatorial de Personalidade – BFP*. Manual Técnico. São Paulo: Casa do Psicólogo.
- Nunes, C. H. S., Hutz, C. S. & Giacomoni, C. H. (2009). Associação entre bem estar subjetivo e personalidade no Modelo dos Cinco Grandes Fatores. *Avaliação Psicológica*. Porto Alegre: vol. 8, pp. 99-108. Recuperado em Outubro de 2011, de <http://goo.gl/H5udmi>.
- Oliveira, C. M. & Nunes, C. H. S. S. (2015). Instrumentos para Avaliação Psicológica de Pessoas com Deficiência Visual: Tecnologias para Desenvolvimento e Adaptação. *Psicologia: Ciência e Profissão*, 2015, 35(3), 886-899. <http://dx.doi.org/10.1590/1982-3703001902013>.
- Oliveira, C. M. (2013). Evidências de Validade de uma Bateria Informatizada para Avaliação da Personalidade adaptada ao Desenho Universal. *Dissertação de Mestrado*, Programa de Pós-graduação em Psicologia, Universidade Federal de Santa Catarina, Florianópolis, SC.
- Oliveira, C. M. ; Nuernberg, A. H. & Nunes, C. H. S. S. (2013). Desenho universal e avaliação psicológica na perspectiva dos direitos humanos. *Avaliação Psicológica*. Vol. 3. <http://goo.gl/qtkhx3>.
- Paim, M. C. C., Rossi, D. & Richter, D. S. (2008). *Temperamento e traços de personalidade: uma abordagem*. Revista Digital: efdeportes. Buenos Aires: año 13, n. 124.
- Pasquali, L. (1999). Instrumentos psicológicos: manual prático de elaboração. Brasília: LabPAM/ IBAPP.
- Pasquali, L. (2010). *Instrumentação Psicológica*. Porto Alegre: Artmed.
- Primi, R.; Santos, D.; John, O. P. & De Fruyt, F. (2016). Development of an Inventory Assessing Social and Emotional Skills in Brazilian Youth. *European Journal of Psychological Assessment*; Vol. 32(1):5–16. DOI: 10.1027/1015-5759/a000343. <https://biblio.ugent.be/publication/7280734/file/7280735>.

- Revelle, W. & Condon, D. M. (2015). A model for personality at three levels. *Journal of Research in Personality*, 56 (2015) 70–81. Recuperado em Janeiro de 2016, de: <http://personality-project.org/revelle/publications/models3levels.else.pdf>.
- Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and research* (pp. 114-158). New York, NY: Guilford Press.
- Schultz, D. P. & Schultz, S. E. (2002). *Teorias da Personalidade*. São Paulo: Pioneira. 528 p.
- Shinn, E. & Ofiesh, N. (2012). Cognitive Diversity and Implications for Test Design. *Journal of Postsecondary Education and Disability*, 25 (3), pp. 227-245.
- Silva, R. S., Schlottfeldt, C. G., Rozenberg, M. P., Santos, M. T. & Lelé, A. J. (2007). *Replicabilidade do Modelo dos Cinco Grandes Fatores em medidas da personalidade*. Mosaico: Estudos em Psicologia. Vol. I, n.1, 37-49. Recuperado em Outubro de 2011, de <http://goo.gl/qRbDWx>.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *J Pers Soc Psychol*, 94(4), 718-737. doi:10.1037/0022-3514.94.4.718.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *J Pers Soc Psychol*, 94(4), 718-737. doi:10.1037/0022-3514.94.4.718
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *J Pers Soc Psychol*, 100(2), 330-348. doi:10.1037/a0021717.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *J Pers Soc Psychol*, 100(2), 330-348. doi:10.1037/a0021717
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set

like plaster or persistent change? *Journal of Personality and Social Psychology*, 84, 1041-1053. doi:10.1037/0022-3514.84.5.1041.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, 84, 1041-1053. doi:10.1037/0022-3514.84.5.1041.

Vasconcellos, S. J. L. (2007). *Os pressupostos evolutivos dos fatores Abertura à experiência e Neuroticismo*. Tese de doutorado. Programa de Pós-graduação em Psicologia. Porto Alegre. Recuperado em Outubro de 2011, de <http://goo.gl/ukRYTM>.

Wallis, S. E. (2010). Toward a science of metatheory. *Integral Review*, 6, 73–120. Recuperado em Outubro de 2016, de: <http://www.integral-review.org/documents/old/Wallis,Toward-a-Science-of-Metatheory,Vol.6,No.3.pdf>.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The big three, the big five, and the alternative five. *Journal of Personality and Social Psychology*, 65(4), 757–768. <http://dx.doi.org/10.1037/0022-3514.65.4.757>.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The big three, the big five, and the alternative five. *Journal of Personality and Social Psychology*, 65(4), 757–768. <http://dx.doi.org/10.1037/0022-3514.65.4.757>.

Artigo 4. Calibração dos Itens de um Banco de Itens de Personalidade e Testagem Universal.

Resumo

Na Teoria de Resposta ao Item o desempenho do indivíduo em um teste é predito a partir do cálculo da sua aptidão ou do traço latente. O processo do cálculo das estimativas dos parâmetros dos itens em TRI, calibração, objetiva desenvolver modelos que relacionem as propriedades dos itens e o nível de habilidade dos respondentes. A estimativa dos parâmetros constitui-se do ponto de vista estatístico em dois passos ou etapas. Primeiro, a estimativa das curvas características operacionais e segundo a estimativa das curvas de categorias de resposta em modelos desenvolvidos para itens politômicos. Neste estudo foi utilizado o modelo de resposta gradual, também referido como modelo de Samejima, o qual consiste em uma extensão do modelo de 2 parâmetros da TRI e aplica-se a itens com categorias ordinais. O objetivo consistiu em calibrar um banco de itens de Personalidade e Testagem Universal. Os itens tiveram seus parâmetros estimados, separadamente, em cada um dos cinco fatores resultando em: 64 itens de Abertura para o BI, 58 itens de Extroversão no BI, 82 itens de Neuroticismo, 76 de Realização e 37 de Socialização. Conclui-se, portanto, que 317 itens estão aptos a compor o Banco de Itens de Personalidade e Testagem Universal.

Palavras-chave: Teoria de Resposta ao Item; Banco de Itens; Calibração; Personalidade; Testagem Universal.

Abstract

In Item Response Theory the performance of the individual in a test is predicted from the calculation of his fitness or latent trait. The process of estimation of the parameters of the items in TRI, calibration, aims at models that relate the properties of the items and the skill level of the respondents. The estimation of the parameters is constituted from the statistical point of view in two steps or steps. First, the estimation of the operational characteristic curves and according to the estimation of the curves of response categories. In this study, the gradual response or Samejima model was used, which consists of an extension of the 2-parameter model of the TRI and applies to items with ordinal categories. The objective was to calibrate the items of a bank of Personality and Universal Test items. The items had their parameters estimated separately in each of the five factors, resulting in: 64 BI Opening Items, 58 BI Extroversion Items, 82 Neuroticism Items, 76 Achievement Items,

and 37 Socialization Items. It is concluded, therefore, that 317 items are eligible to compose the Bank of Personality and Universal Testing Items.

Keywords: Item Response Theory; Bank of Items; Calibration; Personality; Universal Testing.

Introdução

A Teoria de Resposta ao Item (TRI) surgiu como uma alternativa para resolução de vários problemas da Teoria Clássica dos Testes (TCT), não veio substituí-la, mas complementá-la. A TRI se insere no âmbito das teorias do traço latente que consistem em assumir que os comportamentos medidos por meio dos itens são uma representação destes traços. Os traços latentes são variáveis hipotéticas não observáveis e causa do comportamento observado o qual é medido pelos itens de um teste (Pasquali, 2010; Pasquali, 2007; Pasquali, 2003).

Na TRI o desempenho do indivíduo em um teste é predito a partir do cálculo da sua aptidão ou do traço latente. Assim, é possível estimar a probabilidade que o indivíduo terá de apresentar certos padrões de respostas em um conjunto de itens considerando seu nível estimado no traço latente, o que pode ser estudado, para cada item, através de um gráfico chamado de Curva Característica do Item – CCI. Algumas das vantagens de utilização da TRI ao invés da TCT segundo Pasquali (2007) são que a aptidão do sujeito e os parâmetros dos itens são calculados de forma relativamente independentemente, permitindo um estudo mais apurado das características dos testes, independente do grupo que faz parte da amostra.

Os estudos envolvendo a TRI têm utilizado principalmente três modelos básicos: 1 parâmetro, 2 parâmetros e 3 parâmetros. No primeiro modelo avalia-se apenas um parâmetro do item – a dificuldade – no cálculo da probabilidade de acerto do item, sendo o modelo mais simples. Já o segundo (2 parâmetros) abarca os parâmetros dificuldade e discriminação, enquanto o terceiro (3 parâmetros) inclui além dos parâmetros já mencionados o acerto casual (Oliveira, 2012; Pasquali, 2007). Existem outras classificações e formas de definir os modelos, como, por exemplo, em modelos cumulativos e de desdobramento. Os modelos cumulativos são classificados quanto ao tipo de itens que compõe o instrumento em dicotômicos ou politômicos, quanto à dimensionalidade (unidimensional ou multidimensional) e, ainda, quanto ao número de populações consideradas (uma ou mais) (Castro, Trentini & Riboldi, 2010).

Neste estudo foi utilizado o modelo de resposta gradual ou de Samejima que consiste em uma extensão do modelo de 2 parâmetros da TRI e aplica-se a itens com categorias ordinais semelhantes as utilizadas nesta pesquisa. Os modelos de TRI para itens politômicos obtêm mais informação do que os dicotômicos. Deste modo, a aplicação da TRI para itens com mais de duas categorias potencializa o acesso a informação, porém exige que a amostra de calibração do instrumento seja maior, pois há mais parâmetros a serem estimados (Andrade, Tavares & Valle, 2000; Embretson & Reise, 2000).

No modelo de resposta gradual (vide Equação 1) o parâmetro a (inclinação da curva ou discriminação) é o mesmo para todas as curvas características dos itens das categorias. As categorias de cada item, por seu caráter ordinal, deverão ter valores de dificuldades graduais $b1 < b2 < b3 \dots < bi$. Deste modo, para um item com cinco possibilidades de resposta obtêm-se 4 categorias, nas quais a dificuldade de uma determinada opção será sempre maior que sua antecessora, e, concomitantemente ter-se-á 4 valores diferentes de b (dificuldade) (Embretson & Reise, 2000).

$$P_{ix} * (\theta) = \frac{\exp[\alpha_i(\theta - \beta_{ij})]}{1 + \exp[\alpha_i(\theta - \beta_{ij})]}$$

Equação 1. Modelo de resposta gradual (Embretson & Reise, 2000).

$P_{ix}*(\theta)$ = é a curva característica operacional que expressa a probabilidade de um indivíduo x com habilidade θ responder corretamente (endossar) o item.

α = é o parâmetro inclinação (discriminação).

β = é o parâmetro dificuldade.

A estimativa dos parâmetros constitui-se do ponto de vista estatístico em dois passos ou etapas. Primeiro, a estimativa das curvas características operacionais e segundo a estimativa das curvas de categorias de resposta. No cálculo das curvas operacionais cada par de categorias é tratado dicotomicamente estimando qual a probabilidade da resposta do sujeito cair dentro desta ou daquela categoria. Após o cálculo das curvas operacionais, as curvas dos níveis de resposta são obtidas pelo cálculo das diferenças entre as curvas características operacionais (vide Equação 2) (Pasquali, 2007).

$$P_{ix}(\theta) = P_{ix}^*(\theta) - P_{i(x+1)}^*(\theta), \quad \text{com, } x=0 \dots 4$$

Equação 2. Curvas dos níveis de resposta (Embretson & Reise, 2000).

Por definição, a probabilidade de responder a categoria mais baixa ou acima desta é $P_{i0}(\theta) = 1$ e a probabilidade de responder a categoria mais alta ou acima desta é $P_{i5}(\theta) = 0$. Deste modo a probabilidade de responder a cada uma das cinco categorias será:

$$\begin{aligned} P_{i0} &= 1 - P_{i1}^*(\theta) \\ P_{i1}(\theta) &= P_{i1}^*(\theta) - P_{i2}^*(\theta) \\ P_{i2}(\theta) &= P_{i2}^*(\theta) - P_{i3}^*(\theta) \\ P_{i3}(\theta) &= P_{i3}^*(\theta) - P_{i4}^*(\theta) \\ P_{i4}(\theta) &= P_{i4}^*(\theta) - 0 \end{aligned}$$

A TRI complementa e amplia a TCT, modificando o paradigma da elaboração de instrumentos e seus parâmetros psicométricos, bem como fazendo surgir novas perspectivas no campo dos testes psicológicos e educacionais. Contribui, por meio de seus modelos, para a resolução de problemas dificilmente solucionáveis com os recursos estatísticos que a antecederam.

A Testagem Universal respeita todos os pressupostos teóricos instituídos no tocante aos testes psicológicos preocupando-se de forma mais específica com o formato dos testes e sua influência na realização do mesmo. Foca as adaptações e os recursos de tecnologias assistivas enquanto proporcionadores de acessibilidade analisando sua repercussão na qualidade dos instrumentos (Ketterlin-Geller, 2005; Thompson, Johnstone & Thurlow, 2002). Os princípios do Desenho Universal (Story; Mueller & Mace, 1998), na sua aplicação à testagem, foram sistematizados em sete princípios da Testagem Universal por Thompson, Johnstone e Thurlow (2002) – um teste será considerado de Testagem Universal se atender a estes sete princípios total ou parcialmente:

1. *População de avaliação ampla e inclusiva*

Os testes psicológicos são desenvolvidos ou adaptados objetivando uma população ampla, geralmente pessoas com e sem deficiências. Atende a uma demanda crescente por maior inclusão da diversidade humana.

2. *Definição precisa do construto*

Os testes psicológicos são desenvolvidos ou adaptados visando à máxima definição dos construtos a serem medidos evitando que possam ser afetados por outros construtos alheios ao que se pretende

medir. Para tanto se faz imprescindível que os instrumentos empreguem teorias fortemente embasadas.

3. *Itens acessíveis e não tendenciosos*

Os testes psicológicos são desenvolvidos ou adaptados para que indivíduos pertencentes a diferentes grupos (deficientes, estrangeiros, sem deficiência, entre outros) com a mesma habilidade no construto avaliado tenham a mesma chance de acertar os itens do teste psicológico (em testes de desempenho) ou a apresentarem certos padrões de respostas (em testes de personalidade, interesse, entre outros), e que todos estes grupos compreendam as instruções dos procedimentos envolvidos na testagem.

4. *Testes flexíveis a acomodações*

Os testes psicológicos são desenvolvidos ou adaptados de tal forma que apliquem os princípios do Desenho Universal para os tornarem flexíveis a acomodações, como, por exemplo, as tecnologias assistivas.

5. *Instruções e procedimentos simples, claros e intuitivos*

Os testes psicológicos são desenvolvidos ou adaptados evitando-se que instruções e procedimentos em linguagem complexa ou confusa prejudiquem a compreensão do que é questionado, evitando que o testando incorra em erros por não compreender o que lhe é pedido mesmo possuindo habilidade no construto medido.

6. *Leitura agradável e de máxima inteligibilidade*

Os testes psicológicos são desenvolvidos ou adaptados para reduzir a complexidade verbal e de organização textual dos itens e instruções preservando seu conteúdo essencial.

7. *Máxima legibilidade*

Os testes psicológicos são desenvolvidos ou adaptados de forma que o texto, os gráficos, tabelas e ilustrações, bem como o formato das respostas possam ser percebidos e decifrados com facilidade.

Ressalta-se que no desenvolvimento de testes psicológicos de Testagem Universal a serem utilizados pelas pessoas com deficiência faz-se imprescindível a utilização de tecnologias assistivas as quais criam um lócus de acessibilidade. A tecnologia assistiva, quando bem empregada, viabiliza a realização dos instrumentos por pessoas com deficiência de forma equitativa às pessoas sem deficiência. Porém, tais recursos exigem que o instrumento seja planejado desde o princípio com vistas a garantir a exequibilidade técnica do mesmo. Os testes informatizados se tornam grandes aliados na construção de testes de Testagem Universal por permitirem que os instrumentos possuam maior flexibilidade agregando várias possibilidades de usos e recursos

adicionais para as pessoas com deficiência (Thurlow, Lazarus, Albus & Hodgson, 2010).

Método

O presente estudo enquadra-se no âmbito dos estudos de desenvolvimento de instrumentos em Avaliação Psicológica. O objetivo desta pesquisa consistiu em calibrar os itens de um banco de itens de Personalidade e Testagem Universal. Este compõe um estudo mais amplo de elaboração de um banco de itens (BI) de Personalidade e Testagem Universal para Testagem Adaptativa Computadorizada (*Computerized Adaptive Testing – CAT*).

Participantes

A amostra compôs-se de indivíduos maiores de idade de ambos os sexos, com e sem deficiência (n total = 1.382; n dos Itens Âncoras= 638 – vide amostra por caderno na Tabela 1), sendo destes 76% do sexo feminino. Os participantes foram de 24 estados brasileiros e do Distrito Federal. Os estados com o maior número de participantes foram Santa Catarina (37%) e Paraíba (24%). Os estados com menor número de participantes foram Amapá, Acre e Amazonas cada qual com 1 participante somente; Rondônia e Roraima foram os únicos estados sem participantes. A média de idade foi de 33 anos (DP=11), com valor mínimo de 18 anos e máximo de 74 anos. Quanto à escolaridade apenas 13% possui até o Ensino Médio Completo, 70% da amostra possui Ensino Superior Incompleto e Completo ou Especialização, enquanto 17% possui mestrado ou doutorado.

Tabela 1.

*Amostra por Caderno*⁴.

Caderno	Amostra
Caderno 1	236
Caderno 2	244
Caderno 3	235
Caderno 4	235
Caderno 5	270

⁴ A soma da amostra por caderno excede a amostra total, pois os itens seguiram o modelo de Blocos Incompletos Balanceados (BIB).

Caderno	Amostra
Caderno 6	316
Caderno 7	236
Caderno 8	256
Caderno 9	249
Caderno 10	220
Caderno 11	291
Caderno 12	231
Caderno 13	227
Caderno 14	239
Caderno 15	281

Instrumento

Construção do Instrumento

Os itens são de autorrelato com cinco opções de respostas (escala tipo Likert) e foram desenvolvidos segundo os Cinco Grandes Fatores e os pressupostos da Testagem Universal, totalizando 525 itens. A amostra almejada tratou-se de indivíduos maiores de idade sem e com deficiência (visual e motora). Organizaram-se, os itens, em uma Bateria de Personalidade e Testagem Universal (BPTU), na qual foram acrescentados 40 itens com parâmetros de qualidade já avaliada em outros estudos do Laboratório de Pesquisa em Avaliação Psicológica (LPAP/UFSC) – itens âncoras. Tal bateria foi alocada na Plataforma *Concerto* (*Open-source Online Rbased Adaptive Testing Platform*), a qual é uma plataforma para criação de instrumentos de código aberto que permite que os usuários criem avaliações *on-line*. Acresce-se que tais itens passaram por todas as etapas de construção de itens já sedimentadas na literatura científica da área (Pasquali, 2010), como a análise de juízes e ainda por avaliação dos critérios oriundos da Testagem Universal (Oliveira & Nunes, 2015; Oliveira; Nuernberg & Nunes, 2013).

Os itens foram construídos usando os métodos citados no parágrafo anterior e a pré-testagem destes foi realizada por meio do método de Blocos Incompletos Balanceados (BIB). O BIB é uma solução estatística que permite a coleta de um grande número de informações de uma quantidade extensa de itens, sendo que um mesmo testando responde a apenas uma fração destes (Fisher & Yates, 1963; Andrade, Tavares & Valle, 2000; Beckman, 2001; Moreira Junior,

2011). No modelo de BIB adotado os itens foram separados em 35 cadernos (c), 15 blocos (b), com um número de 3 (k) blocos por caderno e um número de 7 (r) repetições de cada bloco. Ainda, o fator de utilização (*fu*) foi de 0,2, ou seja, 20% dos testandos foi submetido a cada bloco (Fisher & Yates, 1963; Andrade, Tavares & Valle, 2000; Beckman, 2001; Moreira Junior, 2011). Assim, os 525 itens foram distribuídos no banco em 35 itens por bloco e 105 itens em cada caderno ($525/15=35$; $3*35=105$). Os itens passaram pela análise fatorial e aqueles que apresentaram valores de cargas fatoriais satisfatórias e que carregaram nos cinco fatores de forma teoricamente defensável persistiram no banco. Assim a BPTU utilizada neste artigo de calibração será de 392 itens novos somados aos 40 itens âncoras.

Procedimentos

As coletas foram realizadas *on-line*, o *link*⁵ para tal coleta foi amplamente divulgado por *e-mails* e redes sociais (institucionais e individuais). Ainda, contatou-se a Associação Catarinense para Integração do Cego (ACIC) e divulgou-se o *link* em listas de *e-mails* desta associação. Complementa-se que este estudo foi submetido ao comitê de ética em pesquisas com seres humanos (CEPSH/UFSC), por meio da Plataforma Brasil/DATASUS, estando aprovado e com número CAAE 51788015.6.0000.012.

Análise de Dados

Foi realizada a calibração dos itens em TRI de dois parâmetros para itens politômicos no Modelo de Resposta Gradual de Samejima (*Samejima's Graded Response Model* – SGRM). Utilizou-se como método para a estimativa do *theta* a estimativa da máxima verossimilhança (*Maximum Likelihood Estimation* – MLE). Os programas utilizados nas análises de dados foram o STATA 14 (*Data Analysis and Statistical Software*), o Xcalibre 4.1 (*Item Response Theory Calibration Software*) e o Microsoft Excel 2010 – destaca-se que a calibração dos itens foi feita no Xcalibre.

Resultados da calibração

⁵ Link para acessar o instrumento:
<http://concerto4.labape.com.br/?wid=4&tid=1>

O processo de estimativa dos parâmetros dos itens em TRI, calibração, objetiva gerar modelos que relacionem as propriedades dos itens e o nível de habilidade dos respondentes (Nunes & Primi, 2009; Andrade, Tavares & Valle, 2000; Embretson & Reise, 2000). Os itens tiveram seus parâmetros estimados, separadamente, em cada um dos cinco fatores: Abertura, Extroversão, Neuroticismo, Realização e Socialização, os quais passaram por análise fatorial anteriormente a este estudo, mantendo, assim, a assunção de dimensionalidade apropriada ao modelo de TRI escolhido (SGRM) (Embretson & Reise, 2000). Na sequência, serão relatados os resultados obtidos nas análises para cada um dos fatores.

Abertura

O número de itens para o fator Abertura foi de 86, sendo dentre estes 78 itens novos e 8 itens que já possuíam os seus parâmetros conhecidos (itens âncora). Contudo, no processo de calibração para os cinco fatores que iremos apresentar nesta seção deste artigo, a estimativa foi realizada para todos os itens, ou seja, os parâmetros dos itens âncora não foram fixados. No referido fator, 81 itens foram calibrados, mas o item i215c7f1fc2 foi removido por não possuir respostas em uma das categorias⁶ “Descreve-me muito mal” (vide Figura 1). Outros 4 itens foram removidos i11c1f1fc, i71c3f1fc3, i73c3f1fc2 e i421c13f1fc1 por possuírem baixa frequência em uma das categorias aliada à pouca variabilidade entre as categorias (vide exemplo do item i11c1f1fc na Figura 2). Acresce-se que o item i11c1f1fc é um item intenso/ostensivo, assim seria necessária uma amostra específica, por exemplo, clínica, ou a ampliação desta amostra para alcançarem-se uma distribuição suficiente para calibração; tal fato se repete para os itens i71c3f1fc3 e i421c13f1fc1 nos quais as temáticas abordadas variam entre abertura à opinião dos outros e a busca por novidades. Já o item i73c3f1fc2, que envolve aspectos relacionados ao gosto por assuntos científicos, pode ter sido influenciado pelo alto nível de escolaridade da amostra (menor variabilidade) e também não apresentou bons parâmetros de calibração (Vide Figura 3).

⁶ Categorias utilizadas: 1. Descreve-me muito mal; 2. Descreve-me mal; 3. Descreve-me mais ou menos; 4. Descreve-me bem; 5. Descreve-me muito bem

i215c7f1fc2	Freq.	Percent	Cum.
2	2	0.85	0.85
3	12	5.11	5.96
4	90	38.30	44.26
5	131	55.74	100.00
Total	235	100.00	

Figura 1. Frequência das categorias do item i215c7f1fc2.

i11c1f1fc	Freq.	Percent	Cum.
1	67	28.39	28.39
2	88	37.29	65.68
3	60	25.42	91.10
4	16	6.78	97.88
5	5	2.12	100.00
Total	236	100.00	

Figura 2. Frequência das categorias do item i215c7f1fc2.

i73c3f1fc2	Freq.	Percent	Cum.
1	6	2.55	2.55
2	20	8.51	11.06
3	49	20.85	31.91
4	78	33.19	65.11
5	82	34.89	100.00
Total	235	100.00	

Figura 3. Frequência das categorias do item i73c3f1fc2.

Os valores de *theta* da amostra para o fator Abertura apresentaram valor máximo de 7,00 e mínimo de -4,027 (valor mínimo -7 e máximo +7) com média de -0,010. O Gráfico 1 mostra a distribuição das estimativas de *theta* das pessoas que responderam ao teste para todos os itens calibrados.

O Gráfico 2 exibe a Função de Informação do Teste (*Test Information Function* – TIF) para todos os itens calibrados. A TIF é uma representação gráfica de quanta informação o teste está fornecendo em cada nível de *theta*. A informação máxima ocorreu no valor de *theta* de -0,800. Observa-se que o teste apresentou-se mais informativo para os

níveis mais baixos de θ , ou seja, menor nível de abertura e pouca informação nos níveis mais altos.

O Gráfico 3 exibe a Função de Erro Padrão Condicional de Medida (CSEM). O CSEM é uma função invertida do TIF e estima a quantidade de erro na estimativa do θ para cada nível de θ . O CSEM mínimo foi 0,187 em $\theta = -0,800$ – maior informação do teste.

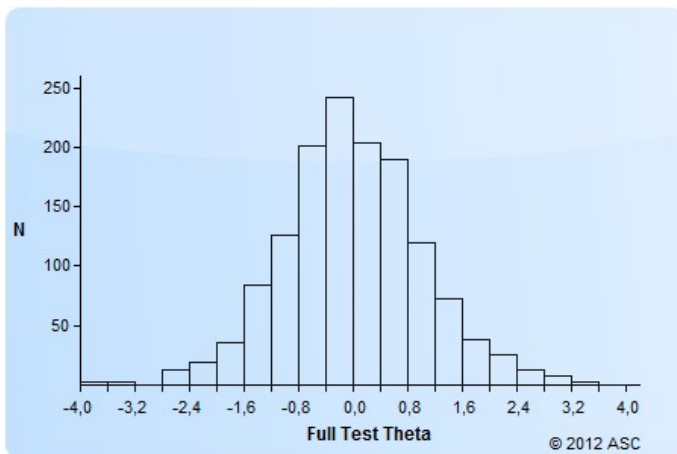


Gráfico 1. Distribuição dos θ s para o fator Abertura.

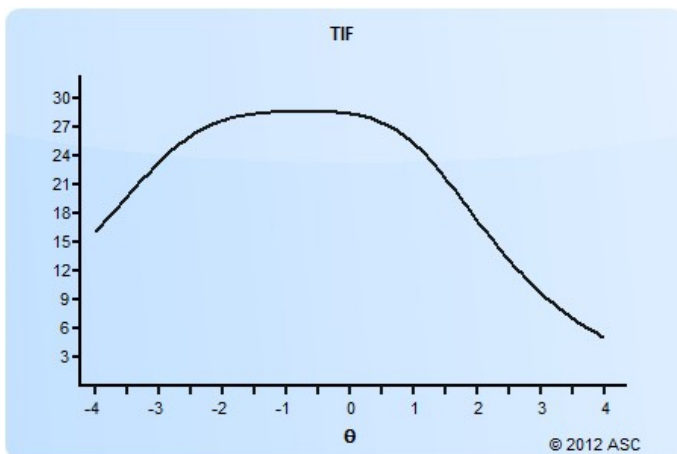


Gráfico 2. Função de Informação do Teste – fator Abertura.

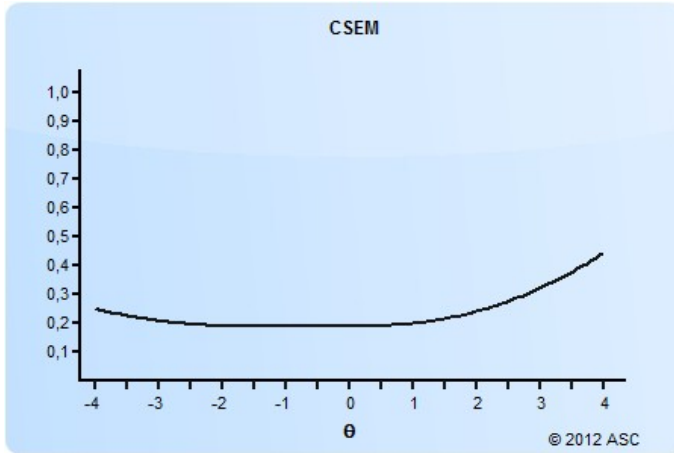


Gráfico 3. Erro Padrão Condicional de Medida (CSEM) – fator Abertura.

A Tabela 2 apresenta os parâmetros dos itens que foram calibrados. O a é o parâmetro discriminação e o b o parâmetro dificuldade do item. O parâmetro a reflete a inclinação das curvas características dos itens; nos itens politômicos o ponto de medida da inclinação é onde a probabilidade de resposta correta aquela categoria é 0,50 e a inclinação das curvas características está no máximo (Guyer & Thompson, 2011). Os itens com maiores valores de a são considerados os melhores, pois eles diferenciam de forma mais eficaz os testandos. Valores baixos de a são indicativos de revisão ou remoção de itens. O manual do XCalibre, programa utilizado na calibração dos itens, destaca como valores baixos $a < 0,50$ quando $D=1$. Tais itens foram excluídos da base (Vide Tabela 2).

Tabela 2.

Parâmetros dos itens para todos os itens calibrados – fator Abertura.

Seq.	Item ID	Item Mean	R	a	b
42	i2c1f1fc1	3,748	0,471	1,054	-3,76, -2,10, -0,37, 0,75
44	i4c1f1fc3	3,833	0,396	0,859	-4,21, -2,52, -0,74, 0,76
45	i5c1f1fc3	3,744	0,592	1,331	-2,90, -1,67, -0,40, 0,67
46	i6c1f1fc2	4,316	0,605	1,544	-3,73, -2,85, -1,62, 0,04
76	i36c2f1fc1	4,247	0,512	1,251	-4,01, -3,57, -1,65, 0,12
77	i37c2f1fc1	3,579	0,517	1,117	-4,79, -2,50, -0,21, 1,78
78	i38c2f1fc2	2,979	0,420	0,924	-2,42, -0,98, 0,84, 2,15
80	i40c2f1fc2	2,517	0,467	1,040	-1,28, 0,15, 1,20, 2,30
81	i41c2f1fc3	3,423	0,608	1,395	-2,35, -1,38, -0,07, 1,11

Seq.	Item ID	Item Mean	R	a	b
82	i42c2fl fc3	2,754	0,487	1,022	-1,55, -0,40, 0,72, 2,37
115	i75c3fl fc3	3,877	0,607	1,455	-3,97, -2,61, -0,64, 0,74
126	i86c3fl fc1	3,068	0,438	0,998	-2,59, -1,00, 0,72, 1,87
128	i88c3fl fc3	3,739	0,491	1,099	-3,98, -2,65, -0,55, 1,15
146	i106c4fl fc1	3,598	0,557	1,245	-4,02, -2,06, -0,19, 1,39
147	i107c4fl fc1	2,996	0,502	1,099	-2,66, -1,07, 0,89, 2,71
148	i108c4fl fc1	3,553	0,587	1,385	-2,97, -1,88, -0,25, 1,44
149	i109c4fl fc2	3,413	0,453	0,970	-2,78, -1,39, -0,13, 1,49
151	i111c4fl fc3	3,617	0,341	0,744	-3,69, -2,28, -0,50, 1,23
183	i143c5fl fc2	3,755	0,395	0,928	-3,12, -2,00, -0,84, 0,80
184	i144c5fl fc2	2,216	0,448	0,991	-0,33, 0,61, 1,50, 2,39
185	i145c5fl fc2	2,550	0,343	0,731	-1,42, -0,05, 1,68, 3,19
187	i147c5fl fc3	3,216	0,523	1,094	-2,07, -0,97, 0,08, 1,53
218	i178c6fl fc2	4,087	0,571	1,328	-3,69, -2,87, -1,05, 0,55
219	i179c6fl fc2	3,805	0,451	1,025	-3,49, -1,98, -0,60, 0,85
240	i200c6fl fc	2,487	0,279	0,645	-1,89, 0,15, 2,51, 4,51
248	i208c6fl fc	3,175	0,449	0,908	-2,73, -1,32, 0,62, 2,35
249	i209c6fl fc	2,672	0,499	1,074	-1,95, -0,28, 1,62, 3,60
250	i210c6fl fc1	2,006	0,185	0,473	0,31, 1,91, 3,52, 5,66
252	i212c7fl fc1	3,479	0,382	0,778	-4,52, -2,05, 0,05, 1,93
254	i214c7fl fc2	4,145	0,263	0,627	-6,64, -4,69, -2,37, 0,60
267	i227c7fl fc	3,624	0,415	0,955	-4,26, -2,67, -0,34, 2,01
285	i245c7fl fc1	1,373	0,028	0,402	3,36, 5,64, 7,87, 9,65
286	i246c8fl fc1	3,311	0,318	0,777	-4,24, -1,89, 0,47, 2,26
287	i247c8fl fc1	4,059	0,602	1,399	-4,55, -2,68, -1,13, 0,60
288	i248c8fl fc2	4,394	0,369	0,863	-6,73, -4,62, -2,27, -0,37
301	i261c8fl fc	3,839	0,309	0,739	-5,09, -3,02, -0,96, 0,97
303	i263c8fl fc	3,659	0,351	0,767	-4,73, -2,52, -0,54, 1,56
322	i282c9fl fc1	3,411	0,263	0,648	-3,70, -1,99, -0,13, 2,09
323	i283c9fl fc2	3,474	0,755	1,917	-1,90, -1,04, -0,04, 1,01
324	i284c9fl fc2	3,665	0,717	1,808	-2,17, -1,44, -0,33, 0,99
325	i285c9fl fc3	3,331	0,459	1,040	-1,90, -0,86, -0,08, 1,17
326	i286c9fl fc3	3,714	0,566	1,298	-2,54, -1,79, -0,51, 0,98
330	i290c9fl fc	3,414	0,394	0,826	-2,98, -2,00, 0,23, 1,79
331	i291c9fl fc	3,251	0,367	0,791	-4,12, -2,44, 0,84, 3,43
355	i315c9fl fc1	3,786	0,308	0,749	-3,25, -2,68, -0,90, 0,88
357	i317c10fl fc1	4,174	0,373	0,897	-4,37, -3,18, -1,65, -0,05
358	i318c10fl fc1	3,740	0,589	1,446	-2,77, -1,56, -0,56, 0,66
361	i321c10fl fc3	3,274	0,602	1,365	-1,87, -1,03, 0,12, 1,11
362	i322c10fl fc3	2,706	0,336	0,753	-1,71, -0,30, 1,07, 2,92
392	i352c11fl fc1	3,784	0,512	1,177	-3,36, -2,03, -0,52, 0,91
393	i353c11fl fc2	3,652	0,607	1,369	-2,33, -1,24, -0,31, 0,78
394	i354c11fl fc2	3,621	0,424	0,928	-2,92, -1,63, -0,56, 1,17
395	i355c11fl fc2	3,876	0,457	1,068	-2,98, -2,14, -0,60, 0,47
396	i356c11fl fc3	3,928	0,486	1,108	-2,62, -1,87, -0,79, 0,28

Seq.	Item ID	Item Mean	R	a	b
425	i385c11flfc1	3,804	0,136	0,428	-4,81, -3,38, -1,44, 0,29
426	i386c12flfc1	4,374	0,412	0,988	-4,55, -3,15, -2,04, -0,58
427	i387c12flfc1	3,667	0,488	1,043	-3,96, -2,26, -0,36, 1,15
428	i388c12flfc2	4,074	0,491	1,102	-4,42, -3,16, -1,13, 0,30
429	i389c12flfc2	3,416	0,507	1,142	-2,72, -1,54, -0,01, 1,34
430	i390c12flfc3	2,365	0,468	0,982	-0,57, 0,36, 0,94, 2,40
432	i392c12flfc3	3,061	0,491	1,046	-2,02, -0,87, 0,38, 1,84
463	i423c13flfc1	3,668	0,294	0,703	-3,28, -2,08, -0,56, 0,71
474	i434c13flfc	4,828	0,179	0,872	-6,50, -5,23, -4,07, -2,53
497	i457c14flfc1	3,008	0,368	0,804	-2,31, -0,76, 0,84, 2,62
499	i459c14flfc2	2,517	0,355	0,791	-1,10, 0,43, 1,47, 2,95
500	i460c14flfc2	3,874	0,373	0,953	-3,65, -2,47, -0,76, 0,87
501	i461c14flfc3	3,483	0,220	0,601	-3,61, -2,26, -0,32, 2,16
502	i462c14flfc3	3,181	0,498	1,162	-2,27, -0,88, 0,67, 1,65
529	i489c14flfc1	2,945	0,407	0,937	-1,28, -0,28, -0,61, 1,66
531	i491c15flfc1	3,057	0,562	1,262	-1,99, -0,78, 0,54, 1,80
533	i493c15flfc2	3,767	0,295	0,750	-4,26, -2,55, -0,68, 0,81
534	i494c15flfc2	3,014	0,604	1,394	-1,95, -0,79, 0,75, 1,72
547	i507c15flfc	3,662	0,171	0,493	-6,27, -3,55, -0,61, 1,84

Já o parâmetro b , dificuldade do item, no modelo adotado SGRM é o parâmetro de localização do item. Assim, um conjunto comum de parâmetros de localização de limiares de categoria é estimado para um dado domínio ou categoria. O parâmetro b indexa a distância em θ (θ) de um dado item a partir do centro do conjunto comum dos parâmetros dos limiares. Portanto, o parâmetro b está relacionado com a média do item (Guyer & Thompson, 2011). No Gráfico 4 são apresentados os valores mínimos e máximos dos limiares de categoria (*thresholds*) dos bs para cada um dos itens. Observa-se maior amplitude da categorias de bs negativos abarcando valores extremos invertidos de dificuldade em Abertura, bem como muitos itens próximos ao valor 0 de dificuldade e um número menor de categorias de itens intensos afirmativos; pode-se inferir que esta seja uma consequência de alguns itens desta natureza não terem sido calibrados. Ainda, tal fato pode dever-se ao caráter da amostra – população em geral – ao passo que tais itens possuem características medidas em populações específicas, sobretudo clínicas. Sugere-se, em novos estudos, a elaboração de mais itens desta natureza, a ampliação da coleta em cada um dos cadernos objetivando a calibração dos itens mais intensos, bem como a busca por populações específicas.

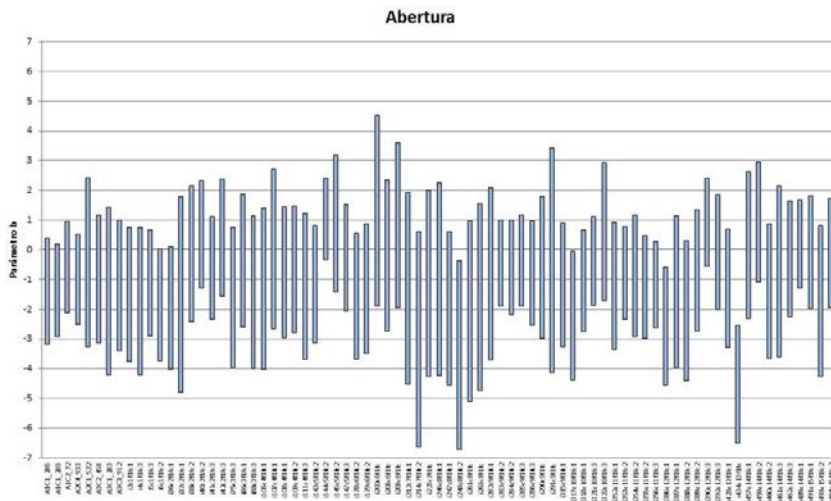


Gráfico 4. Parâmetros *bs*: Limiares de categorias máximos e mínimos dos itens – fator Abertura.

Extroversão

O número de itens para o fator Extroversão foi de 75, sendo dentre estes 67 itens novos e 8 itens que já possuíam os seus parâmetros conhecidos. No referido fator, 71 itens foram calibrados. O item *i161c5f3fc3*, que versa sobre os níveis mais altos de extroversão em atividades sociais, foi removido por não possuir respostas em uma das categorias “Descreve-me muito bem”. Outros 3 itens que eram mais intensos/ostensivos e cujas temáticas versavam sobre níveis elevados de extroversão em atividades de liderança e sociais (*i83c3f3fc2*, *i235c7f3fc* e *i400c12f3fc1*) foram removidos por possuírem baixo valor em uma das categorias aliados à baixa variabilidade entre as categorias. Deste modo, recomenda-se a ampliação da amostra com vistas ao alcance da calibração.

Os valores de *theta* da amostra para o fator Extroversão apresentaram valor máximo de 7 e mínimo de -7 (valor mínimo -7 e máximo +7) com média de 0,001. O Gráfico 5 mostra a distribuição das estimativas de *theta* das pessoas que responderam ao teste para todos os itens calibrados.

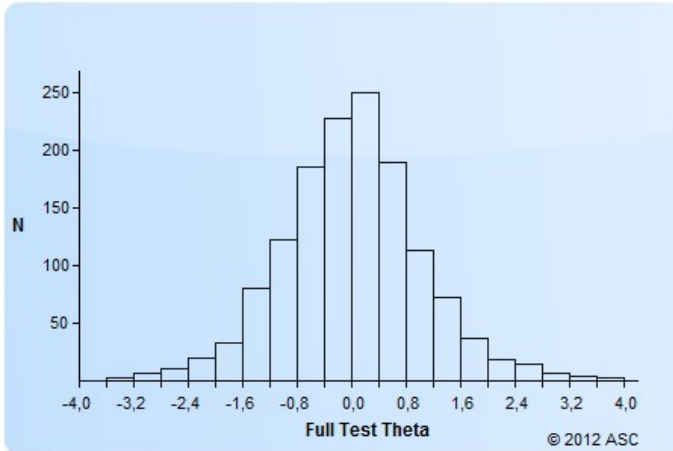


Gráfico 5. Distribuição dos *thetas* para o fator Extroversão.

O Gráfico 6 apresenta a TIF para todos os itens calibrados no fator Extroversão. A informação máxima foi de 35.287 em um valor de *theta* de -0,350. O gráfico aponta para uma extensão considerável de *theta*, na qual há bastante informação com predominância dos valores negativos (próximos de $\theta = -2$).

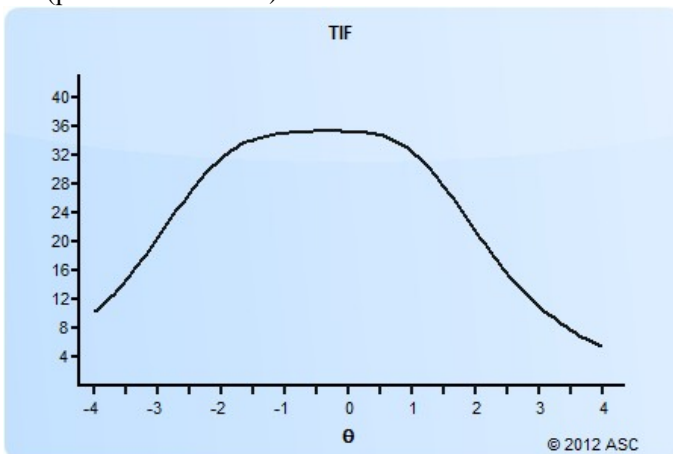


Gráfico 6. Função de Informação do Teste.

O Gráfico 7 expõe o CSEM. No fator Extroversão o CSEM mínimo foi de 0,168 no *theta* de -0,350. Os maiores erros de medida localizam-se nos valores elevados e afirmativos de *theta*.

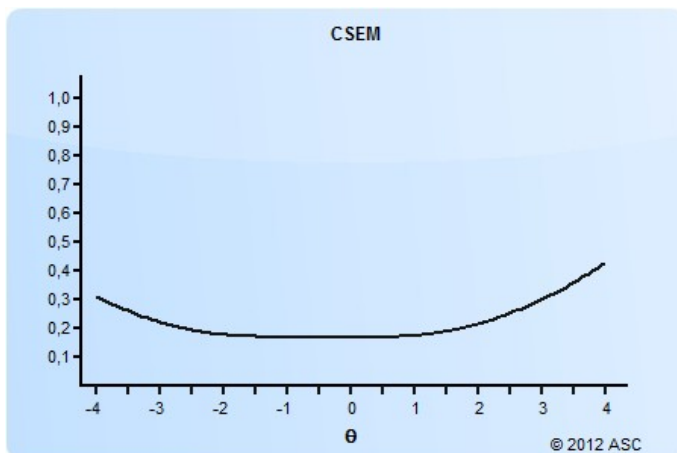


Gráfico 7. Erro Padrão Condicional de Medida (CSEM) – fator Extroversão.

Na Tabela 3 (Parâmetros dos Itens Calibrados) são apresentados os valores de a para o fator Extroversão. O item i226c7f3fc1 ($a = 0,258$) $a < 0,5$ foi retirado do BI.

Tabela 3.

Parâmetros dos itens para todos os itens calibrados – fator Extroversão.

Seq.	Item ID	Item Mean	R	a	b
53	i13c1f3fc1	3,894	0,467	1,106	-4,78, -2,82, -0,79, 0,84
54	i14c1f3fc1	3,591	0,524	1,265	-3,16, -2,12, -0,32, 1,56
56	i16c1f3fc2	3,640	0,628	1,679	-2,62, -1,56, -0,30, 0,95
57	i17c1f3fc2	2,391	0,565	1,219	-1,10, 0,24, 1,52, 2,77
59	i19c1f3fc3	3,421	0,551	1,253	-2,92, -1,32, 0,00, 1,38
61	i21c1f3fc	2,315	0,368	0,804	-1,29, 0,49, 2,16, 4,08
91	i51c2f3fc1	3,477	0,540	1,247	-2,36, -1,25, -0,22, 1,17
93	i53c2f3fc2	3,317	0,348	0,779	-4,05, -1,89, 0,24, 2,74
95	i55c2f3fc3	3,896	0,695	1,716	-2,79, -1,57, -0,52, 0,38
120	i80c3f3fc1	3,803	0,472	1,044	-4,07, -2,24, -0,58, 0,77
122	i82c3f3fc2	4,132	0,399	0,938	-4,96, -3,50, -1,48, 0,39
124	i84c3f3fc3	2,996	0,290	0,699	-2,83, -1,17, 1,09, 2,96
127	i87c3f3fc3	3,200	0,654	1,507	-1,50, -0,83, 0,27, 1,12
160	i120c4f3fc1	3,282	0,413	0,901	-4,25, -2,10, 0,68, 2,44
164	i124c4f3fc2	3,196	0,350	0,770	-3,18, -1,57, 0,54, 2,33
165	i125c4f3fc3	3,197	0,457	0,981	-2,66, -1,10, 0,52, 1,55
166	i126c4f3fc3	3,430	0,402	0,851	-3,32, -2,15, 0,13, 1,66
195	i155c5f3fc1	3,356	0,393	0,927	-4,45, -3,31, 0,49, 3,42

Seq.	Item ID	Item Mean	R	a	b
196	i156c5f3fc1	3,420	0,586	1,407	-2,18, -1,27, -0,05, 1,10
197	i157c5f3fc2	3,910	0,778	2,237	-2,55, -1,82, -0,63, 0,52
198	i158c5f3fc2	3,737	0,682	1,733	-2,34, -1,60, -0,47, 0,73
200	i160c5f3fc3	3,071	0,371	0,897	-2,62, -0,97, 0,62, 2,27
231	i191c6f3fc1	3,140	0,445	1,002	-1,95, -1,08, 0,40, 1,79
234	i194c6f3fc3	2,850	0,486	1,038	-1,84, -0,62, 1,16, 2,17
235	i195c6f3fc3	2,783	0,355	0,791	-2,04, -0,32, 1,15, 2,78
236	i196c6f3fc3	3,460	0,712	1,864	-1,80, -1,00, -0,07, 0,89
266	i226c7f3fc1	3,137	-0,033	0,258	-6,48, -3,12, 1,53, 5,34
269	i229c7f3fc2	3,693	0,356	0,813	-4,49, -3,02, -0,70, 1,91
271	i231c7f3fc3	2,907	0,521	1,110	-1,91, -0,57, 0,88, 2,10
302	i262c8f3fc2	3,133	0,363	0,749	-3,28, -1,40, 0,85, 2,67
305	i265c8f3fc3	2,125	0,415	0,879	-0,61, 0,87, 2,24, 3,83
338	i298c9f3fc2	3,169	0,709	1,857	-1,54, -0,75, 0,51, 1,31
339	i299c9f3fc3	3,193	0,494	1,124	-2,76, -1,16, 0,68, 2,16
340	i300c9f3fc3	2,707	0,641	1,538	-1,41, -0,23, 1,32, 2,06
341	i301c9f3fc3	2,698	0,370	0,862	-1,89, -0,06, 1,51, 2,81
371	i331c10f3fc1	4,092	0,546	1,352	-4,06, -3,01, -1,11, 0,43
372	i332c10f3fc1	3,452	0,709	1,958	-2,19, -1,03, -0,10, 0,98
373	i333c10f3fc2	2,936	0,499	1,109	-1,54, -0,40, 0,45, 1,67
374	i334c10f3fc2	3,251	0,741	2,040	-1,74, -1,05, 0,22, 1,29
375	i335c10f3fc3	3,388	0,701	1,837	-1,81, -0,89, 0,00, 0,78
376	i336c10f3fc3	2,653	0,505	1,126	-1,20, -0,22, 0,99, 1,91
405	i365c11f3fc1	3,313	0,661	1,683	-1,78, -0,94, 0,12, 1,18
406	i366c11f3fc1	3,605	0,680	1,811	-2,42, -1,46, -0,21, 1,07
407	i367c11f3fc2	3,796	0,649	1,632	-3,22, -2,04, -0,43, 0,95
408	i368c11f3fc2	3,436	0,495	1,179	-2,87, -1,72, 0,08, 1,62
411	i371c11f3fc3	3,775	0,518	1,213	-3,22, -2,11, -0,58, 1,04
443	i403c12f3fc2	3,873	0,708	1,884	-2,29, -1,49, -0,62, 0,25
445	i405c12f3fc3	3,635	0,583	1,347	-2,27, -1,67, -0,46, 0,79
477	i437c13f3fc1	3,419	0,635	1,511	-2,53, -1,28, 0,17, 1,36
478	i438c13f3fc2	3,484	0,632	1,456	-2,32, -1,28, 0,00, 1,15
480	i440c13f3fc3	2,584	0,410	0,941	-1,30, 0,05, 1,40, 2,56
481	i441c13f3fc3	3,088	0,424	0,919	-1,90, -0,82, 0,63, 1,66
495	i455c13f3fc	2,740	0,391	0,907	-2,02, -0,42, 1,29, 3,28
511	i471c14f3fc1	3,286	0,387	0,834	-3,53, -1,39, 0,31, 2,30
512	i472c14f3fc2	2,508	0,562	1,257	-0,97, 0,27, 1,25, 1,96
513	i473c14f3fc2	3,242	0,551	1,356	-1,84, -0,79, 0,30, 1,30
514	i474c14f3fc2	3,198	0,387	0,868	-2,51, -1,24, 0,45, 2,05
515	i475c14f3fc3	2,633	0,300	0,698	-1,56, -0,27, 1,54, 3,81
536	i496c15f3fc	2,262	0,388	0,933	-0,53, 0,43, 1,65, 2,89
548	i508c15f3fc2	3,452	0,555	1,304	-2,45, -1,48, 0,01, 1,20
549	i509c15f3fc3	2,762	0,603	1,518	-1,34, -0,32, 0,88, 1,72
550	i510c15f3fc3	3,964	0,318	0,830	-4,25, -2,88, -1,21, 0,48
551	i511c15f3fc3	2,761	0,698	1,818	-1,11, -0,31, 0,70, 1,53

Os limiares de categoria (*thresholds*) do parâmetro *b* para Extroversão são apresentados no Gráfico 8. Entre -2 e +2 encontram-se as maiores concentrações de itens ao passo que há um menor número de itens intensos e afirmativos do que itens intensos invertidos para o fator extroversão. Há, complementarmente, redundância de itens com valores próximos a 0 tanto com valores positivos quanto negativos (no intervalo $+1 \leq 0 \leq -1$). Caso objetive-se criar uma versão reduzida, em novos estudos, poder-se-ia retirar alguns itens deste intervalo.

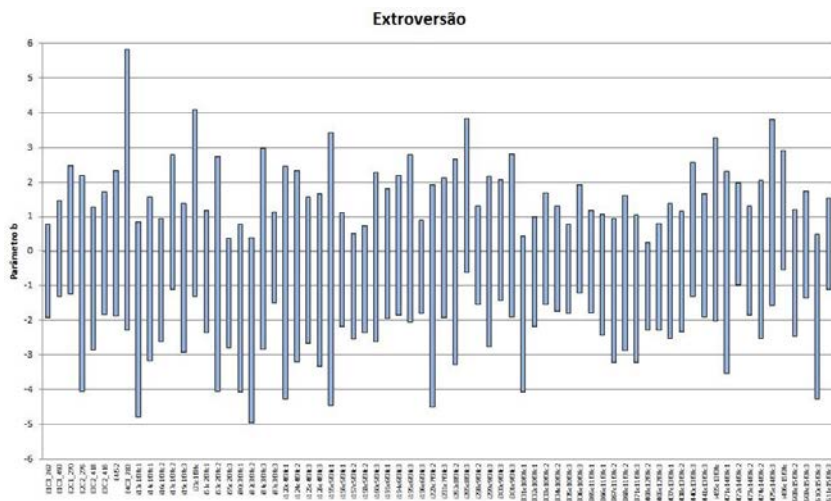


Gráfico 8. Parâmetros *bs*: Limiares de categorias máximos e mínimos dos itens – fator Extroversão.

Neuroticismo

O número de itens para o fator Neuroticismo foi de 97, sendo dentre estes 89 itens novos e 8 itens que já possuíam os seus parâmetros conhecidos. No referido fator, 94 itens foram calibrados. Os itens *i15c1f4fc* e *i50c2f4fc*, cuja temática predominante foi a assertividade apresentaram categorias com poucas respostas e pouca variabilidade entre as categorias, assim não calibraram; espera-se que com o aumento da amostra, tais itens, os quais são mais intensos/ostensivos, calibrem. O item *i192c6f4fc* apresentou pouca variabilidade nas respostas (vide figura 4) e também não calibrou. Tal item foi elaborado para ser de Neuroticismo invertido, abordou o inverso do humor depressivo, assim

sendo, como tal item não alcançou o objetivo desejado, ele foi retirado do BI.

i192c6f4fc	Freq.	Percent	Cum.
1	34	10.90	10.90
2	35	11.22	22.12
3	123	39.42	61.54
4	91	29.17	90.71
5	29	9.29	100.00
Total	312	100.00	

Figura 4. Frequência das categorias do item i192c6f4fc.

Os valores de *theta* para o fator Neuroticismo apresentaram valor máximo de 4,128 e mínimo de -7 (valor mínimo -7 e máximo +7) com média de -0,002. O Gráfico 9 mostra a distribuição das estimativas de *theta* das pessoas que responderam ao teste para todos os itens calibrados.

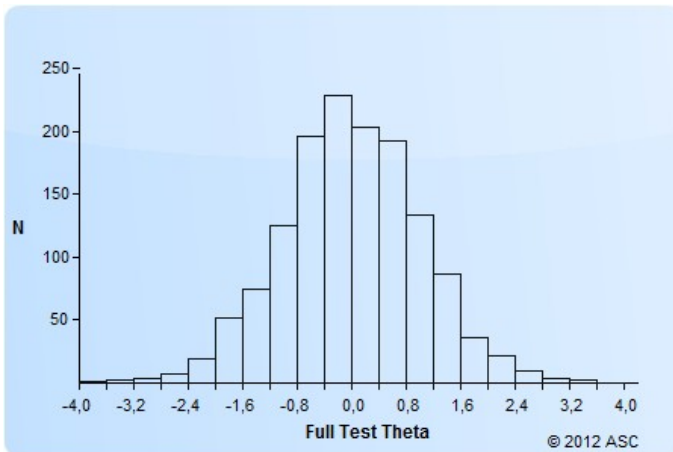


Gráfico 9. Distribuição dos *thetas* para o fator Neuroticismo.

No Gráfico 10 apresenta-se a Função de Informação do Teste para todos os itens calibrados. A informação máxima foi de 46.500 no *theta* = 0,850. Observa-se que os itens são mais informativos para níveis de *theta* médios e altos em Neuroticismo.

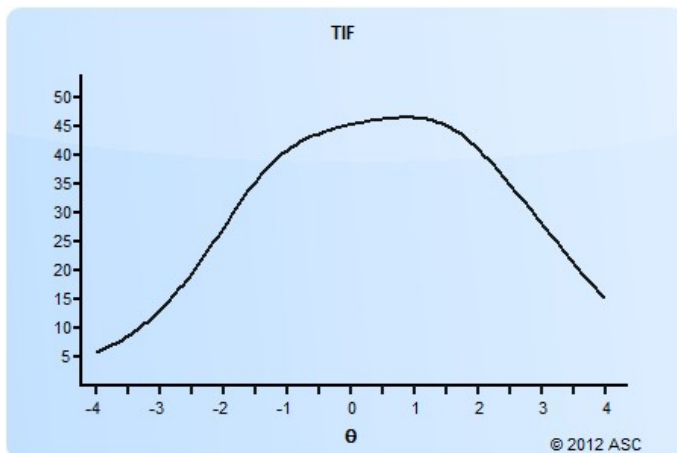


Gráfico 10. Função de Informação do Teste – fator Neuroticismo.

No gráfico 11 o Erro Padrão Condicional de Medida mínimo foi de 0,147 no θ de 0,850.

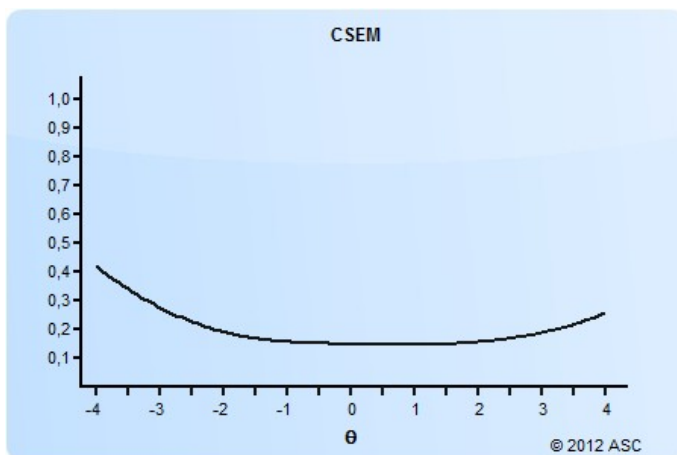


Gráfico 11. Erro Padrão Condicional de Medida (CSEM) – fator Neuroticismo.

Na Tabela 4 são apresentados os valores de a para o fator Neuroticismo. O item i233c7f4fc1 ($a = 0,477$) $a < 0,5$ foi retirado do BI.

Tabela 4.

Parâmetros dos itens para todos os itens calibrados – fator Neuroticismo.

Seq.	Item ID	Item Mean	R	a	b
25	N1C1_174	2,827	0,450	0,994	-1,96, -0,51, 1,08, 2,82
26	N1C3_446	2,860	0,416	0,885	-1,81, -0,31, 0,79, 2,24
27	N2C2_476	2,831	0,689	1,798	-1,29, -0,27, 0,80, 1,53
28	N2I77	2,797	0,740	2,083	-1,19, -0,18, 0,81, 1,48
29	N3C2_700	3,109	0,523	1,175	-2,22, -0,89, 0,62, 1,89
30	N3C3_793	2,782	0,530	1,257	-1,68, -0,32, 1,14, 2,22
31	N4C3_429	2,773	0,754	2,235	-1,11, -0,14, 0,80, 1,50
32	N4C3_425	2,633	0,563	1,351	-1,28, -0,03, 1,23, 2,19
63	i23c1f4fe1	2,800	0,641	1,428	-2,13, -0,40, 1,12, 2,08
65	i25c1f4fe2	3,140	0,341	0,729	-2,56, -0,93, 0,26, 2,06
66	i26c1f4fe2	3,102	0,688	1,623	-1,63, -0,74, 0,40, 1,38
67	i27c1f4fe3	2,923	0,670	1,508	-1,56, -0,59, 0,71, 1,63
68	i28c1f4fe3	2,855	0,507	1,133	-2,25, -0,64, 1,10, 2,49
94	i54c2f4fe	2,959	0,625	1,533	-1,61, -0,78, 0,74, 1,62
97	i57c2f4fe1	1,820	0,440	1,305	0,14, 1,16, 2,22, 2,79
98	i58c2f4fe1	2,086	0,500	1,180	-0,42, 0,76, 1,75, 2,72
99	i59c2f4fe2	2,741	0,717	1,802	-1,33, -0,42, 0,87, 1,74
100	i60c2f4fe2	2,463	0,725	1,856	-1,03, 0,15, 1,02, 1,72
101	i61c2f4fe2	2,918	0,636	1,483	-1,75, -0,42, 0,59, 1,60
103	i63c2f4fe3	3,096	0,462	1,014	-2,37, -1,17, 0,43, 2,14
131	i91c3f4fe1	2,236	0,600	1,318	-1,25, 0,53, 2,03, 3,21
133	i93c3f4fe2	2,983	0,237	0,612	-3,28, -0,58, 1,11, 2,32
134	i94c3f4fe2	2,308	0,599	1,403	-0,92, 0,17, 1,69, 2,83
135	i95c3f4fe3	2,575	0,669	1,582	-1,05, 0,06, 1,01, 1,57
136	i96c3f4fe3	2,761	0,391	1,023	-2,52, -0,39, 1,56, 2,73
137	i97c3f4fe3	2,671	0,556	1,238	-1,87, -0,24, 1,34, 2,23
156	i116c4f4fe	2,782	0,425	0,948	-2,43, -0,71, 1,81, 3,18
157	i117c4f4fe	2,815	0,304	0,693	-2,53, -0,40, 1,45, 3,15
159	i119c4f4fe	1,457	0,216	0,773	1,32, 3,02, 4,19, 5,93
169	i129c4f4fe1	2,655	0,457	0,995	-1,89, -0,24, 1,51, 3,19
170	i130c4f4fe2	2,902	0,217	0,598	-3,62, -0,80, 1,38, 4,17
171	i131c4f4fe2	2,409	0,623	1,518	-0,99, 0,22, 1,41, 2,51
172	i132c4f4fe3	1,585	0,353	0,944	1,00, 1,92, 2,94, 4,21
173	i133c4f4fe3	2,675	0,486	1,084	-0,94, 0,09, 0,77, 2,07
191	i151c5f4fe	3,259	0,474	1,041	-2,37, -1,00, 0,23, 1,50
194	i154c5f4fe	2,056	0,333	0,694	-0,05, 1,00, 2,94, 3,81
203	i163c5f4fe1	1,883	0,627	1,534	0,16, 0,85, 1,94, 2,41
204	i164c5f4fe2	2,726	0,631	1,412	-1,40, -0,30, 1,10, 1,91
205	i165c5f4fe2	1,182	0,233	0,856	2,56, 3,91, 5,60, 6,90
207	i167c5f4fe3	2,985	0,458	1,057	-1,95, -0,85, 1,01, 2,12
208	i168c5f4fe3	2,033	0,383	0,892	0,02, 1,08, 1,93, 3,43

Seq.	Item ID	Item Mean	R	a	b
237	i197c6f4fc1	1,862	0,339	0,816	0,02, 1,88, 3,04, 4,37
238	i198c6f4fc1	2,333	0,533	1,200	-1,28, 0,39, 2,29, 3,17
239	i199c6f4fc2	2,411	0,427	0,914	-1,15, 0,35, 1,92, 3,16
241	i201c6f4fc3	2,558	0,674	1,756	-0,81, 0,12, 1,15, 1,64
242	i202c6f4fc3	2,662	0,625	1,474	-1,49, -0,28, 1,44, 2,70
243	i203c6f4fc3	2,411	0,498	1,089	-0,75, 0,33, 1,46, 2,57
264	i224c7f4fc	2,184	0,342	0,816	-1,26, 0,78, 2,87, 4,99
272	i232c7f4fc1	2,825	0,336	0,801	-3,01, -0,82, 1,58, 3,88
273	i233c7f4fc1	2,392	0,166	0,477	-1,58, 0,61, 2,81, 5,18
274	i234c7f4fc1	1,498	0,620	1,751	0,55, 1,59, 2,41, 3,04
276	i236c7f4fc2	2,448	0,669	1,653	-1,29, 0,04, 1,71, 2,46
277	i237c7f4fc3	2,236	0,674	1,775	-0,61, 0,34, 1,42, 2,25
278	i238c7f4fc3	2,149	0,611	1,476	-0,64, 0,68, 1,63, 2,63
309	i269c8f4fc2	2,031	0,422	0,943	-1,09, 1,18, 3,54, 4,53
311	i271c8f4fc2	3,424	0,571	1,296	-2,82, -1,52, 0,10, 1,34
312	i272c8f4fc3	1,796	0,512	1,322	0,05, 1,38, 2,05, 3,11
313	i273c8f4fc3	2,070	0,613	1,583	-0,75, 0,65, 2,17, 3,03
343	i303c9f4fc1	2,133	0,608	1,447	-0,61, 0,56, 2,00, 3,10
344	i304c9f4fc2	1,746	0,460	1,168	0,63, 1,35, 2,04, 2,78
345	i305c9f4fc2	2,605	0,603	1,347	-1,56, -0,19, 1,44, 2,65
346	i306c9f4fc3	2,362	0,538	1,192	-1,19, 0,20, 2,06, 3,05
347	i307c9f4fc3	2,992	0,597	1,353	-1,67, -0,40, 0,64, 1,56
348	i308c9f4fc3	2,372	0,391	0,897	-2,08, 0,40, 2,89, 4,68
367	i327c10f4fc	3,479	0,333	0,786	-4,94, -2,53, 0,20, 2,52
377	i337c10f4fc1	2,125	0,605	1,484	-0,57, 0,63, 2,05, 2,95
378	i338c10f4fc1	2,186	0,598	1,496	-0,60, 0,57, 1,89, 2,88
380	i340c10f4fc2	2,553	0,556	1,288	-1,52, 0,10, 1,59, 2,76
382	i342c10f4fc3	2,301	0,686	1,784	-1,03, 0,50, 1,60, 3,11
383	i343c10f4fc3	1,777	0,328	0,839	0,72, 1,79, 2,46, 3,72
400	i360c11f4fc	3,790	0,210	0,554	-7,44, -4,72, -0,86, 1,94
412	i372c11f4fc1	2,886	0,257	0,660	-4,51, -1,39, 2,39, 4,40
414	i374c11f4fc2	3,208	0,431	0,997	-2,53, -1,02, 0,31, 1,82
416	i376c11f4fc2	3,362	0,733	1,952	-1,54, -0,73, 0,08, 0,83
417	i377c11f4fc3	2,979	0,612	1,510	-1,60, -0,44, 0,58, 1,47
418	i378c11f4fc3	2,647	0,553	1,295	-1,01, -0,03, 0,96, 1,80
447	i407c12f4fc1	1,922	0,493	1,211	-0,05, 1,11, 2,12, 2,66
448	i408c12f4fc1	2,186	0,471	1,125	-0,73, 0,46, 2,23, 3,46
449	i409c12f4fc2	2,693	0,430	1,050	-1,29, -0,22, 1,00, 2,40
451	i411c12f4fc3	2,563	0,603	1,483	-1,00, 0,03, 1,13, 2,02
452	i412c12f4fc3	3,494	0,524	1,210	-2,71, -1,28, -0,01, 1,02
453	i413c12f4fc3	2,571	0,587	1,478	-0,75, 0,11, 0,94, 1,79
457	i417c12f4fc	3,000	0,212	0,572	-3,38, -1,11, 1,17, 3,29
459	i419c12f4fc	2,407	0,442	1,030	-1,47, 0,21, 2,20, 3,16
483	i443c13f4fc1	1,911	0,410	1,004	-0,45, 1,45, 3,42, 4,39
484	i444c13f4fc1	2,075	0,521	1,231	-0,45, 0,78, 2,29, 3,28

Seq.	Item ID	Item Mean	R	a	b
485	i445c13f4fc2	2,885	0,545	1,309	-1,29, -0,32, 0,75, 1,76
486	i446c13f4fc2	2,583	0,597	1,444	-0,98, 0,08, 1,23, 1,97
522	i482c14f4fc3	1,633	0,545	1,368	0,25, 1,52, 2,92, 3,30
523	i483c14f4fc3	2,131	0,563	1,288	-0,54, 0,54, 1,75, 2,82
540	i500c15f4fc	3,272	0,359	0,815	-3,73, -1,59, 0,46, 2,42
554	i514c15f4fc2	3,349	0,664	1,697	-2,37, -1,24, 0,34, 1,27
555	i515c15f4fc2	3,464	0,282	0,776	-3,17, -1,86, -0,05, 1,36
557	i517c15f4fc3	3,114	0,699	1,866	-1,72, -0,58, 0,45, 1,28

Os limiares de categoria (*thresholds*) do parâmetro *b* para Neuroticismo são apresentados no Gráfico 12. Observa-se no referido gráfico, uma maior concentração de itens afirmativos para Neuroticismo. Portanto, sugere-se, em novos estudos com vistas à ampliação do BI, a elaboração de mais itens invertidos para Neuroticismo.

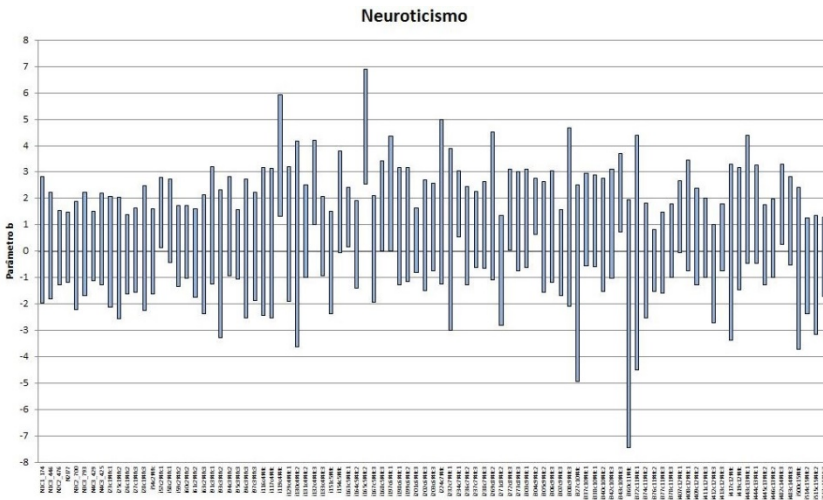


Gráfico 12. Parâmetros *bs*: Limiares de categorias máximos e mínimos dos itens – fator Neuroticismo.

Realização

O número de itens para o fator Realização foi de 97, sendo dentre estes 89 itens novos e 8 itens que já possuíam os seus parâmetros conhecidos. No referido fator, dentre os 97 itens que o compunham, 88 foram calibrados. Quatro itens apresentaram categorias que não foram endossadas (i176c6f5fc3, i344c10f5fc1, i350c10f5fc3 e i523c15f5fc3) e,

portanto, não calibraram. Já os itens i32c1f5fc2, i49c2f5fc, i264c8f5fc, i279c8f5fc3 e i436c13f5fc apresentaram categorias com frequência de respostas semelhantes – baixa variabilidade – e também não calibraram. Estes itens ou são intensos ou abordam temáticas relacionadas ao desempenho no trabalho/estudo, portanto, hipotetiza-se a influência da amostra com alto nível de escolarização e senso de realização nas respostas apresentadas a estes itens. Sugestiona-se, em novos estudos, a ampliação da amostra em níveis mais baixos de escolarização.

Os valores de *theta* da amostra para o fator Realização apresentaram valor máximo de 7 e mínimo de -7 (valor mínimo -7 e máximo +7) com média de 0,005. O Gráfico 13 mostra a distribuição das estimativas de *theta* das pessoas que responderam ao teste para todos os itens calibrados.

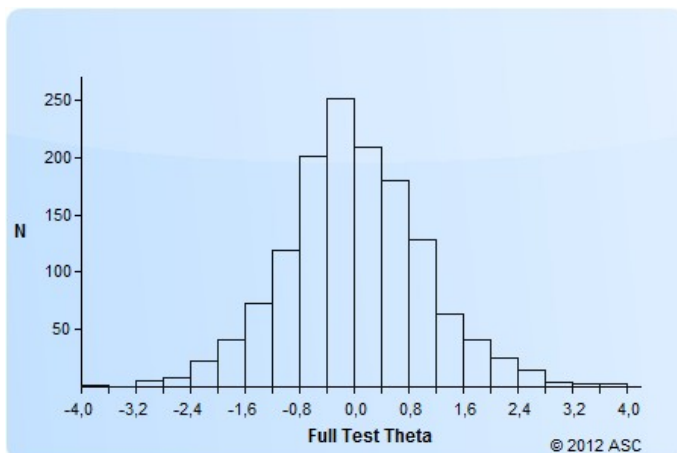


Gráfico 13. Distribuição dos *thetas* para o fator Realização.

O Gráfico 14 exibe a Função de Informação do Teste para todos os itens calibrados. No TIF observa-se que a informação máxima foi de 39.042 em *theta* = -0,90. Para o fator realização, tal gráfico demonstra que há mais informação nos níveis mais baixos de *theta*.

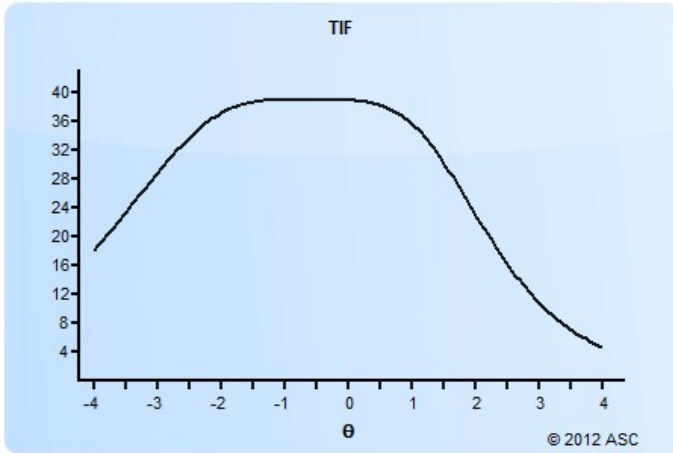


Gráfico 14. Função de Informação do Teste – fator Realização.

O CSEM mínimo é de 0,160 em $theta = -0,90$ conforme observa-se no Gráfico 15.

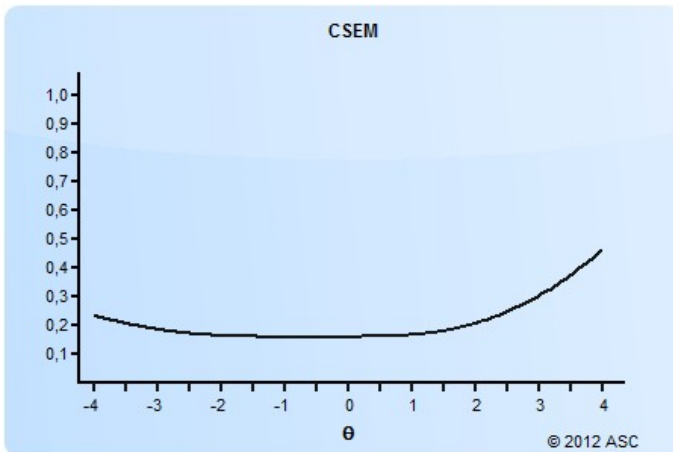


Gráfico 15. Erro Padrão Condicional de Medida (CSEM) – fator Realização.

Na Tabela 5 são apresentados os valores de a para o fator Realização. Os itens i481c14f5fc ($a=0,265$), i420c12f5fc3 ($a=0,494$) e i171c5f5fc2 ($a=0,296$) têm $a < 0,5$ e, portanto, foram retirados do BI.

Tabela 5.

Parâmetros dos itens para todos os itens calibrados – fator Realização.

Seq.	Item ID	Item Mean	R	a	b
69	i29c1f5fc1	4,078	0,659	1,713	-3,40, -2,34, -1,00, 0,36
70	i30c1f5fc1	2,754	0,442	1,015	-1,85, -0,40, 1,17, 2,45
71	i31c1f5fc1	3,517	0,469	1,098	-3,46, -2,03, -0,10, 1,62
73	i33c1f5fc2	3,145	0,382	0,891	-2,38, -0,99, 0,50, 1,50
74	i34c1f5fc3	2,894	0,673	1,537	-1,35, -0,54, 0,56, 1,67
75	i35c1f5fc3	3,295	0,496	1,221	-2,28, -1,18, 0,03, 1,41
88	i48c2f5fc	4,276	0,307	0,823	-6,98, -5,64, -2,46, 0,36
92	i52c2f5fc	3,111	0,200	0,605	-2,67, -1,52, 0,59, 2,40
104	i64c2f5fc1	3,967	0,586	1,444	-3,85, -2,90, -1,03, 0,98
105	i65c2f5fc1	3,634	0,579	1,381	-2,91, -1,38, -0,25, 0,81
106	i66c2f5fc2	3,921	0,301	0,764	-4,67, -3,06, -1,24, 0,77
107	i67c2f5fc2	3,638	0,476	1,075	-3,39, -2,31, -0,36, 1,42
108	i68c2f5fc2	2,246	0,552	1,257	-0,86, 0,59, 1,72, 2,43
109	i69c2f5fc3	3,872	0,541	1,259	-3,01, -2,11, -0,70, 0,63
110	i70c2f5fc3	3,630	0,613	1,522	-2,41, -1,72, -0,33, 1,08
132	i92c3f5fc	3,177	0,420	0,969	-3,46, -1,59, 0,91, 2,69
139	i99c3f5fc1	3,321	0,361	0,809	-3,49, -1,61, 0,37, 2,17
140	i100c3f5fc2	2,429	0,252	0,675	-1,27, 0,42, 1,94, 4,26
142	i102c3f5fc2	2,567	0,279	0,689	-2,08, -0,08, 2,20, 4,46
143	i103c3f5fc3	4,227	0,446	1,168	-4,40, -2,64, -1,64, 0,22
153	i113c4f5fc	2,872	0,294	0,726	-1,37, -0,31, 0,80, 1,88
155	i115c4f5fc	2,897	0,309	0,690	-2,21, -0,84, 1,16, 2,95
175	i135c4f5fc1	3,442	0,543	1,238	-2,24, -1,43, -0,12, 1,50
176	i136c4f5fc1	3,187	0,678	1,670	-2,27, -1,11, 0,47, 1,86
177	i137c4f5fc2	3,496	0,672	1,590	-2,23, -1,22, -0,03, 1,03
180	i140c4f5fc3	4,047	0,341	0,810	-5,29, -3,46, -1,42, 0,70
202	i162c5f5fc	3,796	0,589	1,441	-3,49, -2,51, -0,56, 1,11
209	i169c5f5fc1	3,704	0,421	0,916	-3,89, -2,11, -0,40, 0,95
210	i170c5f5fc1	3,996	0,575	1,404	-3,56, -3,38, -1,01, 0,87
211	i171c5f5fc2	3,563	0,052	0,296	-9,67, -5,00, -0,54, 3,35
212	i172c5f5fc2	3,202	0,523	1,212	-2,06, -0,87, 0,38, 1,17
213	i173c5f5fc2	3,341	0,689	1,745	-1,82, -0,97, 0,04, 1,01
214	i174c5f5fc3	3,675	0,552	1,315	-2,86, -2,05, -0,37, 1,15
215	i175c5f5fc3	3,937	0,573	1,409	-3,74, -3,35, -0,97, 1,06
230	i190c6f5fc	3,432	0,360	0,763	-4,38, -2,23, 0,17, 2,15
244	i204c6f5fc1	3,905	0,501	1,178	-3,33, -2,12, -0,70, 0,51
245	i205c6f5fc1	3,863	0,538	1,244	-2,77, -1,73, -0,74, 0,56
246	i206c6f5fc2	3,844	0,613	1,609	-2,21, -1,45, -0,61, 0,36
247	i207c6f5fc2	3,376	0,612	1,484	-1,82, -1,08, 0,15, 1,04
253	i213c7f5fc3	4,051	0,269	0,675	-5,41, -3,91, -1,74, 0,57
279	i239c7f5fc1	3,721	0,498	1,181	-3,11, -1,67, -0,44, 0,81
280	i240c7f5fc1	3,369	0,740	1,920	-1,78, -0,86, 0,00, 1,08
281	i241c7f5fc1	3,148	0,542	1,313	-2,01, -1,12, 0,46, 2,00

Seq.	Item ID	Item Mean	R	a	b
282	i242c7f5fe2	3,313	0,701	1,709	-1,85, -1,33, 0,31, 1,33
316	i276c8f5fe2	3,237	0,610	1,509	-1,80, -0,91, 0,11, 1,07
317	i277c8f5fe2	3,736	0,594	1,559	-2,33, -1,41, -0,55, 0,47
318	i278c8f5fe2	3,180	0,474	1,049	-2,39, -1,25, 0,39, 1,68
321	i281c9f5fe3	4,097	0,346	0,863	-5,81, -3,20, -1,40, 0,50
336	i296c9f5fe	3,629	0,467	1,001	-3,78, -2,47, 0,02, 1,45
337	i297c9f5fe	4,371	0,369	0,921	-6,22, -3,77, -2,31, -0,10
342	i302c9f5fe	4,389	0,523	1,295	-4,69, -3,52, -2,02, 0,10
350	i310c9f5fe1	3,375	0,450	1,022	-2,91, -1,47, 0,34, 1,71
351	i311c9f5fe2	3,418	0,480	1,077	-1,87, -1,08, -0,19, 1,19
353	i313c9f5fe3	3,702	0,339	0,775	-5,47, -3,07, -0,47, 1,81
354	i314c9f5fe3	3,801	0,591	1,476	-3,45, -2,33, -0,45, 1,15
385	i345c10f5fe1	3,693	0,651	1,605	-2,86, -1,56, -0,29, 1,10
386	i346c10f5fe1	3,509	0,720	1,836	-1,99, -1,28, 0,08, 1,01
387	i347c10f5fe2	3,182	0,506	1,174	-1,79, -0,85, 0,33, 1,43
388	i348c10f5fe2	2,753	0,413	0,949	-1,84, -0,26, 1,23, 2,62
389	i349c10f5fe3	3,355	0,352	0,788	-2,92, -1,47, 0,08, 1,78
391	i351c11f5fe3	3,873	0,561	1,284	-3,37, -2,12, -0,56, 0,68
419	i379c11f5fe1	3,400	0,612	1,378	-2,23, -1,26, 0,05, 1,53
421	i381c11f5fe2	3,879	0,536	1,262	-3,27, -2,06, -0,70, 0,66
422	i382c11f5fe2	3,117	0,394	0,847	-3,09, -1,26, 0,86, 2,49
423	i383c11f5fe2	3,443	0,684	1,697	-1,77, -0,84, -0,05, 0,79
455	i415c12f5fe1	2,693	0,722	1,769	-1,30, -0,27, 0,83, 2,04
456	i416c12f5fe2	3,489	0,486	1,119	-2,97, -1,60, -0,10, 1,31
460	i420c12f5fe3	3,169	0,170	0,494	-4,35, -2,13, 1,15, 3,06
465	i425c13f5fe	3,652	0,306	0,713	-4,19, -2,50, -0,61, 1,35
489	i449c13f5fe1	3,407	0,671	1,619	-2,15, -1,18, -0,01, 0,99
490	i450c13f5fe1	3,841	0,494	1,148	-4,65, -2,62, -0,98, 1,19
491	i451c13f5fe1	3,537	0,291	0,669	-5,68, -3,29, -0,22, 2,62
494	i454c13f5fe3	3,203	0,391	0,876	-3,07, -1,50, 0,40, 2,34
521	i481c14f5fe	2,600	0,012	0,265	-5,55, 0,00, 5,06, 8,53
559	i519c15f5fe1	3,785	0,602	1,456	-3,37, -2,46, -0,33, 0,96
560	i520c15f5fe1	3,393	0,567	1,436	-2,47, -1,36, 0,07, 1,45
561	i521c15f5fe2	3,580	0,610	1,499	-2,29, -1,57, -0,18, 1,03
562	i522c15f5fe2	3,607	0,567	1,306	-2,41, -1,44, -0,32, 0,97
564	i524c15f5fe3	3,879	0,539	1,259	-3,66, -2,32, -0,81, 0,92
565	i525c15f5fe3	3,636	0,602	1,543	-2,34, -1,55, -0,29, 0,95

Os limiares de categoria (*thresholds*) do parâmetro *b* para Realização são apresentados no Gráfico 16, no qual observa-se uma concentração de itens próximos ao valor 0. Sugere-se a ampliação do BI com a elaboração de itens ostensivos, bem como a diversificação da amostra em novos estudos.

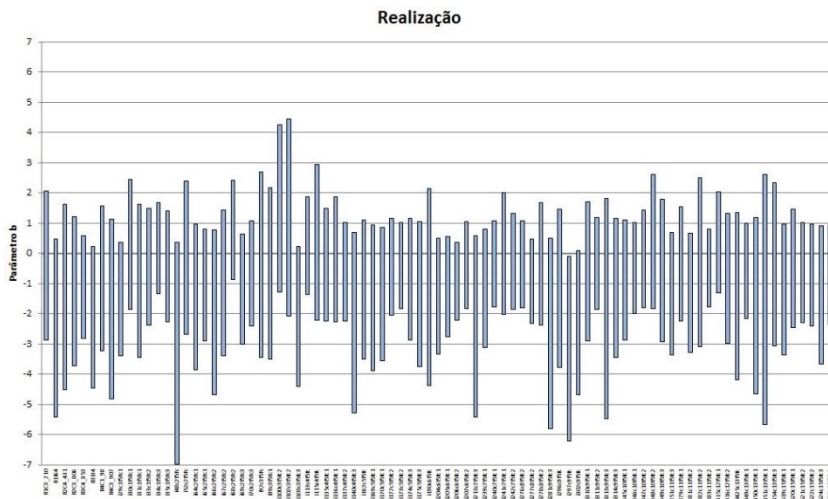


Gráfico 16. Parâmetros bs: Limiars de categorias máximos e mínimos dos itens – fator Realização.

Socialização

O número de itens para o fator Socialização foi de 77, sendo dentre estes 69 itens novos e 8 itens que já possuíam os seus parâmetros conhecidos. No referido fator, dentre os 77 itens que o compunham, 62 foram calibrados. Sete itens apresentaram categorias que não foram endossadas (i43c2f2fc1, i258c8f2fc3, i292c9f2fc3, i293c9f2fc3, i328c10f2fc3, i329c10f2fc3 e i398c12f2fc3). Todos estes eram itens mais intensos, sendo tanto invertidos para socialização quanto afirmativos para o mesmo fator. As temáticas abordadas nos itens com categorias sem resposta foram: violência, maus tratos e gentileza. Já os itens i12c1f2fc3, i114c4f2fc1, i128c4f2fc, i153c5f2fc3, i255c8f2fc2, i362c11f2fc2 e i468c14f2fc3 apresentaram categorias com frequência de respostas semelhantes – baixa variabilidade – e também não calibraram. Contudo, tais itens possuem valores muito próximos dos valores esperados para calibração, acredita-se, portanto que com a ampliação da amostra em cada caderno tais itens calibrariam. O item i289c9f2fc1 (vide Figura 5) apresenta uma alta frequência de respostas na categoria 3 e desordem entre as categorias, assim, o mesmo não calibrou; esperava-se que tal item medisse altos níveis de socialização, pois o mesmo refere-se a resolução de problemas sociais. Hipotetiza-se que seu sentido ficou muito abrangente e sugere-se a retirada deste item do BI.

i128c4f2fc	Freq.	Percent	Cum.
1	12	5.11	5.11
2	21	8.94	14.04
3	114	48.51	62.55
4	74	31.49	94.04
5	14	5.96	100.00
Total	235	100.00	

Figura 5. Frequência das categorias do item i128c4f2fc.

Os valores de *theta* para o fator Socialização apresentaram valor máximo de 7 e mínimo de -7 (valor mínimo -7 e máximo +7) com média de 0,00. O Gráfico 17 mostra a distribuição das estimativas de *theta* das pessoas que responderam ao teste para todos os itens calibrados.

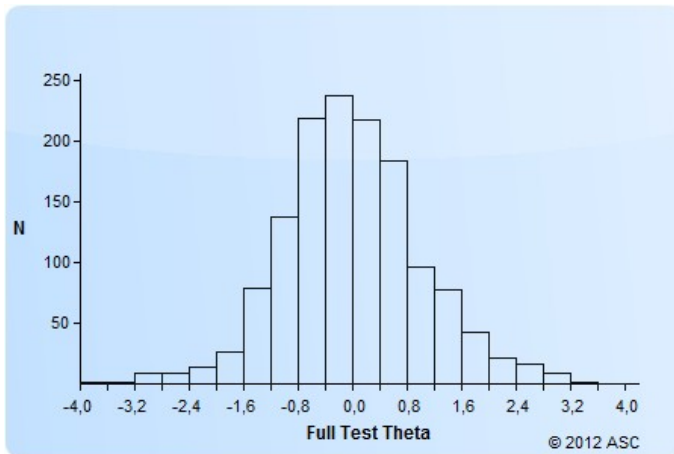


Gráfico 18 – Distribuição dos *thetas* para o fator Socialização.

O Gráfico 19 exibe a Função de Informação do Teste para todos os itens calibrados. A informação máxima foi de 27.229 em *theta* = -1.850. Observa-se que o teste para o fator socialização é mais informativo nos níveis mais baixos de *theta*, concomitantemente possui pouca informação para níveis mais elevados de *theta*.

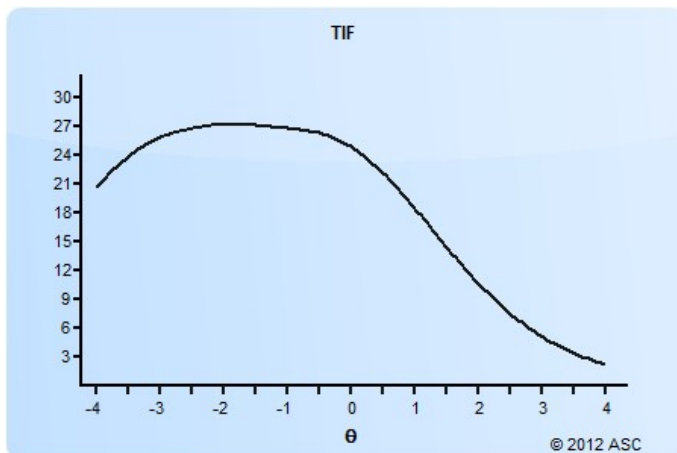


Gráfico 19. Função de Informação do Teste – fator Socialização.

No Gráfico 20 o Erro Padrão Condicional de Medida (CSEM) foi de 0,192 em $\theta = -1,850$.

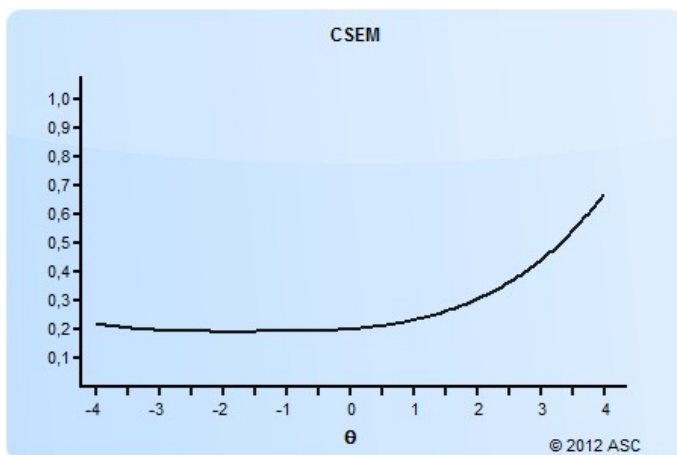


Gráfico 20. Erro Padrão Condicional de Medida (CSEM) – Socialização.

Na Tabela 6 são apresentados os valores de a para o fator Socialização. Os itens S2C3_252 ($a=0,29$) e i447c13f2fc ($a=0,492$) possuem $a < 0,5$ e foram retirados do BI. O item S2C3_252 era item âncora e, portanto, será reanalisado em estudos posteriores.

Tabela 6.

Parâmetros dos itens para todos os itens calibrados – fator Socialização.

Seq.	Item ID	Item Mean	R	a	b
47	i7c1f2fc1	4,645	0,471	1,354	-4,42, -3,58, -2,62, -0,76
48	i8c1f2fc1	3,889	0,488	1,158	-3,81, -2,23, -0,70, 0,70
49	i9c1f2fc2	3,339	0,500	1,071	-2,58, -1,27, 0,27, 1,41
50	i10c1f2fc2	3,796	0,600	1,348	-3,61, -2,07, -0,32, 0,80
58	i18c1f2fc1	4,123	0,258	0,719	-6,79, -3,53, -1,76, 0,33
85	i45c2f2fc2	2,725	0,555	1,242	-1,54, -0,38, 1,29, 2,49
86	i46c2f2fc2	3,115	0,580	1,309	-2,12, -1,06, 0,70, 2,08
87	i47c2f2fc2	2,594	0,479	1,057	-1,66, -0,29, 1,73, 3,61
117	i77c3f2fc2	2,974	0,586	1,318	-1,95, -0,94, 0,98, 2,17
125	i85c3f2fc1	4,479	0,471	1,327	-4,55, -3,30, -1,80, -0,46
141	i101c3f2fc2	3,366	0,570	1,244	-2,15, -1,33, 0,06, 1,47
158	i118c4f2fc3	4,511	0,429	1,114	-5,36, -4,67, -2,64, -0,40
188	i148c5f2fc1	4,171	0,458	1,220	-4,58, -3,56, -1,30, 0,45
189	i149c5f2fc1	4,175	0,272	0,695	-6,66, -4,35, -1,79, 0,17
192	i152c5f2fc2	4,277	0,511	1,547	-2,90, -2,33, -1,06, -0,21
223	i183c6f2fc1	4,325	0,554	1,367	-4,72, -3,26, -1,45, -0,06
224	i184c6f2fc1	3,627	0,524	1,159	-2,86, -1,76, -0,25, 1,05
225	i185c6f2fc2	3,125	0,503	1,109	-2,64, -1,31, 0,84, 2,15
226	i186c6f2fc2	4,035	0,422	0,990	-3,58, -2,47, -0,98, 0,06
227	i187c6f2fc3	4,349	0,227	0,705	-5,82, -3,80, -2,29, -0,80
260	i220c7f2fc1	4,701	0,390	1,614	-4,03, -3,33, -2,12, -1,03
261	i221c7f2fc2	4,426	0,388	1,193	-5,02, -3,84, -1,65, -0,46
284	i244c7f2fc1	4,506	0,420	1,452	-3,56, -2,82, -1,90, -0,54
293	i253c8f2fc1	3,651	0,555	1,480	-2,43, -1,66, -0,37, 1,01
299	i259c8f2fc3	4,311	0,420	1,138	-3,53, -3,08, -1,83, -0,08
300	i260c8f2fc	4,366	0,514	1,512	-4,33, -2,91, -1,78, 0,02
334	i294c9f2fc3	3,980	0,499	1,397	-2,77, -1,73, -0,65, 0,26
363	i323c10f2fc1	4,000	0,454	1,109	-4,31, -3,51, -1,04, 0,60
364	i324c10f2fc1	4,155	0,520	1,278	-3,58, -3,39, -1,53, 0,40
365	i325c10f2fc1	3,548	0,483	1,138	-3,57, -2,06, -0,23, 1,44
398	i358c11f2fc1	4,192	0,409	1,193	-3,61, -2,79, -1,52, 0,05
399	i359c11f2fc1	4,368	0,565	1,655	-3,47, -3,29, -1,94, 0,01
401	i361c11f2fc2	4,498	0,392	1,235	-4,24, -3,66, -2,13, -0,63
403	i363c11f2fc3	3,708	0,574	1,413	-2,61, -1,71, -0,60, 0,95
404	i364c11f2fc3	3,348	0,570	1,333	-2,53, -1,45, 0,15, 1,63
433	i393c12f2fc1	4,078	0,377	0,877	-6,50, -5,20, -1,69, 1,01
434	i394c12f2fc1	4,113	0,517	1,226	-4,89, -3,71, -1,12, 0,50
435	i395c12f2fc2	3,598	0,514	1,197	-3,16, -2,05, -0,18, 1,37
436	i396c12f2fc2	4,494	0,519	1,353	-4,52, -3,42, -2,23, -0,33
437	i397c12f2fc3	4,377	0,612	1,607	-3,97, -3,03, -1,59, -0,06
439	i399c12f2fc3	4,429	0,398	0,999	-5,76, -4,11, -2,32, -0,33

Seq.	Item ID	Item Mean	R	a	b
468	i428c13f2fc1	4,471	0,350	1,021	-4,68, -3,56, -2,18, -0,80
469	i429c13f2fc1	4,667	0,420	1,104	-5,38, -4,74, -2,76, -1,18
470	i430c13f2fc1	3,291	0,472	1,069	-2,75, -1,45, 0,24, 1,77
471	i431c13f2fc2	3,811	0,754	2,008	-3,05, -2,04, -0,54, 0,86
472	i432c13f2fc2	4,174	0,598	1,560	-4,18, -3,26, -1,05, 0,22
473	i433c13f2fc3	3,692	0,577	1,335	-3,10, -1,92, -0,48, 1,14
487	i447c13f2fc	3,150	0,167	0,492	-3,73, -1,61, 0,51, 3,04
503	i463c14f2fc1	4,293	0,219	0,608	-6,21, -4,85, -2,38, -0,45
504	i464c14f2fc1	3,076	0,327	0,708	-3,97, -1,35, 1,24, 3,08
505	i465c14f2fc2	4,377	0,450	1,097	-4,12, -3,89, -1,67, -0,32
506	i466c14f2fc2	3,979	0,410	0,982	-3,88, -2,54, -1,16, 0,69
539	i499c15f2fc1	4,439	0,459	1,396	-3,42, -2,49, -1,52, -0,69
544	i504c15f2fc3	4,333	0,325	0,991	-4,40, -3,15, -1,68, -0,59

Os limiares de categoria (*thresholds*) do parâmetro *b* para Socialização são apresentados no Gráfico 21, no qual se observa a escassez de itens com valores acima de 1. O grande número de itens que não calibraram (14) podem ter contribuído para este panorama. Acresce-se que para Socialização, apesar de termos níveis de *theta* variados, tanto mais quanto menos elevados (visto no Gráfico 18), não há itens correspondentes (Gráfico 21). Sugere-se, em novos estudos, a elaboração de itens com dificuldades acima de 1 e a ampliação da amostra com vistas à calibração dos itens que ainda não calibraram.

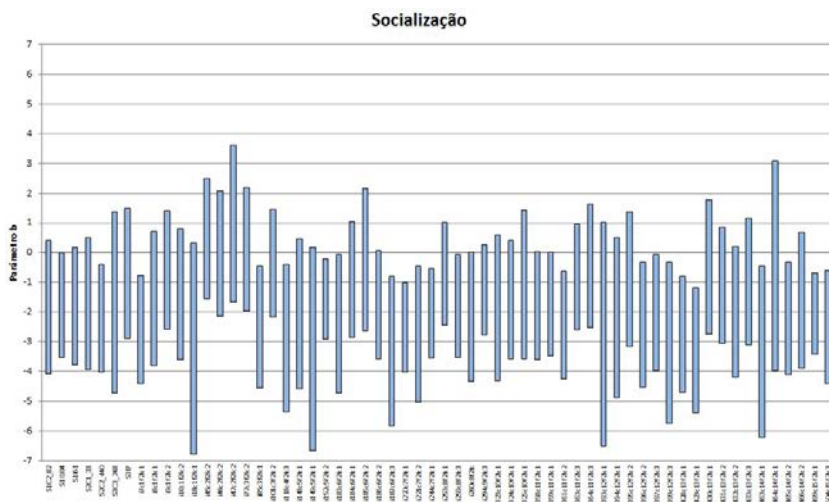


Gráfico 21 - Parâmetros *bs*: Limiares de categorias máximos e mínimos dos itens – fator socialização.

Considerações Finais

A Tabela 7 resume os resultados da calibração. No fator Abertura, 73 itens novos calibraram – os itens que não calibraram (5 itens) foram analisados e verificou-se que tratavam-se de itens mais intensos, deste modo foram sugeridos novos estudos e estes foram temporariamente retirados do BI. Dos itens com a baixo ($a < 0,5$), 4 no total, 2 foram excluídos do BI e 2 deles apresentaram valores mais próximos de 0,5 e foram encaminhados para novos estudos, pois há a possibilidade de que com o aumento da amostra o valor de a sofrer um incremento. Assim, para Abertura têm-se 64 itens no BI.

Para Extroversão calibraram 63 itens novos, destes 1 item foi retirado ($a < 0,5$) e 4 foram encaminhados para novos estudos, totalizando 58 itens de Extroversão no BI. Para Neuroticismo 86 itens novos calibraram e 1 item foi retirado por não ter alcançado o objetivo esperado, 1 foi encaminhado para novos estudos por ter a próximo de 0,5 e 2 também foram encaminhados para novos estudos por tratarem-se de itens intensos, restando, assim 82 itens novos do fator Neuroticismo no BI.

No fator Realização 77 itens calibraram. Dentre estes 9 itens foram encaminhados para novos estudos com características similares aos demais que já foram encaminhados devido, sobretudo, as hipóteses de intensidade e necessidade de especificidade ou incremento da amostra nos cadernos. Em Realização, 3 itens apresentaram a baixos. Destes, 2 foram excluídos e 1 foi encaminhado para novos estudos por apresentar um valor próximo de 0,5.

Já Socialização foi o fator que apresentou o maior número de itens que não calibraram, 15 itens. Dentre estes, 14 foram encaminhados para novos estudos pelas mesmas razões dos demais fatores, como a intensidade do item, e 1 item não funcionou conforme o esperado e foi excluído do BI. Apenas 1 item novo apresentou $a < 0,5$, contudo este valor se aproxima de 0,5 e o item foi encaminhado para novos estudos. Ainda, inesperadamente 1 item âncora apresentou $a < 0,5$ e o mesmo será reanalisado e submetido a novos estudos questionando-se seu caráter de âncora. Para o fator Socialização 37 itens novos foram para o BI.

Conclui-se, portanto, que após este estudo de calibração 317 itens estão aptos a compor o Banco de Itens de Personalidade e Testagem Universal.

Tabela 7.

Resumo dos Resultados da Calibração.

	Análise fatorial	Calibração	Itens retirados	Itens para novos estudos	BI parcial
Abertura	78	73	2	7	64
Extroversão	67	63	1	4	58
Neuroticismo	89	86	1	3	82
Realização	89	88	2	10	76
Socialização	69	53	1	15	37
Total	392	363	7	39	317

Como este estudo fez parte de um estudo maior de desenvolvimento de um BI de Personalidade e Testagem Universal, o mesmo não possui um caráter estático. Deste modo, os próximos passos em novos estudos serão a equiparação do BI de Personalidade e TU com o BI já utilizado no Laboratório de Pesquisa em Avaliação Psicológica por meio dos itens âncoras agregando-o a Bateria Adaptativa de Personalidade⁷.

Referências

- Andrade, D. F. Tavares, H. R. & Valle, (2000). Teoria da Resposta ao Item: Conceitos e Aplicações. São Paulo: SINAPE 2000. 154 p.
- Beckman, R. M. (2001). Aplicação dos Blocos Incompletos Balanceados na teoria de resposta ao item. *Estudos em Avaliação Educacional*, nº 24. Recuperado de <http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/990/990.pdf>, em Agosto de 2016.
- Castro, S. M. de J., Trentini, C., & Riboldi, J. (2010). Teoria da resposta ao item aplicada ao Inventário de Depressão Beck. *Rev Bras Epidemiol*. 13(3): 487-501. Recuperado em Fevereiro de 2014, de http://www.scielo.br/scielo.php?pid=S1415-790X2010000300012&script=sci_arttext.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates. p. 371.
- Fisher, S. R. A.; Yates, F. (1963). *Statistical Tables for biological, agricultural and medical research*. Sixth Edition. Recuperado de

⁷ Link para acessar o instrumento:
<http://concerto4.labape.com.br/?wid=4&tid=1>

https://hekyll.services.adelaide.edu.au/dspace/bitstream/2440/10701/1/stat_tab.pdf, junho de 2015.

- Guyer, R., & Thompson, N.A., (2011). *User's Manual for Xcalibre 4.1*. St. Paul MN:
- Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning and Assessment*, 4(2). Recuperado em agosto de 2011, de <http://www.jtla.org>.
- Moreira Junior, F. de J. (2011) Sistemática para Implantação de Testes Adaptativos Informatizados Baseados na Teoria da Resposta ao Item. Tese. Recuperado de <https://repositorio.ufsc.br/handle/123456789/95506>, em junho de 2015.
- Nunes, C. H. S. S. & Primi, R. (2009). Teoria de Resposta ao Item: Conceitos e Aplicações na Psicologia e na Educação. In: Hutz, S. C. (org.) *Avanços e Polêmicas em Avaliação Psicológica: em homenagem a Jurema Alcides Cunha* (2009). São Paulo: Casa do Psicólogo.
- Oliveira, C. M. & Nunes, C. H. S. S. (2015). Instrumentos para Avaliação Psicológica de Pessoas com Deficiência Visual: Tecnologias para Desenvolvimento e Adaptação. *Psicologia: Ciência e Profissão*, 2015, 35(3), 886-899. <http://dx.doi.org/10.1590/1982-3703001902013>.
- Oliveira, C. M. & Nunes, C. H. S. S. (2015). Instrumentos para Avaliação Psicológica de Pessoas com Deficiência Visual: Tecnologias para Desenvolvimento e Adaptação. *Psicologia: Ciência e Profissão*, 2015, 35(3), 886-899. <http://dx.doi.org/10.1590/1982-3703001902013>.
- Oliveira, C. M. ; Nuernberg, A. H. & Nunes, C. H. S. S. (2013). Desenho universal e avaliação psicológica na perspectiva dos direitos humanos. *Avaliação Psicológica*. Vol. 3. <http://goo.gl/qtkhx3>.
- Oliveira, C. M. ; Nuernberg, A. H. & Nunes, C. H. S. S. (2013). Desenho universal e avaliação psicológica na perspectiva dos direitos humanos. *Avaliação Psicológica*. Vol. 3. <http://goo.gl/qtkhx3>.

- Oliveira, S. M. da S. S. (2012). *O Modelo de Rasch para Avaliar o Inventário de Ansiedade na Escola*. Tese (doutorado). Programa de pós-graduação *Stricto Sensu* em Psicologia da Universidade São Francisco, Itatiba. 141 p.
- Pasquali L. (2003). *Psicometria: Teoria dos testes na Psicologia e na Educação*. Petrópolis, RJ: Vozes. p. 397.
- Pasquali, L. (2007). *Teoria de Resposta ao Item*. Brasília: LabPAM/UnB. p. 236.
- Pasquali, L. (2010). *Instrumentação Psicológica*. Porto Alegre: Artmed.
- Story, M. F., Mueller, L. J. & Mace, R. L. (1998). The Universal Design File: Designing for People of all Ages and Abilities. *Design Research and Methods Journal*. Recuperado em Agosto de 2011, de http://www.ncsu.edu/ncsu/design/cud/pubs_p/pudfiletoc.htm.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Recuperado em Agosto de 2011, de <http://goo.gl/EmWQZ3>.
- Thurlow, M., Lazarus, S. S., Albus, D., & Hodgson, J. (2010). *Computer-based testing: Practices and considerations* (Synthesis Report 78). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Artigo 5. Funcionamento Diferencial dos itens de um teste para avaliação da personalidade baseado na Testagem Universal.

Resumo

Uma das técnicas utilizadas para verificar a invariância da medida é o Funcionamento Diferencial do Item (DIF), a qual é derivada da Teoria de Resposta ao Item (TRI). No âmbito dos testes desenvolvidos ou adaptados para as pessoas com deficiência sob a perspectiva do Desenho Universal/Testagem Universal, o DIF permite verificar se indivíduos com a mesma característica psicológica, valores semelhantes de θ , pertencentes a grupos diferentes, apresentam padrões semelhantes de respostas; por isso a sua utilidade nos estudos de construção e adaptação de testes é destacada. O objetivo da pesquisa consistiu em verificar a invariância dos parâmetros dos itens de um banco de itens de Personalidade e Testagem Universal entre indivíduos com e sem deficiência. Considerou-se, nestas análises, itens passíveis de exclusão aqueles nos quais o DIF apresentou valores de moderados a altos ($|\text{DIF}| \geq 0,64$). Os resultados da aplicação da Testagem Universal foram bastante promissores na busca pela invariância dos itens, uma vez que poucos itens apresentaram valores de DIF de moderados a altos. Para novos estudos sugere-se a ampliação da amostra de pessoas com deficiência e a ampliação do número de itens do banco que lhes sejam aplicados.

Palavras-chave: Funcionamento Diferencial dos itens; Testagem Universal; Invariância dos parâmetros; Personalidade.

Abstract

One of the techniques to verify the invariance of the measure is the Differential item Function (DIF). The DIF is part of the Item Response Theory (TRI). There are no results for this article. Please contact your world representative. ; for this reason its usefulness in the studies of construction and adaptation of testicles. The objective of the research was to verify the invariance of the parameters of the items among the individuals with deficiency of a bank of Personality and Universal Test items. Consider, in these analyzes, as indicative items of exclusion in which DIF presented moderate to high values ($|\text{DIF}| \geq 0.64$).The results of the application of the Universal Testing were very promising in the search for the invariance of the items since few items presented moderate to high DIF values. For further studies it is suggested an expansion of the sample of people with disabilities and an increase in the number of bank items that are applied.

Keywords: Differential item Functioning; Universal Testing; Invariance of parameters; Personality.

Introdução

A preocupação sobre as condições em que os testes são aplicados para grupos diferentes de pessoas trouxe consigo perguntas como “Os itens desse teste medem o mesmo atributo para os grupos?” ou “Os escores dos testes favorecem mais um grupo do que outro?” (McDonald, 1999). O conceito de invariância da medida abarca estas perguntas. Matematicamente, o conceito de invariância da medida é atrelado à probabilidade condicional. Ou seja, uma variável randômica X é tida como uma medida invariante com relação à seleção de V se

$$F(x|w, v) = F(x|w)$$

para todos (x, w, v) na amostra. Então, X é uma variável randômica da dimensão n com realização x , W é uma variável latente com dimensão p e realização w que X mede e V é uma variável randômica com dimensão m e realização v que acomoda a base de seleção da subpopulação (Meredith, 1993). A partir dessa notação, a invariância da medida existe quando dois indivíduos de grupos diferentes, mas igualmente hábeis, apresentam a mesma expectativa (ou probabilidade) de receberem o mesmo escore.

Uma das técnicas utilizadas para verificar a invariância da medida é o Funcionamento Diferencial do Item (DIF)⁸. O DIF faz parte da Teoria de Resposta ao Item (TRI) e é uma metodologia de análise dos itens. O DIF acontece quando o item de um teste não tem a mesma relação com o traço latente, ou habilidade, em dois ou mais grupos de respondentes (Embretson & Reise, 2000). Para Zumbo (2007), o uso do DIF está passando por diversas mudanças metodológicas e teóricas nos últimos anos, ampliando a perspectiva de que o DIF não se refere apenas ao viés do item, a perspectiva deste autor será detalhada subsequentemente neste artigo.

O Funcionamento Diferencial do Item do inglês *Differential Item Functioning* (DIF) e a Funcionamento Diferencial do Teste (DTF) correspondem à aplicação do conceito de viés (*bias*) na Teoria de Resposta ao Item (TRI), como já referido no parágrafo anterior. Surgiram da preocupação dos especialistas (Holland & Wainer, 1994;

⁸ Neste artigo “ocorrência de DIF” indica DIF significativo.

Holland & Thayer, 1988; Wright, 1968; Rasch, 1966; Zumbo 2007, entre outros) com a possibilidade das medidas em Psicologia funcionarem de forma diferente para grupos particulares de indivíduos (ex. homens e mulheres, estrangeiros e nativos).

Os estudos do DIF partem do pressuposto que para que um teste tenha validade ele deverá atender a três características, sendo a terceira delas inerente a presença ou ausência de DIF (Sisto, 2006): 1. A pessoa com maior habilidade (maior quantidade em um dado traço latente) deverá apresentar maior possibilidade de acerto ou de endossar um item que uma pessoa com menor habilidade em um traço latente; 2. As pessoas deverão ter maior probabilidade de acertos de itens comuns (ou fáceis) do que itens raros (ou difíceis); e, 3. Essas condições devem ser produto da pessoa e da posição do item no traço e não devem se alicerçar na raça, sexo, ou outras características. Assim, o DIF significativo ocorre quanto características irrelevantes interferem neste sistema, admite-se uma nova dimensão alheia e indesejada ao construto/traço latente avaliado pelo teste.

O Funcionamento Diferencial do Item ocorre quando sujeitos com aptidão igual, mas de grupos diferentes (culturalmente/socialmente), não tem a mesma probabilidade de responder a um item específico e a Funcionamento Diferencial do teste refere-se ao mesmo raciocínio, mas aplica-se ao teste como um todo. Os testes devem fornecer medidas relativamente idênticas, considerando que toda medida possui algum erro, para pessoas com um mesmo nível de habilidade no construto ao qual o teste se propõe a medir. Contudo, quando em decorrência de características irrelevantes como ser do sexo feminino ou masculino, pertencer a diferentes grupos étnicos e/ou culturais e até mesmo em se tratando de pessoas com deficiência as pontuações do teste como um todo se mostram diferentes fala-se de Funcionamento Diferencial do Teste e, tal resultado, pode comprometer a validade do instrumento (Sisto, 2006; Embretson & Reise, 2000).

No Funcionamento Diferencial do Teste um conjunto de itens subestima sistematicamente a habilidade de certos grupos em função de características diferenciadoras irrelevantes afetando o resultado final do teste – sua pontuação ou escore. Já, no tocante aos itens (DIF), não necessariamente o resultado final será afetado, pois pode ocorrer a existência de itens que proporcionalmente subestimem cada um dos grupos o que de certa forma neutraliza os efeitos (Embretson & Reise, 2000).

A este ponto cabe-se diferenciar DIF e FDT de impacto, uma vez que, com frequência, essas conceituações são confundidas (Sisto,

2006; Andriola, 2001; Embretson & Reise, 2000). O impacto consiste em diferenças reais nos grupos que são confirmadas através da medida. A validade dos testes e concomitantemente sua qualidade é apontada quando se detectam diferenças entre os grupos quando essas diferenças realmente existam. Um dado grupo de indivíduos pode possuir maior habilidade em um construto que outro grupo e, deste modo sistematicamente acertar a um item ou a conjunto de itens por possuir uma maior habilidade neste construto. No tocante aos homens e mulheres, por exemplo, em alguns construtos as mulheres apresentam maiores níveis de habilidade em detrimento dos homens ou vice e versa. Porém, trata-se de diferenças reais não causadas por fontes sistemáticas de variância de erro alheias ao construto medido.

No âmbito dos testes desenvolvidos ou adaptados para as pessoas com deficiência, sob a perspectiva do Desenho Universal/Testagem Universal, o DIF permite verificar se indivíduos com a mesma característica psicológica (mesmo valor de θ nos itens) pertencentes a grupos diferentes apresentam padrões semelhantes de respostas; por isso a sua utilidade nos estudos de construção e adaptação de testes é de grande destaque. Assim, por exemplo, um indivíduo cego e um vidente com o mesmo nível de Extroversão devem apresentar padrões de respostas semelhantes em um teste para avaliação da personalidade adaptado ou desenvolvido segundo os princípios da Testagem Universal. Caso isso não ocorra, o resultado do teste não será justo para avaliar ambos os grupos pois o teste não se propõe a avaliar capacidade visual, mas sim um traço de personalidade.

Considera-se que seja importante notar que não se supõe que os grupos devam ter resultados gerais semelhantes em relação ao traço avaliado. Indivíduos cegos podem apresentar níveis mais elevados, por exemplo, de Extroversão que os videntes - o que significa que há impacto entre os grupos - e ainda assim os itens poderão não ter DIF. A questão é que um cego e um vidente com o mesmo nível de Extroversão devem apresentar padrões de respostas semelhantes. Ressalta-se que os padrões não são idênticos e sim semelhantes por que há sempre que considerar-se alguma variância de erro na medida.

Há na literatura científica da área (Sisto, 2006; Andriola, 2001; Embretson & Reise, 2000) o destaque de dois tipos de abordagem para detecção do DIF, o critério externo e o interno. O critério externo preocupa-se com a correlação entre os resultados do teste e variáveis estranhas ao teste, ou seja, irrelevantes para sua interpretação. O critério interno parte do princípio que sempre se deve comparar a probabilidade de acerto de um item em sujeitos com a mesma magnitude no traço

latente avaliado pelos itens. Os métodos de critério interno para detectar DIF se dividem segundo os recursos estatísticos utilizados em (Sisto, 2006): análise de variância, dificuldades transformadas do item, tabelas de contingência e métodos da Teoria de Resposta ao Item (TRI). Sisto (2006), Andriola (2001) e Embretson & Reise (2000) ressaltam os procedimentos da TRI como os mais promissores na detecção do DIF.

Os métodos para detecção do DIF que não utilizam a TRI baseiam-se em testes de significância, comparações dos valores de p em grupos de referência e focais (p representa a proporção de respostas corretas para itens dicotômicos), tais métodos não são considerados eficazes já que inflam as taxas de erro e são extremamente dependentes da amostra. No Modelo Logístico de Rasch encontra-se a primeira forma de resolver satisfatoriamente o problema da detecção do DIF. Rasch descreve a probabilidade de acerto em um item em função da habilidade do sujeito e da dificuldade do item, esse ponto de partida permite que se tenham medidas mais seguras de DIF. De modo geral para o cálculo de DIF na TRI são necessários alguns processos, como estimar os parâmetros métricos dos itens para os grupos, colocar os parâmetros em uma mesma métrica, representar os parâmetros utilizando as suas curvas características e comparar estatisticamente os parâmetros nos grupos (Pasquali, 2007; Sisto, 2006).

A TRI possui os métodos mais confiáveis para a verificação da DIF. Contudo, para que esses métodos possam demonstrar a sua potência e flexibilidade com qualidade, Sisto (2006) recomenda que: 1. Os tamanhos das amostras devam ser adequados à estimativa dos parâmetros dos itens; 2. Sejam utilizados mais de um método para detectar o DIF; e 3. Alie-se aos recursos estatísticos procedimentos qualitativos. Diversos autores (Zumbo, 2007; Sisto, 2006; Camilli & Shepard, 1994; O'Neill & McPeck, 1993) ressaltam ainda que o foco nos resultados matemático-estatísticos não deve ser o único enfoque a ser utilizado pelos pesquisadores, os mesmos devem tecer considerações teóricas relevantes e aprofundadas sobre as possíveis causas do DIF.

No intuito de alcançar a amplitude necessária nos estudos da DIF busca-se não só constatar a presença de valores de DIF que prejudicam a medida, mas também as causas para que o item ou teste apresentem esses valores. Quando valores prejudiciais são encontrados é recomendada a retirada do item ou mesmo a reformulação do instrumento. Em casos específicos, o teste pode ter seu uso recomendado apenas para avaliar um grupo particular evitando-se que seja utilizado pelo grupo para o qual o teste apresenta DIF. Sobretudo, fazem-se necessários estudos que visem evitar a construção de novos itens com os

mesmos problemas. Assim, conhecendo-se o porquê de o item ser inadequado torna-se possível traçar diretrizes para a construção de itens com valores toleráveis de DIF (Zumbo, Liu, Wu, Shear, Astivia & Ark, 2015; Zumbo, 2007; Sisto, 2006; Andriola, 2001; Embretson & Reise, 2000).

Em consonância com as preocupações levantadas no parágrafo anterior aponta-se na relação entre o DIF e a Testagem Universal. Vários autores (Johnstone, Altman & Thurlow, 2006; Johnstone, Thompson, Moen, Bolt & Kato, 2005; Ketterlin-Geller, 2005; Johnstone, 2003; Thompson, Johnstone & Thurlow, 2002) assinalam a importância do uso da DIF em estudos de Testagem Universal. Uma vez que a Testagem Universal trabalha com grupos amplos e diversos, como pessoas com e sem deficiência, conflui para a aplicação de recursos estatísticos que estudam se um item funciona de forma semelhante em grupos distintos.

Testagem Universal e o DIF

A Testagem Universal busca a acessibilidade plena ou máxima possível aos itens e ao conteúdo do teste como um todo para grupos de testandos amplos e inclusivos, podendo abarcar desde pessoas com ou sem deficiência, indivíduos com déficits cognitivos, grupos étnicos variados, etc. Para tanto se preocupa com o conteúdo dos itens do teste, a sua forma de apresentação, organização estrutural e textos de explicação da aplicação permeando tanto os elementos de padronização do teste como seu conteúdo teórico (Oliveira, 2013; Oliveira; Nuernberg & Nunes, 2013). Ressalta-se que, no tocante aos itens, prima-se pela simplicidade dos mesmos adequando seus conteúdos para que sejam de fácil entendimento evitando-se, por exemplo, palavras rebuscadas e em grande número (Gaster & Clark, 1995; Shinn & Ofiesh, 2012).

A Testagem Universal respeita todos os pressupostos teóricos instituídos no tocante aos testes psicológicos preocupando-se de forma mais específica com o formato dos testes e sua influência na realização do mesmo. Foca as adaptações e os recursos de tecnologias assistivas enquanto proporcionadores de acessibilidade analisando sua repercussão na qualidade dos instrumentos (Ketterlin-Geller, 2005; Thompson, Johnstone & Thurlow, 2002). Os princípios da Testagem Universal são (Thompson, Johnstone & Thurlow, 2002):

1. População de avaliação ampla e inclusiva

Os testes psicológicos são desenvolvidos ou adaptados objetivando uma população ampla, geralmente pessoas com e sem

deficiências. Atende a uma demanda crescente por maior inclusão da diversidade humana.

2. *Definição precisa do construto*

Os testes psicológicos são desenvolvidos ou adaptados visando à máxima definição dos construtos a serem medidos evitando que possam ser afetados por outros construtos alheios ao que se pretende medir. Para tanto se faz imprescindível que os instrumentos empreguem teorias fortemente embasadas.

3. *Itens acessíveis e não tendenciosos*

Os testes psicológicos são desenvolvidos ou adaptados para que indivíduos pertencentes a diferentes grupos (deficientes, estrangeiros, sem deficiência, entre outros) com a mesma habilidade no construto avaliado tenham a mesma chance de acertar os itens do teste psicológico (em testes de desempenho) ou a apresentarem certos padrões de respostas (em testes de personalidade, interesse, entre outros), e que todos estes grupos compreendam as instruções dos procedimentos envolvidos na testagem.

4. *Testes flexíveis a acomodações*

Os testes psicológicos são desenvolvidos ou adaptados de tal forma que apliquem os princípios do Desenho Universal para os tornarem flexíveis a acomodações, como, por exemplo, as tecnologias assistivas.

5. *Instruções e procedimentos simples, claros e intuitivos*

Os testes psicológicos são desenvolvidos ou adaptados evitando-se que instruções e procedimentos em linguagem complexa ou confusa prejudiquem a compreensão do que é questionado, evitando que o testando incorra em erros por não compreender o que lhe é pedido mesmo possuindo habilidade no construto medido.

6. *Leitura agradável e de máxima inteligibilidade*

Os testes psicológicos são desenvolvidos ou adaptados para reduzir a complexidade verbal e de organização textual dos itens e instruções preservando seu conteúdo essencial.

7. *Máxima legibilidade*

Os testes psicológicos são desenvolvidos ou adaptados de forma que o texto, os gráficos, tabelas e ilustrações, bem como o formato das respostas possam ser percebidos e decifrados com facilidade.

Analisando os princípios da Testagem Universal verifica-se que o terceiro princípio está intrinsecamente relacionado à DIF, o próprio título do princípio aponta para os estudos de viés, 3. Itens acessíveis e não tendenciosos (3. *Accessible, non-biased items*) (Oliveira, Nuernberg

& Nunes, 2013). Na busca pela não tendenciosidade dos itens aporta-se no Funcionamento Diferencial do item. Deste modo, a aplicação dos métodos do DIF reveste-se de profunda relevância nos estudos de Testagem Universal.

Em se tratando do DIF e da TRI enquanto técnicas estatísticas aliadas da Testagem Universal realizou-se a busca em bases de dados internacionais, na qual foram encontrados 8 artigos que abordam a Testagem Universal e que especificam as análises estatísticas utilizadas (Tabela 1). Foram pesquisados tanto textos de abordagem mais teórica quanto de relato de pesquisas aplicadas. Os modelos da DIF estavam presentes em cinco artigos e a TRI, de forma geral, isto é, não necessariamente relacionada aos modelos da DIF, em sete dos textos encontrados (Tabela 1). Esse fato reforça a importância do uso da TRI e mais especificamente do DIF nos estudos da Testagem Universal.

Tabela 1.

Análises Estatísticas para Testagem Universal.

Título	Autores	Ano	Análises destacadas
1. <i>Universal Design Applied to Large Scale Assessments</i>	Thompson, Johnstone e Thurlow	2002	DIF e TRI.
2. <i>Improving Validity of Large-scale Tests: Universal Design and Student Performance</i>	Johnstone	2003	DIF, TRI e Teste-T.
3. <i>Analyzing Results of Large-scale Assessments to Ensure Universal Design (Technical Report 41)</i>	Johnstone, Thompson, Moen, Bolt e Kato	2005	Ranking de Itens, Correlação Item Total, DIF usando Tabelas de Contingência e DIF usando TRI.
4. <i>Knowing What all Students Know: Procedures for Developing Universal Design for Assessment</i>	Ketterlin-Geller	2005	DIF e TRI.
5. <i>Applying Principles of Universal Design to Test Delivery: The</i>	Dolan, Hall, Banerjee, Chun e Strangman	2005	Teste-T.

Título	Autores	Ano	Análises destacadas
<i>Effect of Computer-based Read-aloud on Test Performance of High School Students With Learning Disabilities</i>			
6. <i>Inclusive Design for Maximum Accessibility: A Practical Approach to Universal Design</i>	Hanna	2005	TRI.
7. <i>A State Guide to the Development of Universally Designed Assessments</i>	Johnstone, Altman e Thurlow	2006	Ranking de Itens, Correlação Item Total, DIF usando Tabelas de Contingência e DIF usando TRI.
8. <i>Measuring Cognition of Students with Disabilities Using Technology-Enabled Assessments: Recommendations for a National Research Agenda</i>	Bechard, Sheinker, Abell, Barton, Burling, Camacho, Cameto, Haertel, Hansen, Johnstone, Kingston, Murray, Parker, Redeld e Tucker	2010	TRI.

No DIF aplicado a Testagem Universal os aspectos a serem analisados são referentes a todo o conteúdo que em um determinado item é capaz de proporcionar vantagens ou desvantagens a um subgrupo de testandos, desde uma palavra até o item completo. Nos estudos do DIF há, geralmente, uma preocupação com aspectos qualitativos que envolvem o conteúdo dos itens e a verificação se o item está medindo realmente o construto que pretende medir. Deste modo, os estudos de DIF estão também, indiretamente relacionados ao segundo princípio, pois uma 2. *Definição precisa do construto* será essencial para a construção dos itens e a concomitante criação de itens de qualidade e livres de qualquer viés prejudicial à medida (Thompson, Johnstone & Thurlow, 2002).

Algumas possíveis fontes de viés são: 1. Itens com texto longo; 2. Palavras conhecidas apenas por um subgrupo de testandos; 3. Itens que expressem preconceitos; 4. Itens que expressem valores culturais específicos de um subgrupo de testandos; 5. Expressões e frases que possuem significados diferentes para um subgrupo de testandos; 6. Itens que medem uma habilidade, mas, são dependentes de outras habilidades

para serem respondidos corretamente; 7. Itens cujo conteúdo não pertence ao construto avaliado de forma direta e clara; 8. Interferência das opiniões pessoais ou de preconceitos do elaborador do item em seu conteúdo; 9. Desconhecimento da realidade (cultura e hábitos) das pessoas com deficiência. Os pesquisadores e desenvolvedores de itens devem ficar atentos a tais fontes e evitá-las sempre que possível (Oliveira, Nuernberg & Nunes, 2013; Thompson, Johnstone & Thurlow, 2002).

Outras fontes de erro dos testes, mas que não consistem em DIF ou FDT são as instruções e o formato do instrumento. Instruções pobres, equivocadas ou inadequadas afetam o resultado do teste, mas não significa que o item possua viés. Nesse caso o item é mal compreendido não pelo seu conteúdo em si, mas pelas instruções insatisfatórias. Instrumentos sem adaptação das instruções para os diferentes tipos de deficiência, por exemplo, podem ocasionar esse tipo de problema. Do mesmo modo, no tocante ao formato dos itens, dependendo das escolhas perpetradas pelo desenvolvedor, esses proporcionam a acessibilidade para os indivíduos com deficiência ou tornam-se barreiras, principalmente quando não são desenvolvidos ou adaptados visando essa amostra. Tais aspectos se assemelham ao DIF - foram indiretamente incorporados ao modelo ecológico da terceira geração do DIF, o qual será apresentado no próximo tópico (Zumbo, 2015) - e se enquadram no estudo da Testagem Universal em vários princípios desta, principalmente o 5 e o 7 (respectivamente *Instruções e procedimentos simples, claros e intuitivos* e *Máxima legibilidade*) (Oliveira, 2013; Thompson, Johnstone e Thurlow, 2002).

Uma forma aplicada tradicionalmente na construção de itens e que corrobora para a diminuição da probabilidade de ocorrência de DIF é a análise teórica dos itens. Tal análise é realizada por juízes e visa verificar se os itens estão medindo realmente o traço latente que pretendem medir. Os dois tipos de análise teórica são a análise semântica e a análise de juízes. Na primeira, os juízes são indivíduos da própria população alvo do teste e objetivam verificar se os itens são inteligíveis para os níveis de habilidade mais baixos da população, mas, evitando-se também, que as simplificações realizadas para alcançar os níveis mais baixos de habilidade não acarretem deselegância na formulação dos mesmos (Pasquali, 2003). Em se tratando da Testagem Universal a preocupação será em relação aos conteúdos irrelevantes, preconceituosos ou que prejudiquem, por exemplo, os indivíduos com deficiência na execução de algum item. Em se tratando do público com deficiência o

desafio é abarcar a amplitude dos diferentes tipos de deficiências e seus graus de comprometimento sem incorrer em DIF.

Já na análise de juízes, os juízes são peritos na área que avaliam o conteúdo dos itens, ou seja, a adequação dos mesmos como representação comportamental do construto medido. Os juízes possuem a tarefa de avaliar se os itens estão se referindo ao traço latente avaliado. Na Testagem Universal a utilização da análise de juízes corrobora para evitar as fontes de viés já mencionadas em parágrafos anteriores. Contudo, os especialistas conhecedores da Testagem Universal além de analisar os itens quanto ao seu conteúdo no que se refere ao construto são capazes, ainda, de avaliar os itens quanto à pertinência aos princípios da Testagem Universal. Tal julgamento enquadra-se na análise teórica dos itens e é semelhante à análise de juízes, mas caracteriza-se como uma análise de Testagem Universal. Assim, as análises teóricas dos itens contribuem para evitar o viés dos itens (Oliveira, 2013).

Na aplicação do DIF a Testagem Universal busca-se a máxima qualidade dos itens aliando-se os avanços no campo da estatística aos princípios da TU. Ainda, no que se refere aos aspectos quantitativos observa-se a junção de mais de um recurso estatístico para o estudo da DIF como ocorre em três artigos (Tabela 1) (Johnstone, Altman e Thurlow, 2006; Johnstone, Thompson, Moen, Bolt e Kato, 2005; Johnstone, 2003). Porém, além dos aspectos quantitativos, os qualitativos são abordados e aprofundados por meios de técnicas como a análise dos itens abordadas nos parágrafos anteriores. Portanto, os estudos de DIF envolvem aspectos quantitativos e qualitativos de forma complementar possibilitando aos pesquisadores uma visão ampla do problema abordado. Enfim, os resultados de indivíduos com a mesma habilidade, mesmo valor no traço latente tem que ser comparáveis de pessoa para pessoa, de grupo por grupo, de conteúdo a conteúdo.

Testagem Universal e a perspectiva de Zumbo

Zumbo (1999; 2007; 2013; 2015) destaca-se como teórico do estudo de DIF e seus textos têm influenciado os estudiosos desta temática, gerando avanços neste campo. Em seu artigo de 2007, ele descreve três gerações do DIF. A primeira geração refere-se ao surgimento do conceito de viés do item e do teste e ao risco da tomada de decisão no que envolve aptidão, certificação e licenciamento em questões de justiça, por exemplo, assim, a preocupação com a equidade fazia-se primordial. Os estudos de primeira geração centravam-se no desempenho diferencial de grupos com base, principalmente, em sexo e

raça. Em termos de procedimentos estatísticos na primeira geração de DIF, destacam-se: 1) As respostas aos itens modeladas por tabelas de contingências ou por modelos de regressão; e, 2) A teoria de resposta ao item.

Já a segunda geração é marcada pela utilização do termo DIF ao invés de “parcialidade/imparcialidade do item” e a sistematização teórica da separação do que é “impacto” e o que é “viés”, vários métodos estatísticos foram aplicados na tentativa de melhor detectar o DIF. Contudo, ao final da segunda geração, observou-se que o DIF pode ocorrer de tantas maneiras que mesmo utilizando mais de um método estatístico ou mesmo um conjunto deles para sua detecção, tais esforços eram infrutíferos. A segunda geração é marcada pela perspectiva multidimensional do DIF. Ao final da segunda geração surge a pergunta eliciadora da terceira geração: “Por que o DIF está ocorrendo?” (Zumbo, 2015; 2007). Em termos de procedimentos estatísticos na segunda geração de DIF, destacam-se: 1) As respostas aos itens modeladas por tabelas de contingências ou por modelos de regressão; 2) A Teoria de Resposta ao Item; e, 3) Os modelos multidimensionais.

Na terceira geração há a busca pela transcendência dos aspectos geradores de DIF como pertencentes apenas ao teste em si, aportam-se na situação de testagem como geradora de DIF. Para a terceira geração o DIF ocorre devido a alguma característica do item e/ou da situação de testagem que não é relevante para a habilidade medida e para o objetivo do teste. Assim, considerando o teste no seu contexto sócio-político – procura-se detectar quais fatores trazem vantagens e desvantagens para os testandos levando em consideração o contexto social. A visão da terceira geração é ecológica apoiada em Bronfenbrenner (1979; 1994) – objetiva estender-se além do comportamento dos indivíduos abrangendo os sistemas funcionais e suas relações. O modelo ecológico de resposta ao item e avaliação inclui variáveis como: fatores psicológicos, cognitivos, sociológicos, estruturais, comunitários e contextuais (Zumbo, 2015; 2007). Em termos de procedimentos estatísticos na terceira geração de DIF, destacam-se: Classes latentes.

O modelo integrativo de Testagem Universal, do mesmo modo que a terceira geração preocupa-se com as variáveis psicológicas, cognitivas, sociológicas, estruturais, comunitárias e contextuais. Tanto o modelo integrativo quanto a terceira geração preconizam, o que Zumbo (2015 p. 140) descreve como “*it is important to keep in mind that we are adhering to the view that neither the test taker nor the cognitive processes in item responding are isolated in a vacuum*”. Assim, é possível tecer semelhanças e complementariedades uma vez que a

aplicação do modelo integrativo contribui para evitar-se o DIF antes da realização do instrumento. Contudo, ainda é proeminente, no modelo integrativo, o enfoque no instrumento (padronização e itens) e na sua construção semelhantemente a perspectiva da segunda geração do DIF.

Método

O presente estudo enquadra-se no âmbito dos estudos de desenvolvimento de instrumentos em Avaliação Psicológica. O objetivo da pesquisa consistiu em verificar a invariância dos parâmetros dos itens entre os indivíduos com e sem deficiência de um banco de itens de Personalidade e Testagem Universal. Este compõe um estudo mais amplo de elaboração de um banco de itens (BI) de Personalidade e Testagem Universal para Testagem Adaptativa.

Participantes

A amostra compôs-se de indivíduos maiores de idade de ambos os sexos, com e sem deficiência (n total = 1.228), sendo destes 76% do sexo feminino. Os participantes foram de 24 estados brasileiros e do Distrito Federal. Os estados com o maior número de participantes foram Santa Catarina (33%) e Paraíba (27%). Os estados com menor número de participantes foram Amapá, Acre e Amazonas, cada qual com 1 participante somente – Rondônia e Roraima foram os únicos estados sem participantes. A média de idade foi de 33,4 anos (DP=11), com valor mínimo de 18 anos e máximo de 74 anos. Quanto à escolaridade, apenas 13% possui até o Ensino Médio Completo, 70% da amostra possui Ensino Superior Incompleto e Completo ou Especialização, enquanto 16,9% mestrado ou doutorado.

Instrumentos

Construção dos Instrumentos

Os itens são de autorrelato com cinco opções de respostas (escala tipo Likert) e foram desenvolvidos segundo o modelo dos Cinco Grandes Fatores de Personalidade e os pressupostos da Testagem Universal, totalizando 525 itens. A amostra almejada tratou-se de indivíduos maiores de idade sem e com deficiência (visual e motora). Organizaram-se os itens em uma Bateria de Personalidade e Testagem Universal (BPTU) na qual foram acrescentados 40 itens com parâmetros

de qualidade já avaliada em outros estudos (itens âncoras) do Laboratório de Pesquisa em Avaliação Psicológica (LPAP/UFSC). Tal bateria foi alocada na Plataforma *Concerto* (*Open-source Online R based Adaptive Testing Platform*), a qual é uma plataforma para criação de instrumentos de código aberto que permite que os usuários criem avaliações *on-line*. Acresce-se que tais itens passaram por várias etapas de construção de itens sedimentadas na literatura científica da área (Pasquali, 2010), como a análise de juízes e ainda por avaliação dos critérios oriundos da Testagem Universal (Oliveira & Nunes, 2015; Oliveira; Nuernberg & Nunes, 2013).

Os itens foram construídos usando os métodos citados no parágrafo anterior e a pré-testagem destes foi realizada por meio do método de Blocos Incompletos Balanceados (BIB). O BIB é uma solução estatística que permite a coleta de um grande número de informações de uma quantidade extensa de itens, sendo que um mesmo testando responde a apenas uma fração destes (Fisher & Yates, 1963; Andrade, Tavares & Valle, 2000; Beckman, 2001; Moreira Junior, 2011). No modelo de BIB adotado os itens foram separados em 35 cadernos (c), 15 blocos (b), com um número de 3 (k) blocos por caderno e o número de 7 (r) repetições de cada bloco. Ainda, o fator de utilização (fu) foi de 0,2, ou seja, 20% dos testandos foi submetido a cada bloco (Fisher & Yates, 1963; Andrade, Tavares & Valle, 2000; Beckman, 2001; Moreira Junior, 2011). Assim, os 525 itens foram distribuídos no banco em 35 itens por bloco e 105 itens em cada caderno ($525/15=35$; $3*35=105$). Contudo, todos os indivíduos com deficiência responderam ao mesmo caderno, com vistas a permitir as análises posteriores mesmo que os respondentes com deficiência fossem em menor quantidade que os sem deficiência - fato este que se confirmou ao longo do estudo.

Procedimentos

As coletas foram realizadas *on-line* e o *link* para tal coleta foi amplamente divulgado por *e-mails* e redes sociais (institucionais e individuais). Ainda, contactou-se a Associação Catarinense para Integração do Cego (ACIC) e divulgou-se o *link* em listas de *e-mails* desta associação. Complementa-se que este estudo foi submetido ao comitê de ética em pesquisas com seres humanos (CEPSH/UFSC), por meio da Plataforma Brasil/DATASUS, estando aprovado e com número CAAE 51788015.6.0000.012.

Análise de Dados

Estudou-se a invariância dos parâmetros psicométricos dos itens da BPTU a partir da análise da função diferencial do item (DIF) entre os grupos de pessoas com e sem deficiência, bem como entre o sexo. Acresce-se que as análises que serão descritas subsequentemente utilizaram o modelo de Créditos Parciais de Masters, o qual é derivado ao modelo de Rasch, mas que se aplica a escalas politômicas (Wright & Masters, 1982).

Ainda, foram estudados os parâmetros dos itens. Os itens tiveram seus parâmetros estimados, separadamente, em cada um dos cinco fatores: Abertura, Extroversão, Neuroticismo, Realização e Socialização, os quais passaram por análise fatorial anteriormente a este estudo, mantendo, assim, a assunção de unidimensionalidade apropriada ao modelo de TRI escolhido (Embretson & Reise, 2000). O *software* utilizado foi o *Winsteps (Rasch Model Computer Programs)*.

Resultados/Discussão

A calibração foi realizada utilizando o Modelo de créditos Parciais de Masters e será descrita brevemente neste tópico uma vez que o enfoque deste artigo é o estudo do funcionamento diferencial do item.

O *infit* e o *outfit*, índices de desajustes dos itens, representam a presença de respostas inesperadas aos itens pela amostra pesquisada. O *infit* “*inlier-sensitive or information weighted fit*” ocorre quando respostas inesperadas são dadas a itens cujas dificuldades estão próximas à habilidade da pessoa. Já o *outfit* “*outlier-sensitive fit*” busca identificar quando respostas inesperadas são dadas a itens com dificuldade distante da habilidade da pessoa (Linacre, 2002; Nunes, Hutz & Nunes, 2010).

Os valores de *Infit* e *Outfit* e suas consequências para a medida são (Linacre, 2002): 1) $<0,5$: menos produtivo para medida, mas não degradante; 2) 0,5-1,5: Produtivo para a medida; 3) 1,5-2,0: Improdutivo para construção da medida, mas não degradante; e, 4) > 2 : Distorce ou degrada o sistema de medida. Neste estudo recomendou-se a retirada dos itens que apresentarem valores de *Infit* e/ou *Outfit* >2 já que os mesmos distorcem ou degradam a medida.

O fator Abertura, o qual foi avaliado a partir de 121 itens, a precisão – estimada pelo Alfa de Cronbach - foi de 0,76. Tal fator não apresentou nenhum valor de desajuste que distorce a medida e teve *cinco* itens com *Outfit* improdutivos (i453c13f1fc1, i245c7f1fc1, i210c6f1fc1, i385c11f1fc1 e i141c5f1fc1).

No fator Extroversão os itens apresentaram precisão – estimada pelo Alfa de Cronbach - de 0,76. Sendo que nenhum dos 102 itens de extroversão contabilizou valores de *Infit/Outfit* que distorcem a medida, apenas *um* item apresentou *Infit* e *Outfit* improdutivo (i400c12f3fc1) e *um* item apenas *Outfit* (i193c6f3fc2).

Já no fator Neuroticismo, o qual foi avaliado a partir de 120 itens, a precisão – estimada pelo Alfa de Cronbach - foi de 0,83. Em se tratando dos valores de *Infit/Outfit*, *três* itens apresentaram valores de desajustes improdutivos para a medida, sendo destes *um* de *Infit* (i119c4f4fc) e *dois* de *Outfit* (i410c12f4fc2 e i515c15f4fc2). Ainda, *três* itens apresentaram valores que distorcem a medida de *Outfit* (i15c1f4fc e i50c2f4fc) e de *Infit/Outfit* (i192c6f4fc).

Em se tratando de Realização, o qual foi avaliado a partir de 123 itens, a precisão – estimada pelo Alfa de Cronbach - foi de 0,83, nenhum item distorce a medida e *quatro* itens são improdutivos, *dois* de *Infit* e *Outfit* (i414c12f5fc1 e i490c14f5fc3) e *dois* de valores de *Outfit* (i279c8f5fc3 e i264c8f5fc).

No fator Socialização, o qual foi avaliado a partir de 99 itens, a precisão – estimada pelo Alfa de Cronbach - foi de 0,71, apenas *um* item distorce a medida (i79c3f2fc3 – *Outfit*) e *três* são improdutivos, *um* item tanto com *Infit* quanto *Outfit* (i362c11f2fc2) e *dois* com *Outfit* (S2C3_252 e i114c4f2fc1). O pequeno número de itens com desajustes que distorcem a medida nos cinco fatores estudados (apenas *cinco* itens) justifica-se devido, aos referidos itens, já terem sido previamente estudados em outros modelos de TRI e assim, outros itens com alguma espécie de desajuste já foram excluídos do banco de itens.

No mapa dos itens (Figura 1) observa-se a distribuição do nível de habilidade das pessoas (*Person*) e do nível de dificuldade dos itens (*Item*). Os primeiros representados pelo símbolo # à esquerda e os segundos por códigos que representam cada item à direita. O mapa dos itens contribui para análise dos dados, pois permite que se verifique se a amplitude da dificuldade dos itens é capaz de abarcar a distribuição da habilidade (*Theta*) da amostra pesquisada. Além de permitir a identificação de zonas de *theta* em que não se tem itens com dificuldades próximas e, ainda, regiões de *theta* em que ocorra grande concentração de itens (Nunes, Muniz, Nunes, Primi & Miguel, 2010; Lopes, 2011). No mapa de itens segue a distribuição dos itens e pessoas por fator, uma vez que se trata de um banco de itens para testagem adaptativa não há a exclusão de itens redundantes, apenas sugere-se a elaboração de itens ou a inclusão de itens já estudados em outras pesquisas do laboratório de

pesquisa em avaliação psicológica (LPAP) que abarquem os valores de *theta* não contemplados neste estudo.

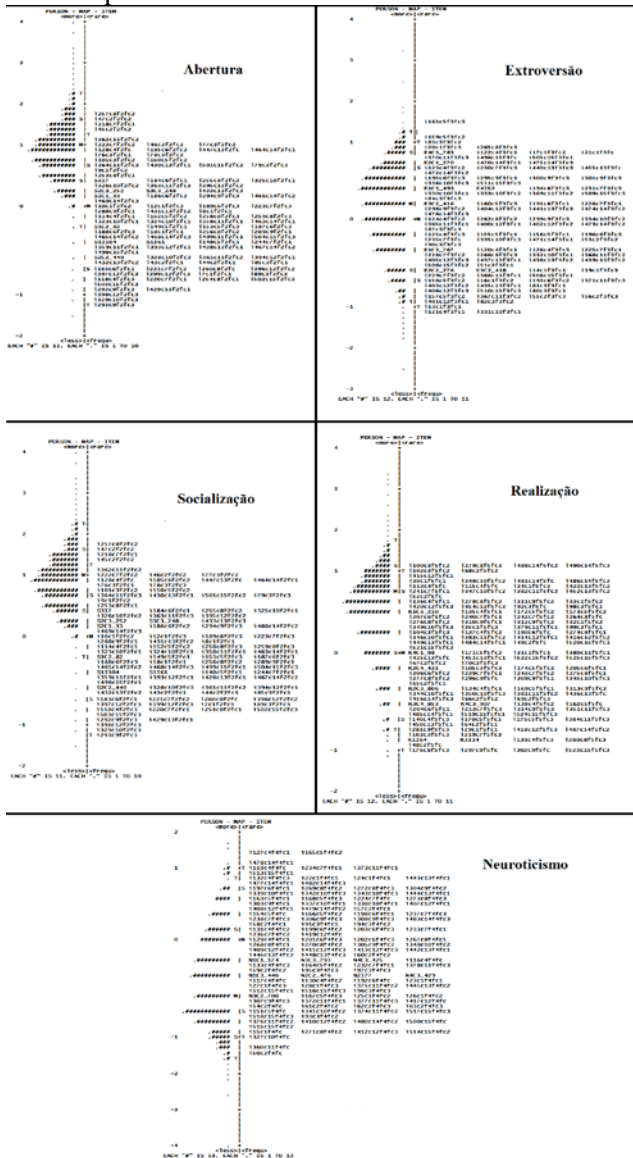


Figura 1. Mapa de itens nos cinco grandes fatores de personalidade.

Funcionamento Diferencial do item

Destaca-se que as pessoas com deficiência responderam a uma bateria fixa com 100 itens, 20 para cada fator. Tal fato decorreu da expectativa de um menor número de pessoas com deficiência como respondentes devido à dificuldade de resposta espontânea desta amostra como já discutido em outros estudos (Shinn & Ofiesh, 2012; Oliveira, 2013). Portanto, o DIF entre pessoas com e sem deficiência foi calculado apenas para esse conjunto de itens, diferentemente do DIF por sexo, que contemplou todos os itens.

No *software* utilizado na análise do DIF, o *Winsteps*, os valores dos contrastes do DIF são considerados de moderados a altos quando $|DIF| \geq 0,64$, de negligenciáveis a moderados quando $|DIF| \geq 0,43$ e $< 0,64$ e desprezíveis quando $< 0,43$ (Linacre & Wright, 1991). Considerou-se nestas análises como itens indicativos de exclusão aqueles nos quais o DIF apresentou valores de moderados a altos ($|DIF| \geq 0,64$), contudo submetem-se tais itens a reanálises de Testagem Universal – pois as mesmas já haviam sido feitas durante a construção dos itens - e quando constatado o não atendimento a algum dos seus princípios, o item também foi indicado para exclusão.

Nas análises de Testagem Universal consideram-se os princípios da TU em sua aplicação aos itens: 1. População de avaliação ampla e inclusiva: Conteúdo do item adequado para pessoas com e sem deficiência; 2. Definição precisa do construto: Simplicidade do conteúdo do item e evitar ambiguidade; 3. Itens acessíveis e não tendenciosos: Equidade para pessoas com e sem deficiência; 4. Testes flexíveis a acomodações: Texto curto (preocupação com o número de palavras). Evitou-se linguajar técnico ou rebuscado e ambiguidade; 5. Instruções e procedimentos simples, claros e intuitivos: Simplicidade e clareza nos textos dos itens; 6. Leitura agradável e de máxima inteligibilidade: Texto curto. Simplicidade do conteúdo dos itens. Evitar ambiguidade, linguajar técnico ou rebuscado; e, 7. Máxima legibilidade: Texto curto (Shinn & Ofiesh, 2012; Oliveira, 2013; Oliveira & Nunes, 2015).

No fator Abertura, para os grupos com e sem deficiência nenhum item apresentou DIF de moderado a alto e apenas um item apresentou DIF de negligenciável a moderado “*Quando falo meu corpo gesticula junto*” – mais fácil para as pessoas sem deficiência. Em se tratando da análise da Testagem Universal, reanalisando este item, observou-se que ele pode não se adequar para pessoas com deficiência motora, sobretudo nos casos de paralisias mais graves (a equidade do item está comprometida – terceiro princípio da TU), portanto excluiu-se tal item do banco. Observou-se ainda que o mesmo item já havia sido

questionado na análise fatorial quanto ao ajustamento ao fator – o qual originalmente era extroversão, fato este que corrobora para a sua exclusão. Uma hipótese possível é que o item não meça nem abertura, nem extroversão e sim algum outro construto, deste modo o problema do item não tem relação direta com o DIF, mas sim se refere a sua construção.

Em relação à ocorrência de DIF em função da variável sexo, para este mesmo fator, 12 itens apresentaram valores que variaram de mínimo a moderado e 4 de moderado a alto. Os itens de negligenciáveis a moderados não serão analisados individualmente, mas sugere-se realizar um equilíbrio no banco de itens entre os sentidos dos vetores destes itens para que não haja favorecimento de um dos grupos em detrimento do outro (DTF – Funcionamento Diferencial do Teste). Os itens que apresentaram DIF de moderado a alto foram: 1) “*Admiro a arte realizada por pessoas com deficiência.*” mais fácil para o sexo feminino; 2) “*A ideia de fazer sexo com várias pessoas ao mesmo tempo me atrai*”. Mais fácil para o sexo masculino; 3) “*Assisto desfiles de moda.*” Mais fácil para mulheres; e, 4) “*Adoro moda*”, mais fácil para mulheres. As diferenças encontradas parecem refletir a seguinte hipótese: diferenças reais entre os sexos, ou seja, mulheres com valores de *theta* menores que os homens, endossam mais facilmente itens referentes à estética e artes, já no tocante as questões sexuais os homens endossam mais facilmente com valores de *theta* menores. O questionamento a ser feito é se temos outros itens que mede as mesmas regiões de *theta*, seria útil permanecer com esses itens no banco? Sugerimos a retirada destes itens e utilizaremos no banco de itens outros tantos que estão nas mesmas regiões de *theta* e apresentam valores de DIF desprezíveis. Como pretende-se utilizar os itens desse banco em um sistema baseado em testagem adaptativa, poderiam ser utilizados outros artifícios para mantermos esses itens como, por exemplo, apresentá-los apenas para as mulheres ou aos homens ou calcularmos parâmetros diferentes como a dificuldade do item, por exemplo, para homens e mulheres.

No fator extroversão tanto no que se refere às pessoas com e sem deficiência quanto ao sexo os itens apresentaram valores de DIF desprezíveis. Já no fator neuroticismo, no tocante as pessoas com e sem deficiência, apenas um item apresentou DIF de negligenciável a moderado e os demais valores desprezíveis. O item “*Tenho esperança que vou superar meus problemas*” mostrou-se mais fácil para as pessoas sem deficiência. Uma hipótese seria que caso a pessoa com deficiência interprete a deficiência como parte de seus problemas ou causadora deles a superação pode não ser possível e sim a convivência com a deficiência,

assim sugere-se a retirada deste item apesar do DIF ser moderado, pois este item não está de acordo com o terceiro princípio da TU o qual prima pela equidade.

No DIF por sexo, ainda para neuroticismo, 4 itens indicaram valores que variaram de negligenciáveis a moderados e um item DIF de moderado a alto “*Evito festejar meus sucessos*”, o qual foi mais fácil para os homens, assim uma hipótese possível é que para os homens em um mesmo nível de *theta* que as mulheres é mais difícil festejar os sucessos, assim poder-se-ia, caso fosse optada a não exclusão, utilizar os recursos já mencionados - como apresentar esse item apenas para homens ou mulheres ou tratá-lo como um item diferente se for apresentado para homens ou mulheres (parâmetros diferentes). Neste estudo recomenda-se a exclusão do item.

Para realização (pessoas com e sem deficiência) apenas um item apresentou DIF de negligenciável a moderado, os demais apresentaram DIF desprezível – “*Sou líder no meu trabalho/estudo*”, o qual é mais fácil para as pessoas sem deficiência. Uma hipótese seria que por uma questão histórica, no passado as pessoas com deficiência não conseguiam se inserir no mercado de trabalho e, portanto, seria mais difícil, atualmente, para uma pessoa com deficiência conseguir alcançar um papel de liderança no seu trabalho. Sugere-se que esse item seja melhor estudado em novas pesquisas. Neste mesmo fator, quanto ao DIF por sexo, 9 itens apresentaram DIF de negligenciável a moderado e nenhum item apresentou valores de moderados a altos.

No fator socialização, o DIF para pessoas com e sem deficiência mostrou 1 item de negligenciável a moderado, os demais foram desprezíveis “*Sempre ofereço o meu lugar para meus amigos*”. Amigos com ou sem deficiência? Seria a pergunta a ser feita, já que se forem amigos sem deficiência o indivíduo não poderia oferecer um lugar próprio (legalmente determinado) para pessoas com deficiência, sugere-se a retirada deste item, pois ele é ambíguo não atendendo aos princípios da TU.

No DIF por sexo para o fator socialização 3 itens têm valores de negligenciáveis a moderados e 4 de moderados a altos: 1) “*Acho que criminosos deveriam perder seus direitos*”; 2) “*Seguro ou abro a porta quando estou caminhando com outra pessoa*”; 3) “*As pessoas dizem que estou sempre calmo*”; e, 4) “*Detesto quando pessoas de outros estados do Brasil vêm morar na minha cidade*” Todos foram mais endossados pelos homens. A hipótese é que todos refletem o que culturalmente é esperado do homem, agressividade, cavalheirismo e a aparência que tem

as situações sob controle “a calma”. Assim, sugere-se a exclusão desses itens do banco de itens.

Conclusões

Neste artigo a aplicação da Testagem Universal permitiu a aproximação teórica, no tocante aos aspectos qualitativos, entre terceira geração de DIF e o modelo integrativo da TU. Este estudo transcende a segunda geração do DIF por considerar vários aspectos que influenciam o processo de testagem em geral e os itens especificamente. Fatores ambientais como o uso de tecnologia assistiva são considerados importantes. Deste modo, tanto para a TU quanto para o modelo ecológico de DIF a situação de testagem é relevante e deve ser considerada desde a elaboração do instrumento até as análises estatísticas aplicadas, contudo nos aspectos estatísticos este estudo ainda permaneceu na segunda geração.

Os resultados da aplicação da Testagem Universal foram bastante promissores na busca pela invariância dos itens uma vez que poucos itens apresentaram valores de DIF de moderados a altos. Para novos estudos sugere-se a ampliação da amostra de pessoas com deficiência e a ampliação do número de itens do banco que lhes sejam aplicados. Na terceira geração de DIF são ressaltadas as limitações dos métodos estatísticos utilizados tanto na primeira como na segunda geração do DIF, assim, acresce-se a importância que sejam realizadas análises estatísticas mais sofisticadas com vistas a se aproximarem da terceira geração ou quiçá alcançá-la (métodos multivariados e a utilização do método de classes latentes).

Referências

- Andrade, D. F. Tavares, H. R. & Valle, (2000). Teoria da Resposta ao Item: Conceitos e Aplicações. São Paulo: SINAPE 2000. 154 p.
- Andriola, W. B. (2001). Descrição dos Principais Métodos para Detectar o Funcionamento Diferencial dos Itens (DIF). *Psicologia: Reflexão e Crítica*, 2001, 14(3), pp. 643-652. Recuperado em Março de 2012, de <http://goo.gl/n0IVti>.
- Bechard, S., Sheinker, J., Abell, R., Barton, K., Burling, K., Camacho, C., Cameto, R., Haertel, G., Hansen, E., Johnstone, C., Kingston, N., Murray, E., Parker, C.E., Red eld, D., and Tucker, B. (2010). Measuring Cognition of Students with Disabilities Using

- Technology-Enabled Assessments: Recommendations for a Research Agenda. *Journal of Technology, Learning, and Assessment*, 10(4). Retrieved [date] from <http://www.jtla.org>.
- Beckman, R. M. (2001). Aplicação dos Blocos Incompletos Balanceados na teoria de resposta ao item. *Estudos em Avaliação Educacional*, nº 24. Recuperado de <http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/990/990.pdf>, em Agosto de 2016.
- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. (1994). Ecological models of human development. In T. Huston & T. N. Postlethwaith (Eds.), *International encyclopedia of education*, 2nd ed., Vol. 3 (pp. 1643–1647). New York, NY: Elsevier Science. Recuperado de <http://titleiii-ptlproject.wikispaces.umb.edu/file/view/35bronfenbrenner94.pdf/455834098/35bronfenbrenner94.pdf>, em fevereiro de 2017.
- Camilli, G. & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousands Oaks, CA: Sage Publications.
- Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment*, 3(7). Recuperado em maio de 2011, de <http://www.jtla.org>.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Psychology Press.
- Fisher, S. R. A.; Yates, F. (1963). *Statistical Tables for biological, agricultural and medical research*. Sixth Edition. Recuperado de https://hekyll.services.adelaide.edu.au/dspace/bitstream/2440/10701/1/stat_tab.pdf, junho de 2015.
- Gaster, L., & Clark, C. (1995). *A guide to providing alternate formats*. West Columbia, SC: Center for Rehabilitation Technology Services. ERIC Document; No. ED 405689. Recuperado de <http://eric.ed.gov/?id=ED405689>, em Junho de 2016.
- Hanna, E. L. (2005). *Inclusive Design for Maximum Accessibility: a practical approach to Universal Design*. Pearson Educational

- Measurement. Recuperado em Março de 2011, de <http://goo.gl/PzBhdu>.
- Holland, P. W. & Thayer, D. T (1988). *Differential item performance and the Mantel-Haenszel procedure*. Em H. Wainer & H. I. Braum (Orgs.). Test Validity. Hillsdale, NJ: Lawrence Erlbaum.
- Holland, P. W. Wainer, Eds. (1994). Differential Item Functioning. *Journal of Educational Measurement*. Spring. Vol. 31, n.1. pp. 88-92. Recuperado em junho de 2014, de <http://goo.gl/4LVTFaf>.
- Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved. Recuperado em Agosto de 2011, de <http://goo.gl/jTwZLk>.
- Johnstone, C. J., Altman, J., Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Disponível em: <http://goo.gl/P4VmkG>. Acesso em: Novembro de 2011.
- Johnstone, C. J., Thompson, S. J., Moen, R. E., Bolt, S., & Kato, K. (2005). *Analyzing results of large-scale assessments to ensure universal design*. Technical Report 41. Minneapolis, MN: University of Minnesota National Center on Educational Outcomes.
- Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning and Assessment*, 4(2). Recuperado em agosto de 2011, de <http://www.jtla.org>.
- Linacre, J. M. (2002) *Differential Item Functioning and Differential Test Functioning (DIF & DTF)*. Rasch Publications. Recuperado em janeiro de 2013, de <http://www.rasch.org/rmt/rmt163g.htm>.
- Linacre, J. M., & Wright, B. D. (1991). *WINSTEPS - Rasch-Model computer programs*. Chicago: MESA Pre.
- McDonald RP. (1999). Test Theory: A Unified Treatment. Mahwah, NJ: Lawrence Erlbaum Associates.
- Merdith, W. (1993). Measurement Invariance, factor analysis and factorial invariance. *Psychometrika*: December 1993, Volume 58,

- p. 525-543. Recuperado em janeiro de 2016 de <https://link.springer.com/article/10.1007/BF02294825>.
- Moreira Júnior, F. de J.; Tezza, R.; Andrade, D. F. de; Bornia, A. C. (2013). Algoritmo de um teste adaptativo informatizado com base na teoria da resposta ao item para a estimativa da usabilidade de sites de e-commerce. *Produção*, v.23, n.3, p.525-536, jul./set. Recuperado em Junho de 2014, de <http://goo.gl/eAX17z>.
- Nunes, C. H. S. da S., Hutz, C. S. & Nunes, M. F. O. (2010). *Bateria Fatorial de Personalidade – BFP*. Manual Técnico. São Paulo: Casa do Psicólogo.
- O'Neill, K. A. & McPeck, W. M. (1993). Item and test characteristics that are associated with differential item functioning. Em P. W. Holland & H. Wainer (Orgs.). *Differential item functioning* (pp. 255-276). New Jersey: Lawrence Erlbaum. Recuperado em junho de 2014, de <http://goo.gl/Ebm3wU>.
- Oliveira, C. M. & Nunes, C. H. S. S. (2015). Instrumentos para Avaliação Psicológica de Pessoas com Deficiência Visual: Tecnologias para Desenvolvimento e Adaptação. *Psicologia: Ciência e Profissão*, 2015, 35(3), 886-899. <http://dx.doi.org/10.1590/1982-3703001902013>.
- Oliveira, C. M. (2013). Evidências de Validade de uma Bateria Informatizada para Avaliação da Personalidade adaptada ao Desenho Universal. *Dissertação de Mestrado*, Programa de Pós-graduação em Psicologia, Universidade Federal de Santa Catarina, Florianópolis, SC.
- Oliveira, C. M. ; Nuernberg, A. H. & Nunes, C. H. S. S. (2013). Desenho universal e avaliação psicológica na perspectiva dos direitos humanos. *Avaliação Psicológica*. Vol. 3. <http://goo.gl/qtkhx3>.
- Oliveri, Maria Elena; Ercikan, Kadriye & Zumbo, Bruno. (2013) Analysis of Sources of Latent Class Differential Item Functioning in International Assessments, *International Journal of Testing*, 13:3, 272-293, DOI: 10.1080/15305058.2012.738266.
- Pasquali L. (2003). *Psicometria: Teoria dos testes na Psicologia e na Educação*. Petrópolis, RJ: Vozes. p. 397.

- Pasquali, L. (2007). *Teoria de Resposta ao Item*. Brasília: LabPAM/UnB. p. 236.
- Pasquali, L. (2010). *Instrumentação Psicológica*. Porto Alegre: Artmed.
- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49-57. Recuperado em junho de 2014, de <http://goo.gl/KlfZNe>.
- Shinn, E. & Ofiesh, N. S. (2012). Cognitive Diversity and the Design of Classroom Tests for All Learners. *Journal of Postsecondary Education & Disability*. Vol. 25. Issue 3, p.227. Recuperado em outubro de 2015, de <http://files.eric.ed.gov/fulltext/EJ994288.pdf>.
- Sisto, F. F. (2006). O funcionamento diferencial dos itens. *Psico-USF*, v. 11, n. 1, p.35- 43, jan./jun. Recuperado em Março de 2012, de <http://goo.gl/NIRIUw>.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Recuperado em Agosto de 2011, de <http://goo.gl/EmWQZ3>.
- Wright, B. D. (1968). Sample-free test calibration and person measurement. Em *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton, NJ: Educational Testing Service. Recuperado em junho de 2014, de <http://www.rasch.org/memo1.htm>.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA.
- Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Going. *Language Assessment Quarterly*, 4(2), p. 223-233. Recuperado de

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.517.3293&rep=rep1&type=pdf>, em agosto de 2014.

Zumbo, Bruno D.; Yan Liu, Amery D. Wu, Benjamin R. Shear, Oscar L. Olvera Astivia & Tavinder K. Ark (2015). A Methodology for Zumbo's Third Generation DIF Analyses and the Ecology of Item Responding, *Language Assessment Quarterly*, 12:1, 136-151, DOI: 10.1080/15434303.2014.972559 To link to this article: <http://dx.doi.org/10.1080/15434303.2014.972559>.

CONSIDERAÇÕES FINAIS

Na presente tese foi detalhado o surpreendente alcance dos princípios do Desenho Universal, inicialmente elaborados com o foco em projetos de arquitetura. Os mesmos evoluíram teoricamente e mostraram a utilidade de sua aplicação em outras áreas. Na Educação, compreendeu um campo vasto com aplicações que vão desde o espaço físico até o currículo e o processo de aprendizagem. As aplicações do Desenho Universal aos testes psicológicos e ao processo de testagem foram os enfoques principais deste trabalho, os quais se tratam de um campo amplo de possibilidades, criador de uma forma diferenciada de vislumbrar a testagem para além do momento ou fração temporal em que ele ocorre, visando, assim, abarcar as inúmeras variáveis que podem interferir neste processo. Faz-se, para tanto, de fundamental importância uma metodologia de construção de testes de Testagem Universal que permitam a segurança técnica necessária a esta tarefa.

Conforme apresentado no Artigo 1, as etapas envolvidas na construção de testes de Testagem Universal são semelhantes às etapas já standardizadas na construção de instrumentos psicológicos. Acresce-se que o conteúdo de cada etapa ou suas sub etapas assumem maior riqueza e detalhamento de aspectos estruturais, sobretudo em relação ao formato e a sua adequação à utilização de tecnologias assistivas. Em se tratando da Testagem Universal, torna-se uma preocupação permanente a acessibilidade do instrumento tanto em relação ao formato quanto ao seu conteúdo, bem como a operacionalização dos princípios da Testagem Universal em cada etapa do processo de construção. Consequentemente, a Testagem Universal permite a ampliação da qualidade dos instrumentos para grupos amplos.

Esta tese evidenciou que a Testagem Universal ultrapassou a visão do testando estático e, em seus desdobramentos atuais, considerou as demandas cognitivas dos indivíduos. Tais demandas possibilitaram a ampliação do grau de detalhamento do processo de realização de testes, dividindo-o no processo de acessar e de responder ao teste, uma vez que apesar destes dois processos possuírem algumas demandas cognitivas que se assemelham, possuem inúmeras outras demandas bastante diversas. Logo, o pensar em uma população ampla e inclusiva de Testagem Universal ultrapassa o pensar apenas em nível de pessoas com e sem deficiência, mas abarca a diversidade cognitiva humana e reforça o compromisso ético dos profissionais de Psicologia com a qualidade destes materiais. Deste modo, espera-se que este trabalho contribua para ampliar o olhar em relação à elaboração de instrumentos em Psicologia e

que a Testagem Universal alcance não só os instrumentos, mas as atitudes e o pensamento dos profissionais de Psicologia, os quais incluem estudos desta natureza quando do processo de desenvolvimento de seus instrumentais.

Na construção do banco de itens para CAT segundo os princípios da Testagem Universal e medindo o construto personalidade (Artigo 2) os princípios da Testagem Universal foram aplicados na elaboração dos itens, em seus conteúdos, preocupando-se com a simplicidade, evitando ambiguidade e primando pela equidade. Igualmente, aplicou-se tais princípios ao instrumento para calibração do banco de itens (Bateria de Personalidade e Testagem Universal) no formato dos itens, nos *Templates* elaborados, cores, tamanho de fonte e no texto das instruções. Tais medidas visaram a ampliar a qualidade do teste para a utilização por pessoas com e sem deficiência abarcando a diversidade cognitiva humana – embora, no presente caso, ainda parcialmente restrito para pessoas sem deficiência e com deficiência visual e motora.

Assim, a Testagem Universal busca agregar as melhores práticas no que se refere aos formatos, as estruturas, aos tipos de mensuração e a avaliação dos instrumentos. Coaduna-se, portanto, plenamente com a testagem adaptativa informatizada, primando pela qualidade do banco de itens e de toda a tecnologia utilizada durante a construção do mesmo.

Na presente pesquisa foram tomadas decisões qualitativas embasadas na literatura científica no que tange à construção de instrumentos de forma ampla e de instrumentos de testagem adaptativa e, mais especificamente, de Testagem Universal. Acresce-se que na construção de testes psicológicos de Testagem Universal, faz-se imprescindível o atendimento aos pressupostos teóricos e técnicos inerentes ao processo de construção de instrumentos sedimentados na literatura científica da área.

Algumas limitações encontradas na aplicação dos princípios da Testagem Universal foram que em se tratando do primeiro princípio (população ampla e inclusiva) os itens foram desenvolvidos apenas para indivíduos com deficiência visual (baixa visão e cegueira) e deficiência motora, não abarcando um espectro maior de deficiências e suas comorbidades. No tocante ao referencial da diversidade cognitiva humana, não foi abarcado no planejamento dos itens indivíduos com TDAH (Transtorno de Déficit de Atenção e Hiperatividade), déficit cognitivo, dificuldades de aprendizagem, depressão, entre outros.

Ainda no âmbito das limitações, uma dificuldade encontrada foi a de não possuir na equipe de pesquisa pelo menos um especialista em

informática e acessibilidade. Tal ausência foi parcialmente suprida pela equipe, porém culminou por limitar as possibilidades de recursos de tecnologias assistivas testados, restringindo-se apenas aos leitores de tela. Complementarmente, os recursos de acessibilidade que a equipe de pesquisa foi capaz de implementar na plataforma *Concerto* foram reduzidos. Sugere-se para novos estudos a presença de um profissional da informática que seja especialista em acessibilidade.

Além disto, mais ainda no âmbito das dificuldades e limitações, destaca-se que os recursos financeiros foram reduzidos. Para minimizar tais efeitos tentou-se utilizar programas já adquiridos pelo LPAP e pelos pesquisadores envolvidos no desenvolvimento do instrumento. Porém, tem-se a convicção que com a disponibilidade de mais recursos financeiros a pesquisa teria avançado com mais qualidade e em menos tempo – as vantagens seriam tanto no tocante as possibilidades de recursos tecnológicos para implementação do desenho universal quanto no investimento na equipe de pesquisa (por exemplo: cursos, capacitações, presença de um programador na equipe).

Uma dificuldade técnica proeminente que prolongou o tempo estimado para as análises estatísticas foi a de adequar os *softwares* de que dispúnhamos e as análises que havíamos planejado a resultados altamente complexos e desafiadores, tais como operar com um grande número de itens e uma amostra bastante diversificada – pessoas com e sem deficiência. Adiciona-se a estes fatos que a modelagem dos dados no BIB escolhido, que permite a coleta de um grande número de informações durante o processo de calibração dos itens, apresenta como consequência um grande número de *missings* requerendo soluções estatisticamente ainda mais robustas. Em novos estudos sugere-se a aplicação de outros recursos estatísticos e *softwares* (como, por exemplo, a análise de equações estruturais e *softwares* de TRI multivariada).

Como aspecto bastante positivo destaca-se o delineamento feito para a análise de juízes que permitiu, apesar do extenso número de itens, que os juízes realizassem a tarefa com qualidade. Assim sendo, a construção de um banco de itens de personalidade e Testagem Universal em testagem adaptativa apresenta-se como uma tarefa de alta complexidade que demanda várias etapas com delineamentos diversos e amplos recursos técnicos, tecnológicos e, sobretudo, metodológicos.

No Artigo 3 foi relatado o estudo da dimensionalidade dos itens. A partir da análise fatorial de um número amplo de itens, 525, foi possível resumi-los em 5 fatores - Abertura, Realização, Extroversão, Socialização e Neuroticismo. O Banco de Itens desenvolvido

demonstrou em seus 15 cadernos a viabilidade de uma solução de cinco fatores, bem como que os itens criados para cada fator relacionaram-se à dimensão para a qual foram construídos. Tal fato demonstra a consistência teórica do modelo e a qualidade da correspondência da mesma com os itens desenvolvidos para cada um dos fatores. Portanto, a análise fatorial acumulou indícios da replicabilidade do modelo dos Cinco Grandes Fatores para a amostra maior de idade. Em novos estudos pretende-se ampliar a amostra em cada um dos 15 cadernos e realizar o estudo da dimensionalidade em nível de facetas.

No Artigo 4 foi realizado o estudo de calibração dos itens, o qual possibilitou a análise dos parâmetros dos itens no modelo de dois parâmetros. Após estas análises concluiu-se que 317 itens estavam aptos a compor o Banco de Itens de Personalidade e Testagem Universal – Abertura (64 itens); Extroversão (58 itens); Neuroticismo (82 itens); Realização (76 itens); e Socialização (37 itens) – ou seja, com parâmetros adequados, sobretudo os valores de a , discriminação, os quais refletem a qualidade do item em medir o construto na faixa de sua dificuldade. Foi sugerido que alguns itens fossem utilizados em novos estudos por serem mais intensos que a amostra de aplicação e, talvez, por este motivo não terem sido calibrados.

A perda de 208 itens (39,6%) do Banco foi considerada alta já que alguns autores da área apontam uma perda em torno de 20%. Sugere-se, para novos estudos a comparação dos melhores itens com os demais a fim de tecer-se o mapa eficaz das perdas de itens e tentar-se minimizá-las. Complementarmente, acresce-se que alguns dos itens que não calibraram apresentaram baixa frequência e respostas nas duas primeiras categorias da escala de cinco pontos implementadas, assim, supõem-se ser interessante verificar a transformação de tal escala em outra de quatro pontos.

Já no Artigo 5 – o último desta tese – a aplicação da Testagem Universal permitiu a aproximação teórica, no tocante aos aspectos qualitativos, entre terceira geração de DIF e o modelo integrativo da TU. Este estudo transcendeu a segunda geração do DIF por considerar vários aspectos que influenciam o processo de testagem em geral e os itens especificamente. Fatores ambientais como o uso de tecnologia assistiva são considerados importantes. Deste modo, tanto para a TU quanto para o modelo ecológico de DIF, a situação de testagem é relevante e deve ser considerada desde a elaboração do instrumento até as análises estatísticas aplicadas, mas nos aspectos estatísticos as análises de DIF perpetradas nessa tese ainda permaneceram na segunda geração descrita por Zumbo.

Acresce-se que construtos alheios ao medido comprometem a validade do teste. Quando isto acontece é comum que o DIF seja significativo – pelo menos em alguns itens. Assim, a análise de DIF contribuiu para a busca de evidências de validade através de consistência interna, uma vez que a presença de viés nos testes torna-os injustos para determinados grupos comprometendo a equidade na avaliação.

Os resultados da aplicação da Testagem Universal foram bastante promissores na busca pela invariância dos itens uma vez que poucos itens apresentaram valores de DIF de moderados a altos. Ou seja, os resultados indicam que dimensões de construtos não relacionados decorrentes da deficiência não afetam de forma significativa o construto medido em uma parcela bastante relevante dos itens utilizados nesta pesquisa.

Para novos estudos sugere-se a ampliação da amostra de pessoas com deficiência e a ampliação do número de itens do banco que lhes sejam aplicados. Na terceira geração de DIF são ressaltadas as limitações dos métodos estatísticos utilizados tanto na primeira como na segunda geração do DIF, assim, destaca-se a importância de que sejam realizadas análises estatísticas mais sofisticadas com vistas a se aproximarem da terceira geração ou, quiçá, alcançá-la (métodos multivariados e a utilização do método de classes latentes).

Sugere-se para pesquisas futuras a equiparação dos itens prontos para o banco de itens com outros itens já existentes no banco de itens de personalidade do LPAP e o desenvolvimento do protótipo de um teste psicológico de personalidade informatizado de testagem adaptativa aplicando os princípios da Testagem Universal. Ainda, seria interessante perpetrarem-se estudos para verificação da associação entre os diferentes grupos, com e sem deficiência, e o atendimento aos princípios da Testagem Universal no instrumento, e, por fim, a elaboração de um manual norteador para a criação de itens/testes de Testagem Universal.

REFERÊNCIAS

- Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2010). Technology-Enabled and Universally Designed Assessment: Considering Access in Measuring the Achievement of Students with Disabilities—A Foundation for Research. *Journal of Technology, Learning, and Assessment*, 10(5). Retrieved [date] from <http://www.jtla.org>.
- Alves, G. A. da S., Souza, M. S. & Baptista, M. N. (2011). Validade e precisão de testes psicológicos. In R. A. M. Ambiel (et al.), *Avaliação Psicológica: Guia de consulta para estudantes e profissionais de Psicologia* (pp. 109-128). São Paulo: Casa do Psicólogo.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Amiralian, M. L. T. M. (1997). Compreendendo o cego: uma visão psicanalítica da cegueira por meio de desenhos-estórias. São Paulo: Casa do Psicólogo. 322p.
- Amorim, M. L. de C. (2006). Construção e Adaptação de um teste de Atenção para indivíduos com deficiência visual: Estudo baseado no Teste de Atenção de Bams. Dissertação. Mestrado em Ciência do Desporto na Área de Especialização em Actividade Física Adaptada. Porto: Faculdade de Desporto da Universidade do Porto. Recuperado em março de 2011, de <http://repositorio-aberto.up.pt/handle/10216/14112>.
- Anastasi, A. & Urbina, S. (2000). *Testagem psicológica* (7ª ed.). Porto Alegre: Artes Médicas.
- Andrade, D. F. Tavares, H. R. & Valle, (2000). Teoria da Resposta ao Item: Conceitos e Aplicações. São Paulo: SINAPE 2000. 154 p.
- Andrade, F. S. de (2008). Aplicação das tecnologias da informação e comunicação em bibliotecas universitárias como recursos auxiliares aos deficientes visuais. Recuperado em julho de 2011, de <http://goo.gl/4JWL8>.

- Andrade, J. M. de. (2008). *Evidências de Validade do Inventário dos Cinco Grandes Fatores de Personalidade para o Brasil*. Tese de Doutorado. Universidade de Brasília. Pós-graduação em Psicologia Social, do trabalho e das organizações. Recuperado em Outubro de 2011, de <http://goo.gl/ApqXo6>.
- Andriola, W. B. (2001). Descrição dos Principais Métodos para Detectar o Funcionamento Diferencial dos Itens (DIF). *Psicologia: Reflexão e Crítica*, 2001, 14(3), pp. 643-652. Recuperado em Março de 2012, de <http://goo.gl/n0IVti>.
- Araújo, R. (2010). *Relação entre bem-estar subjetivo e o fator realização no modelo dos cinco grandes fatores de personalidade*. Recuperado em Outubro de 2011, de <http://goo.gl/tqIgjG>.
- Associação Brasileira de Normas Técnicas, ABNT (2004). *Acessibilidade a edificações, mobiliário, espaços e equipamentos urbanos*. NBR 9050/2004.
- Baptista, N. J. M. (2008). *Teorias da Personalidade*. Recuperado em Setembro de 2011, de www.Psicologia.com.pt.
- Bartholomeu, D. (2008). Traços de Personalidade e Comportamentos de risco no trânsito: um estudo correlacional. *Psicologia Argumento*. Curitiba: jul./set., 26(54), p.193-206. Recuperado em Outubro de 2011, de <http://goo.gl/W6Wb4k>.
- Bechard, S., Sheinker, J., Abell, R., Barton, K., Burling, K., Camacho, C., Cameto, R., Haertel, G., Hansen, E., Johnstone, C., Kingston, N., Murray, E., Parker, C.E., Red eld, D., and Tucker, B. (2010). Measuring Cognition of Students with Disabilities Using Technology-Enabled Assessments: Recommendations for a Research Agenda. *Journal of Technology, Learning, and Assessment*, 10(4). Retrieved [date] from <http://www.jtla.org>.
- Brasil (2006)a. SEESP/MEC. - Brasília : MEC, Secretaria de Educação Especial . *Saberes e práticas da inclusão: desenvolvendo competências para o atendimento às necessidades educacionais especiais de alunos surdos*. 2. ed. 116 p. Recuperado em junho de 2014, de <http://goo.gl/4KrzVT>.
- Brasil (2006)b. SEESP/MEC. - Brasília: MEC, Secretaria de Educação Especial. *Saberes e práticas da inclusão: desenvolvendo competências para o atendimento às necessidades educacionais*

- de alunos com deficiência física/neuro-motora*. 2. Ed. 36 p. Recuperado em junho de 2014, de <http://goo.gl/9gkVxv>.
- Brasil (2013). *Edital nº 01, de 08 de maio de 2013*. Exame nacional do ensino médio. Brasília: Instituto nacional de estudos e pesquisas educacionais Anísio Teixeira. Ministério da educação. Recuperado em dezembro de 2013, de <http://goo.gl/XDbQqw>.
- Burgstahler, S. (2001). *Equal Access: Universal Design of Instruction*. University of Washington. DO-IT (Disabilities, Opportunities, Internetworking, and Technology). Recuperado em Agosto de 2011 de, <http://goo.gl/MYLjUN>.
- Camilli, G. & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousands Oaks, CA: Sage Publications.
- Case, B. J. (2003). *Universal Design*. Pearson: policy report. Recuperado em março de 2011, de www.pearsonassessments.com/UniversalDesig.
- Case, B. J., Zucker, S., & Jeffries, J. L. (2005). *A primer on Assessing the visually impaired*. Pearson Education. Recuperado em março de 2011, de <http://goo.gl/RU9UiM>.
- Castro, S. M. de J., Trentini, C., & Riboldi, J. (2010). Teoria da resposta ao item aplicada ao Inventário de Depressão Beck. *Rev Bras Epidemiol*. 13(3): 487-501. Recuperado em Fevereiro de 2014, de <http://goo.gl/geBIvp>.
- Castro, S. M. de J., Trentini, C., & Riboldi, J. (2010). Teoria da resposta ao item aplicada ao Inventário de Depressão Beck. *Rev Bras Epidemiol*. 13(3): 487-501. Recuperado em Fevereiro de 2014, de <http://goo.gl/kqhCfs>.
- Conde, A. J. M. (2005). *Definindo a cegueira e a visão subnormal*. Instituto Benjamin Constant. Recuperado em julho de 2011, de <http://goo.gl/SJrlez>.
- Conselho Federal de Psicologia - CFP, Sistema de Avaliação de testes psicológicos. (2011). *Testes psicológicos aprovados para uso*. Recuperado em março de 2011, de <http://goo.gl/VKb6SD>.
- Cônsolo, A. T. (2012). *Efeitos do computador, da internet e do celular na comunicação escrita entre surdos*. Dissertação (mestrado). Pontifícia Universidade Católica de São Paulo (PUC-SP). 123 p. Recuperado em junho de 2014, de <http://goo.gl/ObVW7N>.

- Costa, D. R. (2009). Métodos Estatísticos em Testes Adaptativos Informatizados. *Dissertação*. Instituto de Matemática – Departamento de Métodos Estatísticos da Universidade Federal do Rio de Janeiro, UFRJ. Recuperado em Novembro de 2012, de <http://goo.gl/2Ufu3T>.
- Dolan, R. P. & Hall, T. E. (2001). Universal Design for Learning: Implications for Large-Scale Assessment. *IDA Perspectives* 27(4): 22-25. Recuperado em maio de 2011, de <http://goo.gl/aO7GW6>.
- Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment*, 3(7). Recuperado em maio de 2011, de <http://www.jtla.org>
- Eggen, T. J. H. M. & Verschoor, A. J. (2006). Optimal Testing With Easy or Difficult Items in Computerized Adaptive Testing. *Applied Psychological Measurement*, 30(5) 379-393. Recuperado em novembro de 2012, de <http://goo.gl/YOoDK3>.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Psychology Press.
- Fadiman, J. & Frager, R. (2002). *Teorias da Personalidade*. São Paulo: Harbra. 393p.
- Galvão Filho, T. A. A. (2009). Tecnologia Assistiva: de que se trata? In: MACHADO, G. J. C.; SOBRAL, M. N. (Orgs.). *Conexões: educação, comunicação, inclusão e interculturalidade*. 1 ed. Porto Alegre: Redes Editora, p. 207-235, 2009.
- Governo do Estado de São Paulo, Secretaria de estado da habitação & Secretaria de estado dos direitos da pessoa com deficiência (2010). *Desenho Universal: habitação de interesse social*. Recuperado em março de 2011, de <http://goo.gl/SN8GLr>.
- Gray, D. E. (2012). *Pesquisa no mundo real*. 2ed. Porto Alegre: Penso. 488 p.
- Guzzo, R. S. L.; Pinho, C. C. M. de; Carvalho, C. F. de (2002). Construção da Taxonomia para Descritores da Personalidade.

- Psicologia: Reflexão e Crítica*. Porto Alegre, 15(1), pp. 71-75. Recuperado em Outubro de 2011, de <http://goo.gl/WmCdB2>.
- Hanna, E. L. (2005). *Inclusive Design for Maximum Accessibility: a practical approach to Universal Design*. Pearson Educational Measurement. Recuperado em Março de 2011, de <http://goo.gl/PzBhdu>.
- Holland, P. W. & Thayer, D. T (1988). *Differential item performance and the Mantel-Haenszel procedure*. Em H. Wainer & H. I. Braum (Orgs.). *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum.
- Holland, P. W. Wainer, Eds. (1994). *Differential Item Functioning. Journal of Educational Measurement*. Spring. Vol. 31, n.1. pp. 88-92. Recuperado em junho de 2014, de <http://goo.gl/4LVTaf>.
- Hutz, C. S.; Nunes, C. H.; Silveira, A. D.; Serra, J.; Anton, M.; Wieczorek, L. S. (1998). O desenvolvimento de marcadores para a avaliação da personalidade no modelo dos cinco grandes fatores. *Psicologia Reflexão e Crítica*. Porto Alegre: Vol. 11, n.2. Recuperado em Outubro de 2011, de <http://goo.gl/xmyDtm>.
- Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal design and student performance* (Technical Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved. Recuperado em Agosto de 2011, de <http://goo.gl/jTwZLk>.
- Johnstone, C. J., Altman, J., Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Disponível em: <http://goo.gl/P4VmkG>. Acesso em: Novembro de 2011.
- Johnstone, C. J., Thompson, S. J., Moen, R. E., Bolt, S., & Kato, K. (2005). *Analyzing results of large-scale assessments to ensure universal design*. Technical Report 41. Minneapolis, MN: University of Minnesota National Center on Educational Outcomes.
- Kastrup, V., Sampaio, E., Almeida, M.C. de & Carijó, F.H. (2009). O aprendizado da utilização da substituição sensorial visuo-tátil por pessoas com deficiência visual: primeiras experiências e estratégias metodológicas. Recuperado em março de 2011, de <http://goo.gl/5NRW3q>.

- Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning and Assessment*, 4(2). Recuperado em agosto de 2011, de <http://www.jtla.org>
- Marconi, M. de A. Lakato, E. M. (2008). *Fundamentos de metodologia científica*. 6 ed. São Paulo: Atlas. 315 p.
- McCrae, R. R. & Oliver, P. J. (1990). *An Introduction to the Five-Factor Model and its applications*. University of California at Berkeley. Recuperado em Outubro de 2011, de <http://goo.gl/StpU2L>.
- McCrae, R.R. & Terracciano, A. (2004). Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. *Journal of Personality and Social Psychology*, 2005, Vol. 88, No. 3, 547-561. Recuperado em Março de 2014, de <http://goo.gl/n66SWy>.
- Moreira Júnior, F. de J.; Tezza, R.; Andrade, D. F. de; Bornia, A. C. (2013). Algoritmo de um teste adaptativo informatizado com base na teoria da resposta ao item para a estimativa da usabilidade de sites de e-commerce. *Produção*, v.23, n.3, p.525-536, jul./set. Recuperado em Junho de 2014, de <http://goo.gl/eAXI7z>.
- Motti, F. G. Francelin, M. A. S. & Morita, I. (2010). As Implicações Sociais da Deficiência Auditiva Adquirida em Adultos. *Saúde Soc.* São Paulo, v.19, n.1, p.180-192, 2010. Recuperado em junho de 2014, de <http://goo.gl/Eq6tpZ>.
- Muniz, J. (2004). La validación de los tests. *Metodología de las Ciencias del Comportamiento*, Vol. 5, p. 121-141. Recuperado em junho de 2014, de [http://alfama.sim.ucm.es/wwwisis2/wwwisis.exe/\[in=psyke2.in\]/?format=breve&boolean=%5BNR:15759105](http://alfama.sim.ucm.es/wwwisis2/wwwisis.exe/[in=psyke2.in]/?format=breve&boolean=%5BNR:15759105).
- Nunes, C. H. S. da S. (2005). *Construção, normatização e validação das escalas de socialização e Extroversão no modelo dos cinco grandes fatores*. Tese de doutorado, Curso de Pós-graduação em Psicologia do Desenvolvimento. Universidade Federal do Rio grande do Sul. Recuperado em Outubro de 2011, de <http://goo.gl/Uudr0l>.

- Nunes, C. H. S. da S., Hutz, C. S. & Nunes, M. F. O. (2010). *Bateria Fatorial de Personalidade – BFP*. Manual Técnico. São Paulo: Casa do Psicólogo.
- Nunes, C. H. S., Hutz, C. S. & Giacomoni, C. H. (2009). Associação entre bem estar subjetivo e personalidade no Modelo dos Cinco Grandes Fatores. *Avaliação Psicológica*. Porto Alegre: vol. 8, pp. 99-108. Recuperado em Outubro de 2011, de <http://goo.gl/H5udmi>.
- O'Neill, K. A. & McPeck, W. M. (1993). Item and test characteristics that are associated with differential item functioning. Em P. W. Holland & H. Wainer (Orgs.). *Differential item functioning* (pp. 255-276). New Jersey: Lawrence Erlbaum. Recuperado em junho de 2014, de <http://goo.gl/Ebm3wU>
- Oliveira, C. M. (2013). Evidências de Validade de uma Bateria Informatizada para Avaliação da Personalidade adaptada ao Desenho Universal. *Dissertação de Mestrado*, Programa de Pós-graduação em Psicologia, Universidade Federal de Santa Catarina, Florianópolis, SC.
- Oliveira, C. M. ; Nuernberg, A. H. & Nunes, C. H. S. S. (2013). Desenho universal e avaliação psicológica na perspectiva dos direitos humanos. *Avaliação Psicológica*. Vol. 3. <http://goo.gl/qtkhx3>.
- Paim, M. C. C., Rossi, D. & Richter, D. S. (2008). Temperamento e traços de personalidade: uma abordagem. *Revista Digital: efdeportes*. Buenos Aires: año 13, n. 124.
- Pasquali, L. (2000). *Os Tipos Humanos: a teoria da personalidade*. CopyMarket.com. Recuperado em Setembro de 2011, de <http://goo.gl/5RxLui>.
- _____. (2003). *Psicometria: teoria dos testes na Psicologia e na educação*. Petrópolis, RJ: Vozes. p.397.
- _____. (2007). *Teoria de Resposta ao Item*. Brasília: LabPAM/UnB. p. 236.
- Pietro, G. (2010). Testes Informatizados. In: Pasquali, L. (2010). *Instrumentação Psicológica*. Porto Alegre: Artmed.

- Prado, A. R. de A., Lopes, M. E. & Ornstein, S. W. (orgs.) (2010). *Desenho Universal: caminhos da acessibilidade no Brasil*. São Paulo: Annablume. 306p.
- Primi, R. (2010). Psicologia: Teoria e Pesquisa. Avaliação Psicológica no Brasil: Fundamentos, Situação Atual e Direções para o Futuro. *Psicologia: Teoria e Pesquisa*. Vol. 26 n. especial, pp. 25-3. Recuperado em março de 2012, de <http://goo.gl/uCa3Uh>.
- Primi, R., Muniz, M. & Nunes, C.H.S.S. (2009). Definições contemporâneas de validade de testes psicológicos. In C. S. Hurtz (Org.), *Avanços e polêmicas em avaliação psicológica* (pp. 243-265). São Paulo: Casa do Psicólogo.
- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49-57. Recuperado em junho de 2014, de <http://goo.gl/KlfZNe>.
- Sá, E. D. de, Campos, I.M. de & Silva, M.B.C. (2007) Atendimento Educacional Especializado: Deficiente Visual. Brasília-DF: MEC. Recuperado em maio de 2011, de <http://goo.gl/FqwVsR>.
- Schultz, D. P. & Schultz, S. E. (2002). *Teorias da Personalidade*. São Paulo: Pioneira. 528 p.
- Silva, I. B. & Nakano, T. de C. (2011). Modelo dos Cinco Grandes Fatores da Personalidade: Análise de Pesquisas. *Avaliação Psicológica*, 2011, 10(1), pp. 51-62. Recuperado em Março de 2014, em <http://goo.gl/mxp7EP>.
- Silva, R. S., Schlottfeldt, C. G., Rozenberg, M. P., Santos, M. T. & Lelé, A. J. (2007). *Replicabilidade do Modelo dos Cinco Grandes Fatores em medidas da personalidade*. Mosaico: Estudos em Psicologia. Vol. I, n.1, 37-49. Recuperado em Outubro de 2011, de <http://goo.gl/qRbDWx>.
- Sisto, F. F. (2006). O funcionamento diferencial dos itens. *Psico-USF*, v. 11, n. 1, p.35- 43, jan./jun. Recuperado em Março de 2012, de <http://goo.gl/NIRIUw>.
- Story, M. F., Mueller, L. J. & Mace, R. L. (1998). The Universal Design File: Designing for People of all Ages and Abilities. *Design Research and Methods Journal*. Recuperado em Agosto de 2011, de http://www.ncsu.edu/ncsu/design/cud/pubs_p/pudfiletoc.htm.

- Story, M. F. (2001). Principles of Universal Design. In: Preiser, W. F. E. & Ostroff, E. *Universal Design Handbook*. McGraw-Hill Professional. Recuperado em Agosto de 2011, de <http://goo.gl/GvnFrP>.
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Recuperado em Agosto de 2011, de <http://goo.gl/EmWQZ3>.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.
- Vasconcellos, S. J. L. (2007). *Os pressupostos evolutivos dos fatores Abertura à experiência e Neuroticismo*. Tese de doutorado. Programa de Pós-graduação em Psicologia. Porto Alegre. Recuperado em Outubro de 2011, de <http://goo.gl/ukRYTM>.
- Wright, B. D. (1968). Sample-free test calibration and person measurement. Em Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service. Recuperado em junho de 2014, de <http://www.rasch.org/memo1.htm>.