

PROPRIEDADES PSICOMÉTRICAS DOS ITENS DO TESTE WISC-III¹

Vera Lúcia Marques de Figueiredo^{*}
Viviane Leite Dias de Mattos[#]
Luis Pasquali[†]
Aline Poester Freire^æ

RESUMO. O aperfeiçoamento de um teste se dá através da seleção, substituição ou revisão de itens, e quando um item é analisado, aumenta a validade e precisão do teste. Este artigo trata da apresentação dos resultados relativos às propriedades psicométricas dos itens dos subtestes do WISC-III, referentes a dificuldade, discriminação e validade. O WISC-III é um instrumento amplamente utilizado no contexto da avaliação da inteligência, e conhecer a qualidade dos itens é essencial ao profissional que administra o teste. As análises foram efetuadas com base nas pontuações de 801 protocolos do teste, aplicados por ocasião da pesquisa de adaptação a um contexto brasileiro. As análises mostraram que os itens adaptados apresentaram características psicométricas adequadas, possibilitando a utilização do instrumento como meio confiável de diagnóstico.

Palavras-chave: WISC-III, análise de itens, validade de itens.

PSYCHOMETRIC PROPERTIES OF WISC-III ITEMS

ABSTRACT. The improvement of the quality of items by selection, substitution and review will increase a test's validity and reliability. Current essay will present results referring to psychometric properties of WISC-III items, specifically difficulty, discrimination and validity of items. Since WISC-III is a widely used test for the assessment of intelligence, knowledge on the quality of items of the test seems to be essential for professionals using it. Analyses were performed on data 801 protocols obtained in the adaptation research of a test within a Brazilian context. Analyses showed that the adapted items presented adequate psychometric features which make it a reliable diagnostic tool.

Key words: WISC-III, items analysis, validity of items.

PROPIEDADES PSICOMÉTRICAS DE LOS ÍTENES DEL SUBTEST WISC-III

RESUMEN. El perfeccionamiento de un teste ocurre por la selección, sustitución o revisión de ítenes y, cuando un item es analizado, aumenta la validez y fiabilidad del teste. Ese artículo trata de la presentación de los resultados relativos a las propiedades psicométricas de los ítenes del subtest WISC-III, referentes a la dificultad, a la discriminación y a la validez. El WISC-III es un instrumento muy utilizado en el contexto de la evaluación de la inteligencia, y conocer a la calidad de los ítenes es esencial al profesional que administra el teste. Los análisis fueron efectuados con base en los puntajes de 801 protocolos de registro del teste, aplicados por ocasión de encuesta de estandarización a un contexto brasileño. Los análisis enseñaron que los ítenes adaptados apuntaron características psicométricas adecuadas, permitiendo la utilización del instrumento como medio confiable de diagnóstico.

Palabras-clave: WISC-III, análisis de ítenes, validez de ítenes.

Todo teste psicológico, para produzir resultados confiáveis, deve evidenciar sua validade e fidedignidade, parâmetros psicométricos essenciais que determinam a qualidade do instrumento. Tais

¹ Apoio: CNPq.

^{*} Doutora em Psicologia. Docente do Curso de Psicologia da Universidade Católica de Pelotas.

[#] Doutora em Engenharia de Produção. Docente do Curso de Educação da Universidade Católica de Pelotas.

[†] Doutor em Psicologia. Docente do Instituto de Psicologia da Universidade de Brasília.

^æ Acadêmica do Curso de Psicologia da Universidade Católica de Pelotas, bolsista PIBIC-CNPq.

requisitos dependem, em última análise, das características dos seus itens, tanto em relação ao seu conteúdo como em relação a propriedades estatísticas como dificuldade e discriminação.

O aperfeiçoamento de um instrumento se dá quando é feita uma revisão ou substituição de itens. Quando um item é analisado e constata-se que não possui valor, é preciso eliminá-lo ou refazê-lo. Às vezes, é necessário reduzir o número de itens do teste para que a sua validade e precisão aumentem (Anastasi & Urbina, 2000). Para as autoras, a análise de itens assegura a confiança do examinando no início do teste, além de reduzir o desperdício de tempo com itens desnecessários.

Segundo Guilford (1954), o principal objetivo de uma análise de itens é obter informações objetivas sobre os itens apresentados em um teste, o que é válido por vários motivos, entre os quais, permitir checar a subjetividade do autor ao formulá-lo. Os itens devem ser escolhidos por adequado nível médio e variabilidade de dificuldade e grau de intercorrelação. Pasquali (2003) ressalta a relevância de avaliar também a validade.

DIFICULDADE DOS ITENS

A razão principal para medir a dificuldade de itens é a de descartar aqueles que possuem um nível impróprio de dificuldade, não informando diferenças individuais nem proporcionando variabilidade aos resultados de um teste. Para Thordike, Cunningham, Thordike e Hagen citados por Hetzell (1997), a dificuldade do item tem efeito direto tanto na variabilidade dos escores do teste quanto na precisão com que seus escores discriminam grupos diferentes de examinandos. O índice de dificuldade propõe-se a medir as diferenças individuais no que diz respeito ao rendimento alcançado no teste, podendo ser calculado em testes de aptidão, inteligência e conhecimento (Erthal, 1987).

A psicometria tradicional tem tratado da dificuldade do item em testes de aptidão definindo-a como a razão (p) entre a frequência de acertos no item e a quantidade de sujeitos da amostra; assim, quanto menor o índice, maior a dificuldade. Dessa forma, Bunchaft e Cavas (2002) e Pasquali (2003) o denominam de índice de facilidade, pois quanto mais sujeitos acertam o item, mais fácil ele é.

Na construção de um teste de rendimento, pode-se usar o critério proposto por Cerdá citado por Erthal (1987) e Pasquali (1996), que determina proporções adequadas de itens conforme o grau de dificuldade.

Para ele, o objetivo de conhecer o nível de dificuldade é definir a amostra de itens que irá compor o teste.

Todos os testes devem incluir um pequeno número de itens, que ademais devem ser muito fáceis, com vista a treinar os sujeitos na tarefa e aumentar sua motivação e seu grau de cooperação. Em geral, estes itens são arranjados em ordem de dificuldade, de modo que os testandos comecem com itens relativamente fáceis e avancem para itens com dificuldade crescente. Este arranjo permite mais confiança aos testandos e reduz a probabilidade de perderem tempo com itens que ultrapassam sua habilidade (Anastasi & Urbina, 2000).

Segundo McIntire e Miller (2000), itens com nível de dificuldade igual a 0,5 produzem distribuição de escores com mais variabilidade. Por essa razão, os elaboradores de teste descartam ou reelaboram itens com dificuldade entre 0 e 0,2 (muito difícil) ou entre 0,8 e 1,00 (muito fácil). Enryssen citado por Adonez (1999), também aconselha selecionar itens com níveis de dificuldade entre 0,4 e 0,6.

Para Adonez (1999), um item contribui para diferenciar os sujeitos se tiver um $p = 0,5$, dado que $[p.(1-p)]$ é a variância de uma variável dicotômica e ela será máxima para um $p = 0,5$. Supondo-se que se desejem testes com grande variabilidade, todos os itens selecionados devem ter índices de dificuldade médios. Neste caso, dever-se-ia esperar uma distribuição aproximadamente normal e de grande variabilidade, uma vez que se assume que a maioria dos traços psicológicos se distribuem normalmente.

A Teoria de Resposta ao Item (TRI), que o analisa por meio de três parâmetros, denomina o índice de dificuldade de parâmetro b . Esta medida representa a abscissa do ponto de inflexão da curva característica do item, que corresponde a uma probabilidade de 50% de acerto. É expresso em escore z , e, na prática, seus valores variam entre -3 e $+3$. Um valor igual a zero indica uma dificuldade mediana, exigindo do sujeito uma habilidade média (percentil de 50) do construto em questão para acertar o item.

Cabe destacar, ainda, que é dispensável determinar o índice de dificuldade nos testes em que o aspecto que está sendo avaliado é a velocidade, e não o conhecimento (Bunchaft & Cavas, 2002).

Segundo Erthal (1987), é um erro construir um teste selecionando itens de acordo apenas com a sua facilidade. O valor da dificuldade do item proporciona uma indicação precisa de quanto o item é difícil para o testando, mas não proporciona informação da utilidade do item em medir o construto do teste. Essa estimativa

é feita pela determinação do poder de discriminação do item (McIntire & Miller, 2000).

DISCRIMINAÇÃO DOS ITENS

A capacidade de discriminação do item consiste, segundo McIntire e Miller (2000), na medida de quanto ele separa os examinandos que demonstram um alto grau de habilidade, conhecimento, atitude ou característica de personalidade, daqueles que os mostram em baixo grau. O índice expressa o poder do item em diferenciar sujeitos com magnitudes semelhantes no traço latente, evidenciando que quanto mais próximas são as magnitudes do traço que o item puder diferenciar, mais discriminativo ele é.

Existem dezenas de técnicas estatísticas de correlação para estabelecer o índice de discriminação, as quais produzem resultados similares (Anastasi & Urbina, 2000; Pasquali, 2003).

Não obstante, tipo de técnica utilizado dependerá das propriedades dos dados obtidos. Para Erthal (1987), bons itens são aqueles que obtêm uma boa correlação com outros itens e uma correlação elevada com o escore total do teste. Se um item mede um aspecto particular de alguma variável, a correlação entre item e escore total deve ser positiva. Itens ruins apresentam uma correlação nula ou negativa. Uma correlação negativa indica-nos que os indivíduos se saíram bem no teste, embora tenham tido um mau desempenho no item, ou vice-versa. Um item com poder discriminativo perfeito é aquele em que todos os sujeitos com escores altos no teste acertam, enquanto aqueles com escores baixos no teste erram.

O cálculo do índice de discriminação com base no escore total apresenta, segundo Pasquali (1996, 2003), um problema teórico. Busca informações sobre o item baseando-se em todos os itens (escore total), quando ainda não se sabe se os itens do teste são homogêneos.

A TRI tem outra maneira de determinar o índice de discriminação de um item. Seu cálculo é feito pelo algoritmo da função da curva característica do item e é expresso pelo parâmetro a , que consiste no ângulo de incidência da curva no momento da inflexão, ou seja, onde ela corta o ponto de probabilidade de 50% de resposta correta. O valor é expresso numa escala que varia de 0 a 4 e, segundo Hambleton, Swaminathan e Roger (1991), na prática não são usuais valores que extrapolem o intervalo de 0 a 2. Valores altos indicam itens mais capazes para separar examinandos com níveis semelhantes de habilidade.

VALIDADE DOS ITENS

Validade do item refere-se a este estar relacionado com aquilo que pretende medir. Na Teoria Clássica dos Testes (TCT) se diz que o item é válido se tiver correlação com um critério externo; na TRI, se ele for uma representação adequada do traço latente (θ).

Na TCT, o critério externo que é relacionado com o item pode ser um teste paralelo ou uma medida do desempenho do sujeito, numa situação para a qual o teste foi construído. Esse tipo de validade corresponde à validade preditiva do item, isto é, à capacidade do item de prever o critério. Na TRI, a carga resultante da análise fatorial é que melhor expressa tal informação, pois representa a comunalidade ou covariância entre o item e o fator: quanto maior a covariância, maior a validade do item. Tabachnick e Fidell (1996) propõem 0,32 como carga mínima para qualificar o item como um representante útil do construto. No presente trabalho, que teve por intenção examinar a qualidade do maior número possível de itens, o ponto de corte foi mais leniente, considerando-se 0,30 como a carga mínima, uma vez que, segundo Pasquali (1999), este índice demonstra uma covariância de aproximadamente 10% ($0,30^2$) entre o item e o fator - o que não pode ser desconsiderado.

Analisar as características psicométricas dos itens de um instrumento é uma etapa relevante em sua elaboração; entretanto, o manual original do teste WISC-III (Wechsler, 1991) não fornece informações relativas à qualidade dos itens. Este trabalho apresenta os resultados de análises desenvolvidas nos dados da amostra de padronização do teste WISC-III ao contexto brasileiro, dando continuidade aos estudos de Figueiredo (2001).

MÉTODO

Amostra

O estudo foi feito com base nos dados da amostra de padronização, da pesquisa de adaptação do teste WISC-III ao contexto brasileiro (Figueiredo, 2001). Os itens analisados tiveram por base as pontuações obtidas nos protocolos de testes aplicados em 801 crianças e adolescentes de idade entre 6 e 16 anos, matriculados em escolas públicas e privadas da zona urbana do município de Pelotas (RS).

Procedimentos para análise dos dados

Inicialmente foi realizado um estudo exploratório dos dados e da unidimensionalidade dos subtestes

(mais detalhes em Figueiredo, 2001). A análise subsequente dos parâmetros dos itens de cada subteste do WISC-III não considerou os subtestes *código* e *procurar símbolos*, por se caracterizarem como provas de velocidade.

Os parâmetros da dificuldade e da discriminação foram investigados pelos métodos da Teoria Clássica dos Testes (TCT) e pela Teoria de Resposta aos Itens (TRI). Na TRI, utilizou-se o modelo de dois parâmetros de Birnbaum (citado por Pasquali, 1996), admitindo que não haveria respostas corretas dadas ao acaso.

Para o estudo da discriminação, foram considerados o parâmetro *a* da TRI e a correlação ponto bisserial ou ponto polisserial; para o estudo da dificuldade, foram considerados o parâmetro *b* da TRI e a proporção de respostas certas ao item. O critério para estimar a validade dos itens foi o das cargas fatoriais produzidas na solução fatorial de um fator, resultante do estudo da unidimensionalidade dos subtestes.

Para os subtestes com itens dicotômicos (certo/errado), a análise dos parâmetros de discriminação e dificuldade foi realizada no programa *XCALIBRE*. O programa também estima a correlação entre o ponto bisserial e o percentual de acertos dos itens.

Os parâmetros relativos à discriminação e dificuldade dos itens politômicos foram estabelecidos no programa computacional *Parscale* (Muraki & Bock, 1997). O programa faz análise, segundo a TRI, de itens pontuados com categorias múltiplas, estimando os parâmetros e ajustando o modelo. Também oferece as correlações polisseriais e faz estimativas das frequências e respectivas porcentagens de respostas em cada categoria de pontuação.

As cargas fatoriais, utilizadas como referencial da validade dos itens dicotômicos, foram estimadas no programa *TESTFACT*, através da análise fatorial de um fator, do tipo *Full Information*. Para a análise dos itens politômicos, as cargas fatoriais foram produzidas no programa SPSS.

APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Nesta seção apresenta-se um sumário dos principais resultados relativos às características psicométricas dos itens do WISC-III. A Tabela 1 mostra os dados referentes à média dos parâmetros de discriminação (*a*) e dificuldade (*b*) dos itens, segundo a TRI. Para facilitar a interpretação do nível de dificuldade, calculou-se o percentil equivalente. Para

tanto, encontrou-se a área, da curva normal, referente à distância do valor de *z* encontrado para o parâmetro da dificuldade (*b*) até o ponto médio da curva ($z=0$). O valor dessa distância foi acrescentado ao percentil médio (50).

Tabela 1. Média dos Parâmetros de Discriminação e Dificuldade dos Itens

Subtestes	\bar{X}	Dificuldade		
		Dif.	P	Interpretação
<i>Verbais</i>				
Informação	1,49	0,44	67	Médio Superior
Semelhanças	1,00	0,46	68	Médio Superior
Aritmética	1,63	-0,24	41	Médio Inferior
Vocabulário	0,91	-0,36	36	Médio Inferior
Compreensão	0,89	0,08	53	Médio
Dígitos - OD	1,95	0,35	64	Médio Superior
- OI	2,17	1,27	90	Superior
Média	1,43	0,28	60	Médio Superior
<i>Não Verbais</i>				
Completar Figuras	0,97	0,19	57	Médio
Arranjo de Figuras	0,91	0,15	56	Médio
Cubos	2,22	-0,34	37	Médio Inferior
Armar Objetos	0,93	-0,57	31	Médio Inferior
Média	1,25	-0,14	45	Médio Inferior

Legenda: \bar{X} = média da discriminação; Dif.= dificuldade; P=percentil; OD = Ordem Direta; OI = Ordem Inversa

Quanto ao poder médio de *discriminação* dos itens (*a*), os subtestes do conjunto verbal mostraram um valor levemente superior ao do conjunto não-verbal, sugerindo itens mais discriminativos. Contraditoriamente, faz parte do conjunto não-verbal o subteste *Cubos*, com maior média de poder discriminatório, enquanto do conjunto verbal faz parte o subteste *Compreensão*, com a menor média. Em termos gerais, os itens do teste têm capacidade de diferenciar indivíduos com níveis semelhantes de habilidade cognitiva.

Quanto ao grau de *dificuldade* dos itens (*b*), o subteste mais difícil foi *Dígitos* (ordem inversa), enquanto os mais fáceis foram *Armar objetos*, *Vocabulário*, *Cubos* e *Aritmética*. A evidência de ser necessária uma habilidade intelectual dentro da média (percentil entre 25 e 75) para responder a maioria dos subtestes sugere que o teste não avalia com tanta adequação sujeitos dos extremos superior e inferior da habilidade intelectual.

Ainda em relação ao estudo sobre a *dificuldade* dos itens, com base nos valores de *z* encontrados para os índices de dificuldade (parâmetro *b* da TRI), calculou-se o percentual de itens encontrados em cada faixa de dificuldade, segundo critério proposto por

Cerdá (citado por Erthal, 1987 e Pasquali, 1996). Os dados aparecem nas Tabelas 2 e 3.

Tabela 2. Distribuição Ideal e Observada do Nível de Dificuldade dos Itens dos Subtestes Verbais

Interpretação	Ponto de Corte (z)	Percentual Observado					
		Inf	Sem	Arit	Voc	Com	DOD DOI
Extremamente fáceis	< -1,28	10	21	37	20	17	14
Fáceis	-1,28 a -0,51	10	16	13	17	11	00
Medianos	-0,52 a 0,51	21	16	17	28	22	14
Difíceis	0,52 a 1,28	28	21	08	28	11	00
Extremamente difíceis	> 1,28	31	26	25	07	39	49

Legenda: Inf = Informação; Sem = Semelhanças; Arit = Aritmética; Voc = Vocabulário; Com = Compreensão; DOD = Dígitos Ordem Direta; DOI = Dígitos Ordem Inversa.

Tabela 3. Distribuição Ideal e Observada do Nível de Dificuldade dos Itens dos Subtestes Não-Verbais

Faixa	Interpretação	Ponto de Corte (z)	Percentual Observado			
			CFig	AFig	Cubos	AObj
I	Extremamente fáceis	< -1,28	17	14	25	20
II	Fáceis	-1,28 a -0,51	23	7	17	20
III	Medianos	-0,52 a 0,51	20	43	33	60
IV	Difíceis	0,52 a 1,28	7	22	17	00
V	Extremamente difíceis	> 1,28	33	14	8	00

Legenda: CFig = Completar Figuras; AFig = Arranjo de Figuras; Cub= Cubos; AObj = Armar Objetos.

Quanto ao nível de dificuldade, espera-se uma distribuição equilibrada, similar à curva normal, devendo haver itens que cubram toda a extensão de magnitude do traço latente. Observa-se nas Tabelas 2 e 3 que os itens dos subtestes distribuíram-se por todos os níveis de dificuldade, com exceção de *Dígitos* e *Armar objetos*; porém a distribuição das proporções não corresponde ao proposto por Cerdá (Pasquali, 1996). Uma maior concentração de itens medianos foi encontrada somente nos subtestes *Arranjo de figuras*, *Cubos* e *Armar objetos*. A tendência da distribuição dos itens nos outros subtestes foi de uma curva bimodal, onde há concentração de itens extremamente fáceis e de itens extremamente difíceis.

A eliminação de alguns desses itens permitiria uma melhor distribuição do nível de dificuldade e um menor tempo de aplicação. Não obstante, todos os itens foram mantidos, considerando-se que, em estudos posteriores, eles poderão ser úteis para discriminar, com mais exatidão, indivíduos dos extremos inferior e superior da habilidade cognitiva.

Quanto à *validade* dos itens dos subtestes do WISC-III, os valores das cargas fatoriais foram agrupados segundo as categorias apresentadas por Comrey (1973), que propôs um critério bastante rigoroso. Na Tabela 4 aparecem os percentuais de cargas fatoriais obtidos nos itens de cada subteste.

Tabela 4. Percentual das Cargas Fatoriais Segundo as Categorias de Comrey

Subtestes	D	P	R	B	M	E
<i>Verbais</i>						
Informação	00	07	03	00	24	66
Semelhanças	11	21	32	26	05	05
Aritmética	00	00	00	00	08	92
Vocabulário	21	21	27	17	14	00
Compreensão	06	28	33	16	11	06
Dígitos	27	06	27	13	27	00
<i>Não-Verbais</i>						
Completar Figuras	00	03	17	23	23	34
Arranjo de Figuras	07	14	50	22	07	00
Cubos	00	17	17	17	17	32
Armar Objetos	00	00	20	00	60	20

Legenda: D: Desprezível < 0,32; Pobre 0,32 a 0,44; Razoável 0,45 a 0,54; Boa 0,55 a 0,62; Muito Boa 0,63 a 0,70; Excelente \geq 0,71; OD = Ordem Direta; OI = Ordem Inversa.

Observa-se na Tabela 4 que três dos subtestes (*Informação*, *Aritmética* e *Armar objetos*) apresentaram cargas entre muito boas e excelentes. *Completar figuras* e *Cubos* apresentaram itens na categoria excelente, mas também em categorias inferiores. *Semelhanças* e *Arranjo de figuras* mostraram cargas consideradas entre boas e razoáveis, enquanto *Vocabulário* e *Dígitos* têm mais itens com cargas razoáveis e desprezíveis, indicando que seus itens teriam menor representatividade do construto que avaliam.

Considerando-se os requisitos examinados de discriminação e validade de cada item dos subtestes (Figueiredo, 2001), agruparam-se, na Tabela 5, os itens identificados como pouco discriminativos e/ou com baixa representatividade do construto. Utilizou-se um critério de consenso - coeficientes menores que 0,30 - para identificá-los tanto nas análises do parâmetro *a*, da TRI, como nas correlações ponto bisseriais ou polisseriais, da TCT. Também, em relação à validade, foram destacados os itens com cargas fatoriais menores que 0,30.

Tabela 5. Itens do Teste WISC-III com Baixos Índices de Discriminação e/ou Validade

Subtestes	Discriminação		Validade
	Parâmetro <i>a</i>	Correlação	
<i>Verbais</i>			
Informação	--	2, 3, 4, 30, 29	--
Semelhanças	--	--	17
Aritmética	--	1, 21, 22, 23, 24	--
Vocabulário	1, 3, 4, 7, 11, 12, 13, 29	3	1, 2, 3, 28, 29
Compreensão	2	--	2
Dígitos - OD	1, 8	--	1, 2, 8
Dígitos - OI	--	--	7
<i>Não-Verbais</i>			
Completar Figuras	--	1, 2, 26, 27, 28, 29, 30	--
Arranjo de Figuras	--	--	14

Os itens cubos e arma objetos não apresentam baixos índices de discriminação.

Conforme a Tabela 5, os subtestes *Cubos* e *Armar objetos* não apresentaram itens com baixa qualidade psicométrica. Nos demais subtestes, alguns itens foram identificados como de baixa qualidade.

Na escala verbal, os itens 2 (orelhas), 3 (patas), 4 (ferver), 29 (distância São Paulo) e 30 (Darwin) do subteste Informação foram identificados com pouca capacidade discriminativa, segundo a correlação ponto-bisserial, apresentando os maiores percentuais de respostas corretas ou erradas. O item 1 (nariz) não foi avaliado por não apresentar variância de respostas, uma vez que todos os sujeitos acertaram o item. Em *Aritmética*, os itens pouco discriminativos foram o 1 (contar pássaros), 21 (rádio), 22 (viagem), 23 (canetas) e 24 (carros), o mesmo acontecendo com os itens 1 (chapéu), 2 (relógio), 3 (ladrão), 4 (guarda-chuva), 7 (bicicleta), 11 (sair), 12 (alfabeto), 13 (exato), 28 (iminente) e 29 (compelir) no subteste *Vocabulário*. Neste subteste, o item 30 (dilatatório) não foi analisado por falta de variância, uma vez que todos os sujeitos o erraram.

Os coeficientes foram satisfatórios para todos os itens do subteste *Compreensão*, com exceção do 2 (O que você faria se encontrasse uma carteira de alguém em uma loja?), o segundo item mais fácil do subteste. Em *Dígitos* os itens identificados com baixa qualidade foram os mais fáceis - 1 e 2 (respectivamente, memorizar 2 e 3 números na ordem direta) e os mais difíceis - 8 e 7 (respectivamente, memorizar 9 números na ordem direta e 8 números na ordem inversa).

Na escala não-verbal, em *Completar figuras*, os itens 1 (raposa), 2 (gato), 24 (peixe), 27 (mercado), 28 (casa), 29 (guarda-chuva) e 30 (tênis), apontados como pouco discriminativos segundo o critério da

correlação ponto-bisserial, foram os que tiveram maior percentual de acertos ou erros. No *Arranjo de figuras*, o item 14 (sombra) foi o que apresentou o maior índice de dificuldade, podendo a baixa carga fatorial ser atribuída, também, à pouca variância.

Antes de optar pela eliminação dos itens de baixa qualidade psicométrica, observou-se que tais itens se caracterizaram, geralmente, como os que tiveram maior ou menor percentual de acertos. Dessa forma, a pouca variabilidade de respostas pode ter influenciado nas análises. O coeficiente de discriminação, estimado pelas correlações ponto biserial (PBs), apresenta um problema com itens em que o percentual de respostas corretas ou erradas é elevado. No caso de itens muito fáceis ou muito difíceis, nos quais a maioria do grupo acerta ou erra, a falta do grupo-critério irá influenciar no cálculo da correlação. Na sua fórmula é incluída a média no teste dos sujeitos que acertaram o item, a média total do teste, o desvio-padrão do teste e a proporção de sujeitos que acertaram o item. A correlação ponto biserial será sempre pequena e nesse caso oferece uma informação pouco confiável. Assim, nenhum item da Tabela 5 foi eliminado do instrumento, considerando-se que a variância de respostas foi nula ou pequena. Entretanto, os itens 8 do *Vocabulário* e 4 da *Compreensão* merecem atenção em estudos posteriores, pois não estão de acordo com esse raciocínio.

Considerando-se a necessidade de apresentação dos itens segundo a ordem crescente de dificuldade, foi feito um estudo tomando-se por base os valores do percentual de acertos dos itens, resultantes da análise clássica da dificuldade, os quais apontaram para a necessidade de reorganização da ordem original americana, visando a uma melhor adequação à amostra pesquisada (mais detalhes em Figueiredo, 2001).

Tabela 6. Sugestão para Ordem de Apresentação dos Itens dos Subtestes do WISC-III

Subteste	Item
Completar Figuras	1, 2, 3, 5, 4, 7, 6, 9, 15, 8, 11, 16, 10, 14, 12, 18, 13, 19, 17, 22, 23, 20, 26, 21, 25, 24, 29, 27, 30, 28
Informação	1, 2, 3, 5, 4, 6, 7, 8, 9, 10, 11, 13, 22, 12, 15, 25, 14, 20, 16, 18, 17, 19, 21, 24, 26, 30, 23, 27, 29, 28
Semelhanças	2, 4, 3, 1, 5, 6, 7, 9, 8, 10, 11, 13, 15, 12, 14, 16, 19, 18, 17
Arranjo de Figuras	1, 2, 3, 5, 4, 6, 9, 7, 8, 10, 11, 12, 13, 14
Aritmética	1, 2, 3, 4, 5, 7, 6, 8, 10, 9, 12, 11, 13, 14, 15, 16, 18, 17, 20, 19, 23, 21, 22, 24
Cubos	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 11
Vocabulário	2, 1, 8, 3, 7, 5, 4, 10, 18, 12, 9, 6, 16, 19, 23, 11, 15, 14, 17, 24, 27, 13, 25, 22, 29, 21, 20, 28, 26, 30
Armar Objetos	1, 2, 3, 5, 4
Compreensão	1, 4, 5, 3, 2, 6, 8, 9, 10, 11, 7, 13, 15, 12, 16, 18, 14, 17

Os subtestes *Dígitos*, *Código* e *Procurar símbolos* não constam na Tabela 6 porque não se procedeu a nenhuma modificação na ordem de administração de

seus itens em relação à amostra brasileira. Em *Cubos* e *Armar objetos* a alteração proposta implicou no deslocamento de apenas um item, enquanto nos

demais subtestes as mudanças necessárias foram mais substanciais.

Em *Aritmética*, a ordem real de dificuldade dos seis primeiros itens foi: 1 (contar 3 pássaros), 3 (deixar 4 árvores), 2 (contar 12 árvores), 5 (sorvetes), 7 (livros), 4 (deixar 9 árvores) e 6 (maçã). Entretanto, para facilitar o manuseio do Livreto de Estímulos, que na versão americana apresenta o material para administrar os cinco itens iniciais, decidiu-se manter estes itens na mesma ordem da versão original do teste.

Em *Arranjo de figuras*, o item 12 (banhos) teve maior percentual de acertos do que o item 11 (chuva), concluindo-se que aquele foi mais fácil. Entretanto, considerando-se que o tempo médio de execução da amostra brasileira foi o mesmo do teste original (item 12 = 60" e item 11 = 45"), a mudança na ordem desses itens iria alterar o ritmo do examinador no que tange ao controle do tempo, que normalmente segue uma ordem gradativa. Decidiu-se, por essa razão, manter a seqüência original desses dois itens.

A evidência da necessidade de reordenar os itens dos subtestes mostrou que as experiências culturais dos sujeitos influenciam o nível de facilidade das questões de um teste. A ordenação dos itens, na amostra brasileira, configurou-se bastante diversa da apresentada no teste original (Wechsler, 1991), principalmente nos subtestes verbais. Os dados reforçam a necessidade de estudos de adaptação de instrumentos estrangeiros.

CONCLUSÃO

O presente estudo avaliou propriedades psicométricas de itens dos subtestes do WISC-III referentes a dificuldade, discriminação e validade, concluindo que, em termos gerais, os itens deste teste apresentaram qualidade psicométrica adequada para o contexto brasileiro. Constatou-se a tendência de existir maior quantidade de itens com dificuldade mediana, embora em alguns subtestes a distribuição da dificuldade de seus itens se afaste da distribuição ideal que prevê esta maior quantidade. Constatou-se também que a maioria dos subtestes tem pouco ou nenhum item com baixo poder de discriminação, o mesmo acontecendo com os índices de validade. Nessa análise, os itens apontados como duvidosos, coincidentemente, apresentaram pouca variância de respostas, e por essa razão seus resultados podem não ser confiáveis.

Não obstante, o teste parece não estar adequado para discriminar sujeitos de habilidades extremas. Os

dados apontam para a necessidade de incluir na amostra estudada sujeitos deficientes mentais e portadores de altas habilidades, tendo-se em vista que a população investigada mostrou uma concentração de sujeitos com capacidade mediana. A inclusão permitiria checar os resultados encontrados e definir com maior precisão a pertinência de eliminar os itens identificados como de baixa qualidade psicométrica, oportunizando a adequação do teste a tais grupos clínicos.

Nenhuma comparação pode ser feita entre os achados do presente estudo e as características dos itens na versão norte-americana, uma vez que o manual original (Wechsler, 1991) não faz nenhuma referência ao estudo de análises empíricas dos itens.

Os resultados do presente estudo possibilitarão aos usuários do teste, principalmente aos profissionais que trabalham com avaliação neuropsicológica, conhecer a qualidade dos itens e/ou dos subtestes em relação à capacidade discriminativa, ao nível de dificuldade e à representatividade do traço latente.

REFERÊNCIAS

- Adonez, G. P. (1999). Procedimentos de construcción y analisis de tests psicométricos. En S. Wechsler & R. Guzzo (Orgs.), *Avaliação Psicológica: perspectiva internacional* (pp. 74-100). São Paulo: Casa do Psicólogo.
- Anastasi, A. & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artes Médicas.
- Bunchaft, G. & Cavas, C. (2002). *Sob medida*. São Paulo: Vetor.
- Erthal, T. C. (1987). *Manual de psicometria*. Rio de Janeiro: Jorge Zahar.
- Comrey, A. L. (1973). *A first Course in factor analysis*. New York: Academic.
- Figueiredo, V. L. M. (2001). *Uma adaptação brasileira do teste de inteligência WISC-III*. Tese de Doutorado não publicada. Programa de Pós-Graduação em Psicologia, Instituto de Psicologia, Universidade de Brasília, Brasília.
- Guilford, J. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Hambleton, R., Swaminathan, H. & Roger, H. (1991). *Fundamental of item response theory*. London: Sage.
- Hetzell, S. (1997). *Basic concepts in item and test analysis*. EricDatabase, Retrieved July 25, 2000 from <http://www.eric.ed.gov>
- McIntire, S. & Miller, L. (2000). *Foundations of psychological testing*. Boston: McGraw-Hill.
- Muraki, E. & Bock, R. (1997). PARCALE – IRT item analysis and test scoring for rating scale data (version 3.0) [programa de computador]. Chicago: Scientific Software Internacional.
- Pasquali, L. (Org). (1996). *Teoria e métodos de medida em Ciências do Comportamento*. Brasília: Laboratório de pesquisa em avaliação e medida/Instituto de Psicologia/UnB/INEP.

- Pasquali, L. (1999). Testes referentes a construto. Em L. Pasquali (Org), *Instrumentos psicológicos: manual prático de elaboração* (pp. 37-71). Brasília: LABPAM/OBAPP.
- Pasquali, L. (2003). *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis: Vozes.
- Tabachnick, B. & Fidell, L. (1996). *Using multivariate statistics*. California: Harper Collins College.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition Manual*. San Antonio: Psychological Corporation.

Recebido em 16/11/06

Aceito em 16/06/07

Endereço para correspondência : Vera L. M. de Figueiredo. Quinze de novembro, 457/apto 704, Centro, CEP 96015-000, Pelotas-RS. *E-mail:* verafig@terra.com.br