Qualidade de testes psicológicos

Orlete Maria Pompeu de Lima – FMA e Universidade do Paraná Ana Paula Porto Noronha – Universidade São Francisco

Resumo

O presente estudo teve como objetivo avaliar a análise da qualidade de testes promovida por psicólogos, assim como determinar seu consenso. Participaram do estudo seis psicólogos e 14 estudantes de pós-graduação em psicologia. Os dois grupos de sujeitos avaliaram quatro testes utilizando um instrumento de avaliação de testes psicológicos, de forma que houve, além da avaliação de cada grupo, a comparação entre eles. O acerto das respostas dos sujeitos em comparação com o gabarito foi investigado por meio do coeficiente de correlação de Pearson. Os resultados indicaram bons índices de concordância e, em apenas um teste, a avaliação dos alunos da pós-graduação foi superior ao dos avaliadores psicólogos.

Palavras-chave: Testes psicológicos; Avaliação psicológica; Qualidade de testes; Parâmetros psicométricos.

Psychological test qualities

Abstract

The present study aimed to assess the quality analysis of psychological tests done by psychologists, as well to verify their agreement. Six psychologists and 14 psychology graduate students participated on this study. Both subject groups evaluated four tests using an instrument of assessment of psychological tests, allowing a comparison between them and an evaluation of each group. The participants' answers were correlated with the right answers using the Pearson's test. The results showed a good internal consistency, and only in one test, the graduate students' evaluation was better than the psychologists' one.

Keywords: Psychological instruments; Psychological assessment; Test qualities; Psychometric parameters.

Introdução

A busca pela qualidade das ações profissionais dos psicólogos tem sido anseio de pesquisadores, como Bartram (1998), que aponta em seu artigo a necessidade de um guia para estabelecimento de diretrizes internacionais para a padronização e o uso dos testes psicológicos. Ao lado disso, Turner, DeMers, Fox e Reed (2001) organizaram um guia de responsabilidades para o uso de testes padronizados. Os critérios referentes à qualidade são identificados freqüentemente em publicações estrangeiras sobre avaliação psicológica ou instrumentos de avaliação (Canadian Psychological Association, 1996, Adánez, 1999, Oakland, 1999; AERA, APA & NCME, 1999 e International Test Commission, 2001) e mais raramente em trabalhos nacionais (Pasquali, 1998).

No Brasil, a construção de testes foi bastante acelerada entre os anos 1930 e 1960, sem grandes preocupações com os parâmetros psicométricos por parte dos pesquisadores, talvez por causa da precariedade de recursos tecnológicos utilizados nas técnicas estatísticas (Nick, 1988). A pesquisa para construção, adaptação e padronização de testes psicológicos é relativamente recente, e atualmente pode-se contar com o auxílio da informática para a escolha da amostra e para as análises estatísticas de forma mais rápida e precisa. Entretanto, Pasquali (1999) relata que pesquisadores brasileiros, em determinado momento, utilizam instrumentos estrangeiros, sem maiores preocupações com a verificação da qualidade e da aplicabilidade desses no contexto cultural brasileiro.

A relevância do tema justifica-se na medida em que os testes psicológicos são definidos como instru-

mentos de medida em Psicologia e que devem apresentar certas características que justifiquem, ao profissional, utilizar os resultados que produzem. Nesse sentido, Oakland (1996) argumenta que os testes são produzidos e utilizados, em maior quantidade e qualidade, em países que possuem o sistema educacional bem-desenvolvido e que valorizam a tecnologia e as diferenças individuais.

Com o objetivo de verificar até que ponto os testes estrangeiros utilizados pelos profissionais brasileiros, em diferentes campos, têm sido submetidos tanto aos processos de padronização e/ou adaptação como de análise das características metrológicas em amostras brasileiras, Sisto, Codenotti, Costa e Nascimento (1979) publicaram um artigo, que apesar de ter sido realizado há quase três décadas, suas conclusões não se diferenciaram muito do cenário atual. Os resultados indicaram que nenhum dos nove testes analisados cumpriu pelo menos 40% das condições que foram propostas, isto é, não apresentaram comprovações ou informações no que se refere à precisão, validade e padronização dos resultados (normas, tabelas, tipos de amostras, faixa etária, sexo e nível socioeconômico e/ou cultural).

Outras pesquisas têm apontado insatisfações quanto à ausência de instrumentos de qualidade para o profissional poder fazer uso com alguma segurança na elaboração de seus diagnósticos. Entre elas, uma pesquisa com psicólogos de Brasília realizada por Azevedo, Almeida, Pasquali e Veiga (1996), para identificar os instrumentos psicológicos mais utilizados na prática profissional e levantar a opinião deles a respeito deles, concluiu que a maioria dos inquiridos acredita no valor dos testes e que esses medem o que pretendem medir, desde que utilizados como uma ferramenta adicional e de acordo com os princípios éticos, mas apontam a falta de estudos prévios que justifiquem a validade de seu uso.

Essa mesma inadequação quanto à qualidade dos testes psicológicos pôde ser verificada em outros países pelos achados de Almeida, Prieto, Muñiz e Bartram (1998) com psicólogos de Portugal, Espanha e países ibero-americanos. Foram identificados problemas quanto ao uso inadequado dos testes psicológicos, falta de domínio do instrumento, avaliação incorreta, falta de desenvolvimento dos testes, entre outros.

Outros artigos sobre o tema confirmam o descuido em relação à grande parte dos instrumentos psicológicos comercializados no Brasil, que não apresentam a qualidade técnica necessária e indispensável para o exercício profissional. Noronha, Freitas, Sartori e Ottati, (2001) realizaram uma pesquisa para analisar a presença de informações relativas aos dados de identificação e aos dados de aplicação e de avaliação nos manuais de 21 testes de inteligência, que possuíam manual em português. Os resultados foram desanimadores, considerando-se que nos instrumentos foram encontrados alguns que não apresentavam os dados de identificação (nome, autor, editor e data de publicação), embora metade deles tenha apresentado todos os itens dos aspectos avaliados.

Em outra pesquisa, Noronha, Freitas, Sartori e Ottati (2002) estudaram a construção de 26 testes psicológicos de inteligência, comercializados no Brasil, para verificar a presença ou a ausência de estudos de validade e precisão e também identificar se os testes estrangeiros possuíam padronização brasileira. Os resultados indicaram que no que se refere à validade e precisão, 75% dos testes nacionais têm esse tipo de estudo, enquanto nos testes estrangeiros, 90% desenvolveram algum estudo. Em relação à padronização dos instrumentos, 25% dos testes nacionais não informavam os estudos realizados e nos instrumentos estrangeiros, 40%.

No que se refere ao número de sujeitos incluídos nos estudos de padronização dos testes, segundo os dados apresentados nas pesquisas, a variabilidade presente leva a refletir com urgência sobre a questão (Hinkle, Wiersma & Jurs 1994). A ausência de padrões específicos pode, de alguma forma, justificar a disparidade do número de sujeitos encontrados nos instrumentos. Tal fato se justifica pela ausência de parâmetros nacionais sólidos a respeito da construção de instrumentos.

Quanto à fundamentação teórica dos instrumentos, Pasquali (1996) comenta a falta de teorização da medida das ciências psicossociais e dos construtos avaliados, ressaltando a visão positivista da psicometria. Para ele faz-se necessária uma visão mais cognitivista, para a melhor definição psicológica dos traços latentes a serem avaliados. Almeida (1999) aponta também a necessidade de um maior esforço dos pesquisadores quanto aos aspectos teóricos das variáveis avaliadas, isto é, a fundamentação teórica deve explicitar a teoria sobre o construto ou objeto psicológico a ser medido e definir o desenvolvimento do instrumento e a definição constitutiva e operacional do construto.

Outro aspecto importante sobre os testes psicológicos é a questão relativa à interpretação dos resultados. Halperin e McKay (1998), na revisão de testes usados com crianças e adolescentes, detectaram que alguns testes possuem poucos estudos psicométricos, principalmente os projetivos, portanto afirmam que dados isolados pouco contribuem para o entendimento do psicólogo e que é recomendável o uso do teste como um instrumento auxiliar de coleta de dados.

Segundo Garb (2000) pode-se esperar que os recursos da informática fiquem cada vez mais importantes para a qualidade da avaliação psicológica. Embora a validade ainda não tenha sido claramente estabelecida para muitos programas atuais de avaliação automatizados, o autor faz recomendações para a busca de qualidade na construção e avaliação de novos programas de computador, além de comentar, no artigo, sobre a ética no uso de computadores para fazer julgamentos.

Os problemas mais graves e mais freqüentes na avaliação psicológica, com ênfase nos testes, segundo os psicólogos, foram tema da pesquisa realizada por Noronha (1999) em sua tese de doutorado, cujos resultados reafirmam a falta de pesquisas e de instrumentos nacionais para aferição dos recursos utilizados na avaliação psicológica, como o problema mais grave e mais freqüente. Ainda são utilizados muitos testes, criados no começo do século XX, e, ao contrário de todas as outras inovações tecnológicas, como a informática, a saúde e as telecomunicações, os testes psicológicos permanecem, basicamente, imutáveis.

Para Noronha (1999) as críticas atribuídas aos testes psicológicos são inúmeras no que se relaciona à inconsistente fundamentação teórica, à ausência de estudos recentes e de adaptações para diferentes realidades, ao alto custo de determinados materiais e à falta de orientações precisas quanto à construção e ao uso dos instrumentos psicológicos. Isso evidencia a necessidade de avanços na área, como a verificação sistemática de seus parâmetros para a construção, o uso e a comercialização dos testes psicológicos.

Sob essa perspectiva, Sisto, Sbardelini e Primi (2001) ponderam que as dúvidas quanto à qualidade dos testes psicológicos não estimularam a busca das soluções; em contrapartida, induziram ao uso dos recursos escassos disponíveis, assim como ao abuso de métodos pouco científicos. Além disso, os autores denunciam a carência de desenvolvimento tecnológico nacional condizente com a nossa cultura e a falta

de preparo dos profissionais que trabalham com avaliação psicológica.

O Conselho Federal de Psicologia constituiu em 2001 uma comissão consultiva em avaliação psicológica, que teve como tarefa apreciar a qualidade dos testes psicológicos que estavam em uso no Brasil. Essa comissão listou os critérios para a análise dos instrumentos e convidando consultores nas áreas de conhecimento dos testes psicológicos, para auxiliar na elaboração do parecer final de cada instrumento (Conselho Federal de Psicologia, 2003). O presente estudo teve como objetivo avaliar a análise da qualidade de testes promovida por psicólogos, assim como determinar o consenso entre eles no que se refere à análise.

Método

Participantes

Os participantes do presente trabalho foram divididos em dois grupos:

Grupo 1 – Seis psicólogos, sendo 83,3% mulheres (f=5) e 16,7% (f=1) homens, com idade média de 42 anos (DP=6,04) e com uma média 15,16 anos de formados (DP=4,66). Em relação à área de atuação, 16,7% (f=1) atuam na área clínica, escolar e docência, 50% (f=3) na clínica e organizacional, 16,7% (f=1) na clínica, organizacional e trânsito e 16,7% (f=1) na clínica e orientação vocacional.

Grupo 2 – 14 psicólogos, alunos de pós-graduação em psicologia de uma instituição particular do interior de São Paulo, sendo 50% mulheres (f=7) e 50% (f=7) homens, com idade média de 37 anos (DP=9,17) e com média de 8,96 anos de formação (DP=7,83). Em relação à área de atuação 7,1% (f=1) atuam somente na área clínica, 28,7% (f=4) na organizacional, 7,1% (f=1) na docência, 7,1% (f=1) na pesquisa, 14,3% (f=2) na clínica e docência, 21,5% (f=3) na clínica e trânsito, 7,1% (f=1) na clínica, organizacional e pesquisa e 7,1% (f=1) na pesquisa e forense.

Material

Foram utilizados os seguintes materiais para a realização da pesquisa:

- Ficha de dados de identificação, para levantamento de dados, como idade, sexo, tempo de formado, nome da instituição de formação, cursos de pós-graduação, área de atuação profissional e tempo de atuação.
- Quatro testes psicológicos: Bateria de Provas de Raciocínio (BPR-5) Formas A e B (Primi & Almeida

2000); Inventário Fatorial de Personalidade, IFP (Pasquali, Mazzarello & Ghesti 1997); Teste de Desempenho Escolar, TDE, (Stein, 1994); 16 PF (Cattell, Cattell, 2000).

1. Instrumento de análise da qualidade dos testes (Anexo 1)

O instrumento utilizado constituiu-se de 47 questões organizadas em duas partes relativas à descrição geral e à qualidade do teste. No primeiro bloco estiveram relacionadas informações pertinentes a nome, editora, ano de publicação, variável avaliada, entre outras. Já o segundo relacionou informações sobre a construção e os parâmetros psicométricos. Cada teste foi analisado e para o estabelecimento do gabarito de correção, consultou-se o manual do teste. Todas as avaliações efetuadas pelos juízes e avaliadores dos testes foram comparadas com o gabarito respectivo, para verificar o índice de acertos. Na avaliação do instrumento foram considerados como indicadores

positivos quando o avaliador respondeu "sim" para item atendido e quando o item não se aplica, sendo atribuído valor 2. Para as respostas assinaladas com "não" foi atribuído valor zero. Quando o avaliador não respondeu, isto é, não assinalou nenhuma das alternativas para o item, foi atribuído valor 1, significando que nada se pode afirmar, nem que errou e nem que acertou a resposta (entre zero = não atende e 2 = atende).

Resultados

A avaliação da qualidade dos instrumentos estudados permitiu algumas análises. A correção das respostas dos avaliadores em razão do gabarito estabelecido pelas autoras foi investigada por meio do coeficiente de correlação de Pearson. Os resultados são apresentados nas Tabelas 1 a 8, sendo uma por instrumento, e representam as avaliações realizados pelo Grupo 1 (psicólogos).

Tabela 1 – Coeficiente de correlação de Pearson entre os avaliadores psicólogos e o gabarito no teste BPR-5

		Correlação	Sig.
Par 1	Gabarito BPR-5 e Avaliador 1	0,688	0,000
Par 2	Gabarito BPR-5 e Avaliador 2	0,889	0,000
Par 3	Gabarito BPR-5 e Avaliador 3	0,757	0,000
Par 4	Gabarito BPR-5 e Avaliador 4	0,561	0,000
Par 5	Gabarito BPR-5 e Avaliador 5	0,211	0,155
Par 6	Gabarito BPR-5 e Avaliador 6	0,680	0,000

Os resultados de avaliação do teste BPR-5 são apresentados na Tabela 1. As correlações encontradas foram significativas, exceção feita à correção do avaliador 5. As análises dos demais avaliadores se correlacionaram positivamente com o gabarito, com coeficientes entre 0,561 (4º avaliador) a 0,889 (2º avaliador).

A análise da Tabela 2 indica os resultados da correção do TDE. Nesse sentido, todas as correlações foram significativas, exceto a do avaliador 1. As correlações variaram entre 0,293 a 0,895, embora nesse teste apenas a correlação de um avaliador com o gabarito tenha sido superior a 0,50, revelando menos acertos dos avaliadores.

Tabela 2 – Coeficiente de correlação de Pearson entre os avaliadores psicólogos e o gabarito no teste TDE

		Correlação	Sig.
Par 1	Gabarito TDE e Avaliador 1	0,205	0,166
Par 2	Gabarito TDE e Avaliador 2	0,895	0,000
Par 3	Gabarito TDE e Avaliador 3	0,449	0,002
Par 4	Gabarito TDE e Avaliador 4	0,342	0,019
Par 5	Gabarito TDE e Avaliador 5	0,293	0,045
Par 6	Gabarito TDE e Avaliador 6	0,298	0,042

Correlação Siq. Par 1 Gabarito IFP e Avaliador 1 0,290 0,048 Par 2 0,439 0.002 Gabarito IFP e Avaliador 2 Par 3 Gabarito IFP e Avaliador 3 0.090 0,546 Par 4 Gabarito IFP e Avaliador 4 0,224 0,130 Par 5 Gabarito IFP e Avaliador 5 0,138 0,356 Par 6 Gabarito IFP e Avaliador 6 0,171 0.251

Tabela 3 – Coeficiente de correlação de Pearson entre os avaliadores psicólogos e o gabarito no teste IFP

Os resultados da correção do teste IFP que se encontram na Tabela 3 indicam que apenas os coeficientes de Pearson relacionados aos avaliadores 1 e 2 são estatisticamente significativos e, portanto, se correlacionam com o gabarito do teste IFP. Diferentemente dos outros testes analisados, esse não apresentou correlações superiores a 0,50. Nesse teste foram apresentados comentários dos avaliadores nos formulários sobre as informações no manual: "O

autor não apresenta o teste como adaptado, mas no decorrer do manual o cita como versão brasileira". "Na ficha catalográfica o autor diz que o teste foi originado (porque não cita adaptação), porém na página 33 do manual ele se reporta que o teste nasceu de uma adaptação modificada do teste EPPS." Essas observações justificam a possibilidade de ter havido confusão no direcionamento das respostas.

Tabela 4 - Coeficiente de correlação de Pearson entre os avaliadores psicólogos e o gabarito no teste 16PF

		Correlação	Sig.	
Par 1	Gabarito 16PF e Avaliador 1	0,174	0,242	
Par 2	Gabarito 16PF e Avaliador 2	0,079	0,599	
Par 3	Gabarito 16PF e Avaliador 3	0,183	0,219	
Par 4	Gabarito 16PF e Avaliador 4	0,821	0,000	
Par 5	Gabarito 16PF e Avaliador 5	0,353	0,015	
Par 6	Gabarito 16PF e Avaliador 6	0,232	0,117	

A análise da Tabela 4 indica que os avaliadores 4 e 5 obtiveram coeficientes de correlação de Pearson significativos com o gabarito referente à análise da

qualidade do 16PF. Os coeficientes variaram entre 0,079 e 0,821, sendo que a mais alta correlação entre gabarito—correção foi relativa ao avaliador 4.

Tabela 5 – Coeficiente de correlação de Pearson entre os avaliadores alunos da pós-graduação e o gabarito no teste BPR-5

		Correlação	Sig.
Par 1	Gabarito BPR-5 e Avaliador 1	0,683	0,000
Par 2	Gabarito BPR-5 e Avaliador 2	0,555	0,000
Par 3	Gabarito BPR-5 e Avaliador 3	0,565	0,000
Par 4	Gabarito BPR-5 e Avaliador 4	0,688	0,000
Par 5	Gabarito BPR-5 e Avaliador 5	0,644	0,000
Par 6	Gabarito BPR-5 e Avaliador 6	0,540	0,000
Par 7	Gabarito BPR-5 e Avaliador 7	0,126	0,397
Par 8	Gabarito BPR-5 e Avaliador 8	0,766	0,000
Par 9	Gabarito BPR-5 e Avaliador 9	0,180	0,225
Par 10	Gabarito BPR-5 e Avaliador 10	0,547	0,000
Par 11	Gabarito BPR-5 e Avaliador 11	0,736	0,000
Par 12	Gabarito BPR-5 e Avaliador 12	0,780	0,000
Par 13	Gabarito BPR-5 e Avaliador 13	0,544	0,000
Par 14	Gabarito BPR-5 e Avaliador 14	0,807	0,000

No que se refere à avaliação realizada pelo grupo 2, os resultados revelaram que aproximadamente 86% (f=12) dos avaliadores obtiveram resultados estatisticamente significativos na comparação com o gabarito do teste BPR-5, embora os avaliadores 7 e 9 tenham apresentado resultados não significativos. Ainda nesse sentido, o menor coeficiente encontrado nesse instrumento foi

superior a 0,50. Os dados podem ser visualizados na Tabela 5.

Os resultados apresentados da Tabela 6 são referentes à correção do TDE. Nesse sentido os dados revelaram que 78,6% (f=11) dos avaliadores obtiveram resultados significativamente correlacionados com o gabarito do teste TDE, sendo que as correlações significativas variaram entre 0,31 a 0,86.

Tabela 6 – Coeficiente de correlação de Pearson entre os avaliadores alunos da pós-graduação e o gabarito no teste TDE

		Correlação	Sig.
Par 1	Gabarito TDE e Avaliador 1	0,345	0,018
Par 2	Gabarito TDE e Avaliador 2	0,579	0,000
Par 3	Gabarito TDE e Avaliador 3	0,319	0,029
Par 4	Gabarito TDE e Avaliador 4	0,256	0,083
Par 5	Gabarito TDE e Avaliador 5	0,779	0,000
Par 6	Gabarito TDE e Avaliador 6	0,634	0,000
Par 7	Gabarito TDE e Avaliador 7	0,103	0,492
Par 8	Gabarito TDE e Avaliador 8	0,702	0,000
Par 9	Gabarito TDE e Avaliador 9	0,335	0,021
Par 10	Gabarito TDE e Avaliador 10	0,086	0,565
Par 11	Gabarito TDE e Avaliador 11	0,539	0,000
Par 12	Gabarito TDE e Avaliador 12	0,825	0,000
Par 13	Gabarito TDE e Avaliador 13	0,684	0,000
Par 14	Gabarito TDE e Avaliador 14	0,457	0,001

Tabela 7 – Coeficiente de correlação de Pearson entre os avaliadores alunos da pós-graduação e o gabarito no teste IFP

		Correlação	Sig.
Par 1	Gabarito IFP e Avaliador 1	0,518	0,000
Par 2	Gabarito IFP e Avaliador 2	0,228	0,123
Par 3	Gabarito IFP e Avaliador 3	0,260	0,078
Par 4	Gabarito IFP e Avaliador 4	0,234	0,114
Par 5	Gabarito IFP e Avaliador 5	0,474	0,001
Par 6	Gabarito IFP e Avaliador 6	0,349	0,016
Par 7	Gabarito IFP e Avaliador 7	0,425	0,003
Par 8	Gabarito IFP e Avaliador 8	0,404	0,005
Par 9	Gabarito IFP e Avaliador 9	0,356	0,014
Par 10	Gabarito IFP e Avaliador 10	0,451	0,001
Par 11	Gabarito IFP e Avaliador 11	0,597	0,000
Par 12	Gabarito IFP e Avaliador 12	0,644	0,000
Par 13	Gabarito IFP e Avaliador 13	0,246	0,096
Par 14	Gabarito IFP e Avaliador 14	0,284	0,053

A Tabela 7 refere-se à análise da qualidade do IFP. Os achados indicaram que 9 coeficientes são significativos, embora as correlações tenham sido menores do as avaliações dos outros instrumentos, variando entre 0,24 e 0,59. O instrumento avaliado pelo grupo 2 foi o 16PF, cujos resultados se encon-

tram na Tabela 8. Dos 14 coeficientes de correlação de Pearson, 10 se revelaram estatisticamente significativos, sendo que os resultados dos avaliadores 2, 3, 5 e 14 são os não significativos. As correlações nesse caso não foram muito altas, variando entre 0,30 e 0,55.

Tabela 8 – Coeficiente de correlação de Pearson entre os avaliadores alunos da pós-graduação e o gabarito no teste 16PF

		Correlação	Sig.
Par 1	Gabarito 16PF e Avaliador 1	0,492	0,000
Par 2	Gabarito 16PF e Avaliador 2	0,266	0,071
Par 3	Gabarito 16PF e Avaliador 3	-0,142	0,340
Par 4	Gabarito 16PF e Avaliador 4	0,404	0,005
Par 5	Gabarito 16PF e Avaliador 5	-0,010	0,948
Par 6	Gabarito 16PF e Avaliador 6	0,511	0,000
Par 7	Gabarito 16PF e Avaliador 7	0,557	0,000
Par 8	Gabarito 16PF e Avaliador 8	0,529	0,000
Par 9	Gabarito 16PF e Avaliador 9	0,557	0,000
Par 10	Gabarito 16PF e Avaliador 10	0,334	0,022
Par 11	Gabarito 16PF e Avaliador 11	0,383	0,008
Par 12	Gabarito 16PF e Avaliador 12	0,301	0,040
Par 13	Gabarito 16PF e Avaliador 13	0,301	0,040
Par 14	Gabarito 16PF e Avaliador 14	0,252	0,088

Tabela 9 – Porcentagem dos coeficientes de correlação de Pearson significativos entre os avaliadores psicólogos e os avaliadores alunos da pós-graduação e com o gabarito por teste

Testes	Porcentagem de coer correlação significativ gabarito (p<0,05%)		Porcentagem de coeficientes de correlação significativos entre avaliac gabarito (p<0,05%)		
	Psicólogos	Pós-graduados	Psicólogos	Pós-graduados	
BPR-5	83,3	86	17,7	14	
TDE	83,3	78,6	17,7	21,4	
IFP	33,3	64,3	67,7	35,7	
16PF	33,3	71,4	67,7	28,6	

A comparação entre grupos, no que se refere aos coeficientes de correlação entre a correção e o gabarito, pode ser encontrada na Tabela 9, que por sua vez, apresenta um resumo do percentual dos coeficientes de correlação. Em apenas um teste (TDE) o grupo de psicólogos pós-graduados (Grupo 2) obteve menos coeficientes significativos do que o grupo de

psicólogos (Grupo 1). Ainda nesse sentido, considerando as correlações superiores a 0,50, em relação ao BPR-5 o Grupo 1 apresentou 83,3%, enquanto o Grupo 2, 86%; no que diz respeito ao TDE, no Grupo 1 foi encontrada apenas uma correlação superior a 0,50 (17,7%), enquanto no Grupo 2, 50%. A correção do IFP pelo Grupo 1 em comparação ao gabarito não

gerou nenhuma correlação superior a 0,50, enquanto no Grupo 2 foram encontradas 3 (21,4%), e, por último, a avaliação do 16PF resultou em 17,7% de coeficientes superiores a 0,50, já no Grupo 2, 28,6%.

Considerações finais

Considerando o objetivo do presente estudo, ou seja, o de avaliar a análise da qualidade de testes promovida por psicólogos, assim como determinar o consenso entre eles no que se refere à análise, os resultados permitiram reflexões interessantes. A análise da qualidade de testes, disparada oficialmente pelo Conselho Federal de Psicologia em 2001, embora em um primeiro momento estivesse restrita a especialistas na área de avaliação e testagem psicológicas, em longo prazo deve ser uma prática incorporada ao trabalho do psicólogo que faz uso de instrumentos padronizados. Compete ao profissional estabelecer seus critérios de testagem, e em razão disso estabelecer os instrumentos que estejam adequados às necessidades da população, da queixa ou do construto avaliado, entre outros fatores importantes.

Sob essa perspectiva o trabalho de "análise de testes" deve ser administrado pelos próprios psicólogos constantemente. Assim sendo, espera-se que esse profissional possua os domínios necessários para realizar tais análises, o que exigiria dele clareza dos princípios descritos nos manuais dos testes psicológicos, que devem conter informações sobre a fundamentação teórica dos instrumentos, descrição da aplicação, avaliação e interpretação dos resultados e indicação da literatura científica utilizada.

Com o propósito de determinar se o nível de conhecimento dos participantes (psicólogo ou psicólogo pósgraduado) poderia gerar melhores avaliações dos testes, realizou-se o estudo com aluno da pós-graduação, a fim de testar a diferença entre eles. O resultado encontrado indicou que de maneira geral o Grupo 2 (pósgraduação) apresentou mais acertos em relação ao gabarito preparado pelas autoras do que o Grupo 1 (psicólogos), o que reafirma a asserção amplamente discutida por diferentes autores (Almeida, 1999; Noronha, 1999; Pasquali, 1999; Sisto, Sbardelini & Primi, 2001), no que se refere à necessidade de aprimoramento profissional e de formações mais consistentes.

Ao lado disso, o movimento pela busca da qualidade dos testes psicológicos não pode ser desvencilhado da própria formação em psicologia. De alguma forma, os resultados desse trabalho revelaram que há necessidade de implantação de melhor compreensão dos conceitos de qualidade de um teste entre os psicólogos (Noronha, 1999; Noronha, Freitas, Sartori & Ottati, 2001, 2002a, 2002b).

Embora o presente estudo não tenha tido a pretensão de identificar a relação de domínio dos conceitos dos participantes apresentados no instrumento nem verificar o quanto cada um deles compreende a tarefa de análise da qualidade de instrumentos, os achados revelaram alguma dificuldade na execução da tarefa. Isso foi revelado pelos baixos coeficientes de correlação encontrados entre gabaritos e avaliações.

À guisa de considerações finais, registra-se a importância da colaboração dos psicólogos que participaram da pesquisa. Houve diversas manifestações deles acerca do aprendizado adquirido com as repetidas investigações sobre a qualidade dos testes psicológicos avaliados, bem como a sugestão, por parte dos que exercem a docência, de que o instrumento seja usado no ensino das disciplinas de avaliação psicológica. Sugere-se que novos estudos possam ser desenvolvidos com outros instrumentos e grupos de participantes.

Referências

Adánez, G. P. (1999). Procedimientos de construcción y análisis de tests psicométricos. Em S. M. Wechsler & R. S. L. Guzzo (Orgs.). *Avaliação psicológica: perspectiva internacional (pp. 57-100)*. São Paulo: Casa do Psicólogo.

Almeida, L. S. (1999). Avaliação psicológica: exigências e desenvolvimento nos seus métodos. Em S. M. Wechsler & R. S. L Guzzo, (Orgs.). Avaliação psicológica: perspectiva internacional (pp. 41-55). São Paulo: Casa do Psicólogo.

Almeida, L. S., Prieto, G., Muñiz, J. & Bartram, D. (1998). O uso dos testes em Portugal, Espanha e países Ibero-Americanos. *Psychological*, *20*, 27-40.

American Educational Research Association (AERA), American Psychology Association (APA) & National Council on Measurement in Education (NCME) (1999). Standards for Educational and Psychological Testing. New York: American Educational Research Association.

Azevedo, M. M., Almeida, L. S., Pasquali, L. & Veiga, H. M. S. (1996). Utilização dos testes psicológicos no Brasil: dados de estudo preliminar em Brasília. Em L. S. Almeida, S. Araújo, M. M. Gonçalves, C. Machado & M. R. Simões (Orgs.). *Avaliação psicológica: formas e contextos* (Vol. 4, pp. 213-219). Braga, Portugal.

- Bartram, D. (1998). The need for International Guideliness on Standards for Test use a review of European and International Initiatives. *European Psychologist*, *3*(2), 155-163.
- Canadian Psychological Association CPA (1996). Guidelines for Educational and Psychological Testing. Ontario, Canadá: CPA. Disponível em: http://www.cpa.ca/guide9.html
- Cattel, R. B. & Cattel, A. K. S. (2000). *16 PF Maual Técnico*, *5ª Edição*. Rio de Janeiro: CEPA.
- Conselho Federal de Psicologia (2003). Resolução no 02/2003. Disponível em: http://www.pol.org.br. Acessado em: 02/11/2003.
- Garb, H. N. (2000). Computers will become increasingly important for psychological assessment: not that there's anything wrong with that. *Psychological Assessment*, 12(1), 31-39.
- Halperin, J. M. & Mckay, K. E. (1998). Psychological testing for children and adolescent psychiatrists: a review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*(6), 575-584.
- Hinkle, D. E., Wiersma, W. & Jurs, S. G. (1994). *Applied statistics for the behavioral sciences*. Boston, USA: Houghton Mifflin Company.
- International Test Commission (ITC) (2000). Guidelines on adapting tests: International Test Commission. Disponível em: http://www.intestcom.org.
- Nick, E. (1988). Vivências relativas ao trabalho em avaliação psicológica: dificuldades, limites e perspectivas para o Brasil. *Anais do XVIII. Reunião Anual de Psicologia*. Ribeirão Preto, SBPRP.
- Noronha, A. P. P. (1999). *Avaliação psicológica segundo psicólogos: usos e problemas com ênfase nos testes*. Tese de Doutorado. PUCCamp, Campinas.
- Noronha, A. P., Freitas, F. A., Sartori, F. A. & Ottati, F. (2001). Informações contidas nos manuais de testes de inteligência publicados no Brasil. *Psicologia em Estudo*, 6(2), 101-106.
- Noronha, A. P., Sartori, F. A., Freitas, F. A. & Ottati, F. (2002a). Informações contidas nos manuais de testes de personalidade. *Psicologia em Estudo*, 7(1), 143-149.
- Noronha, A. P., Freitas, F. A., Sartori, F. A. & Ottati, F. (2002b). Parâmetros psicométricos de testes de inteligência. *Interação em Psicologia, 6*(2), 195-201.
- Noronha, A. P. P., Vendramini, C. M. M., Canguçu, C., Souza, C. V. R., Cobêro, C., Paula, L. M., Lima, O. M. P., Guerra, P. B. C. & Fillizati, R. (2003). Propriedades psicométricas

- apresentadas em manuais de testes de inteligência. *Psicologia em Estudo, 8*(1), 89-95.
- Oakland, T. (1996). Qualities that will influence test in and assessment practice with children and youth toward the beginning of the twenty-first century: international perspectives. *Psicologia Escolar e Educacional, 1*(1), 11-18.
- Oakland, T. (1999). Developing standardized tests. In S. M. Wechsler & R. S. L Guzzo, (Org.) *Avaliação psicológica: perspectiva internacional (pp. 101-118)*. São Paulo: Casa do Psicólogo.
- Pasquali, L. (1996). Questões epistemológicas na medida psicológica. In L. S. Almeida, S. Araujo, M. M. Gonçalves,
 C. Machado & M. R. Simões (Orgs.), Avaliação psicológica: formas e contextos (Vol. 4). Braga, Portugal: Associação dos Psicólogos Portugueses.
- Pasquali, L. (1998). Princípios de elaboração de escalas psicológicas. *Revista de Psiquiatria Clínica*, 25(5), 206-213.
- Pasquali, L. (1999). (Org.). *Instrumentos psicológicos:* manual prático de elaboração. Brasília: LabPAM/IBAP.
- Pasquali, L., Mazzarello, M. A. & Ghesti, I. (1997). *Inventário Fatorial de Personalidade*. São Paulo: Casa do Psicólogo.
- Primi, R. & Almeida, L. S. (2000). *Bateria de Provas de Raciocínio BPR-5*. São Paulo: Casa do Psicólogo.
- Russell, M. T. & Karol, D. (2000). *Dezesseis PF:* questionário dos 16 fatores de personalidade. Rio de Janeiro: CEPA.
- Sisto, F. F., Codenotti, N., Costa, C. A. J. & Nascimento, T. C. N. (1979). Testes psicológicos no Brasil: que medem realmente. *Revista Educação e Sociedade, 2*, 152-165. São Paulo: Cortez.
- Sisto, F. F., Sbardelini, E. T. B. & Primi, R. (2001). *Contextos* e questões da avaliação psicológica. São Paulo: Casa do Psicólogo.
- Stein, L. M. (1994). *Teste de Desempenho Escolar TDE*. São Paulo: Casa do Psicólogo.
- Turner, S. M., DeMers, S. T., Fox, H. R. & Reed, G. M. (2001). APA's Guideliness for Test user qualifications an executive summary. *American Psychologist*, *56*(12), 1099-1113.

Recebido: abril/2005 Reformulado: junho/2005 Aprovado: junho/2005

Sobre as autoras:

Orlete Maria Pompeu de Lima é psicóloga, mestre em Psicologia pela Universidade São Francisco e docente da FMA e Universidade do Paraná (UNIPAR).

Ana Paula Porto Noronha é psicóloga, mestre e doutora em Psicologia pela Pontificia Universidade Católica de Campinas, docente do Programa de Estudos Pós-Graduados em Psicologia da Universidade São Francisco e bolsista de Produtividade em Pesquisa do CNPq.

Anexo

INS	STRUMENTO DE AVALIAÇÃO DA QUAL ID	ADE DE	TESTE	S PSIC	OLÓGICOS
		Este instrumento tem por objetivo relacionar um conjunto de aspectos básicos que os gigoso devem possuir, de acordo com os parâmetros nacionais e internacionais, para qua lo por Psicólogos, facilitando a escolha de instrumentos apropriados para seus objetivo. ES: Avalie o manual e os materiais do teste e assinale com um "X" na coluna nte. CERIÇÃO GERAL DO TESTE: ATENDIDOS ASPECTOS AVALIADOS AITIMANO NÃO SIM OBIS. APPLICADO 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2			
	Psicológico:				
Data d	a avaliação:				
Avalia	dor:				
estes p	sicológicos devem possuir, de acordo com os parân reciado por Psicólogos, facilitando a escolha de ins	metros nac	ionais e	internaci	onais, para qu
	UÇÕES: Avalie o manual e os materiais do teste e ondente.	assinale co	m um '	'X" na col	luna
	DESCRICÃO CERAL DO TESTE.		ATENI	amas	
1 ITEM		APLICADO	NÃO	SIM	OBS.
1.1.	O teste apresenta nome?	,	_	- 2	
1.2.	Apresenta nome do(s) autor(es)?				
1.3.	Apresenta nome do editor?				
1.4.	Apresenta nome e número do CRP do Psicólogo responsável?				
1.5.	No caso de teste adaptado, apresenta nome do teste na versão original?				
1.6.	Apresenta nome do(s) autor(es) da adaptação?				
1.7.	Apresenta nome do editor da adaptação?				
1.8.	Informa data de publicação?				
1.9.					
1.10.	medir?				
1.11.	identificada(s)?				
1.12.					
1.13.					
1.14.	Informa o tipo de aplicação? (Administração oral,				
1.15.	Descreve as características dos sujeitos nos quais o teste deve ser aplicado? (Idade, nível de				
TEM	ASPECTOS AVALIADOS	APLICADO			OBS.
1.16.	Indica se existem formas diferentes do teste?				
1.17.	Descreve os procedimentos de correção? (Manual mediante crivo, leitura ótica,				
	informatizado, corrigido pela editora, corrigido por				

	especialistas e/ou outros.)				
1.18.	Descreve o procedimento para obter os escores?				
1.19.	Descreve o procedimento para interpretação dos				
	escores? (Normatizada ou não normatizada.)				
1.20.	Apresenta escalas utilizadas para interpretação dos				
	escores?				
1.21.	Informa o tempo estimado para a aplicação individual e /ou coletiva do teste?				
1.22.	Apresenta fundamentação teórica nacional e/ou				
1 22	internacional sobre o construto?				
1.23	Apresenta fundamentação teórica nacional e/ou				
	internacional sobre procedimentos de medição do construto?				
1.24.	Informa as referências bibliográficas utilizadas?				
1.24.	Indica literatura científica complementar				
1.20.	relacionada ao teste?				
A	SUBTOTAL DE PONTOS	البرحا			
Α.	SCHIOTALDETONIOS				
2	QUALIDADES DO TESTE		ATENI	DIDOS	
ITEM	ASPECTOS AVALIADOS	ITEM NÃO	NÃO	SIM	OBS.
		APLICADO 0	1	2	
2.1.	Apresenta qualidade (atrativo e eficiente) no				
	material do teste (impressos e/ou software e outros objetos)?				
2.2.	A descrição das características técnicas está				
	fundamentada em dados e referências científicas?				
2.3.	A fundamentação teórica do construto que se				
	pretende medir está clara e documentada?				
2.4.	O procedimento de medição do construto está claro				
	e documentado?				
2.5.	No caso de adaptação do teste, existe descrição				
	precisa do procedimento de tradução, com estudos				
	de equivalência com a versão original e utilização de normas internacionais?				
2.6.	As instruções são claras, precisas e adequadas para				
2.0.	a população a que se dirige o teste?				
ITEM	ASPECTOS AVALIADOS	ITEM NÃO	NÃO	SIM	OBS.
	nor be room membros	APLICADO 0	1	2	
2.7.	A tarefa a ser executada no teste é clara?				
2.8.	Existe facilidade para registrar as respostas?				
2.9.	Existe informação detalhada de estudos sobre				
	validade dos itens: dificuldade ou variabilidade,				
	discriminação, validade, distratores e/ou outros?				
2.10.	Apresenta estudos de validade para as variáveis				
	que o teste se propõe a medir?				
2.11.	Validade de conteúdo. Apresenta uma precisa				
	definição do conteúdo?				
2.12	Os itens representam adequadamente todas as	1			

2.13.	Informa se há consulta a um número de						
	especialistas maior que 06, mediante um						
	procedimento sistematizado?						
2.14.	Apresenta formas de validade de construto:						
	correlação com outros testes, diferenças entre						
	grupos, matriz multitraço-multimétodo, análise						
	fatorial exploratória, análise fatorial confirmatória						
	e/ou formas experimentais?						
2.15.	Descreve os critérios empregados para validade de						
2.16.	critério e as características das populações?						
2.16.	Apresenta formas de validade com referência ao critério: Concorrente e/ou Preditiva?						
2.17.	Apresenta dados de estudos sobre a precisão:	┢				-	
2.17.	Apresenta dados de estudos sobre a precisao: equivalência (formas paralelas), consistência						
	interna, estabilidade (teste-reteste)?			1			
2.18.	Os coeficientes de equivalência apresentam	H		 		+	
2.10.	mediana acima de 0,60 ?						
2.19.	Os coeficientes de consistência apresentam	Т					
	mediana acima de 0,60 ?			1			
2.20.	Os coeficientes de estabilidade apresentam						
	mediana acima de 0,60 ?						
2.21.	Apresenta tabelas das normas em função da idade,						
	sexo, região, nível cultural e socioeconômico e						
	outras características relevantes?						
2.22.	Tamanho das amostras. O número de sujeitos do						
-	estudo de padronização está acima de 300?						
В	SUBTOTAL DE PONTOS						
		Т	MÁ	KIMO	тот	AL DE	%
			DE PO	ONTOS			,,,
	RESULTADO DA AVALIAÇÃO		10	0%		ESTE	
1. DES	CRIÇÃO GERAL DO TESTE				A =		
OLL	ALIDADES DO TESTE	-					
QO1	LEID/IDES DO TESTE				B =		
Comer	ntários :						
	<u> </u>						
	<u> </u>						