

ESTUDOS DE VALIDADE E DE PRECISÃO EM TESTES DE INTELIGÊNCIA¹

*Ana Paula Porto Noronha²
Universidade São Francisco*

Resumo: O presente estudo abordou a construção dos testes psicológicos, em especial a padronização e os estudos de validade e precisão. O objetivo da pesquisa foi verificar a presença ou a ausência de validade e precisão nos testes de inteligência, bem como identificar se os testes estrangeiros possuem padronização brasileira. Foram analisados 17 testes de inteligência comercializados no Brasil, e os resultados indicaram que grande parte dos testes nacionais e dos internacionais apresenta estudos de validação e de precisão, embora nem todos tenham indicado nos respectivos manuais, os dados referentes à padronização, como amostra de padronização e variáveis relacionadas. Sugere-se a realização de outros estudos com diferentes testes ou novas análises sobre os testes de inteligência. Conclui-se que embora os autores concordem que os testes devam realizar estudos de verificação dos parâmetros psicométricos e que devem possuir normas regionais, tal prática ainda não se encontra totalmente difundida na avaliação psicológica brasileira.

Palavras-chave: parâmetros psicométricos; inteligência; testes psicológicos.

VALIDITY AND RELIABILITY STUDIES OF INTELLIGENCE TESTS

Abstract: The aim of the present study was to analyze the construction of psychological tests, focusing on standardization and studies of validity and reliability. The objective of the research was to verify whether the tests of intelligence are valid and reliable or not, as well as to identify if the foreign tests present Brazilian standardization. Seventeen tests of intelligence commercialized in Brazil were analyzed and the results indicated that most tests, local and foreign ones, present studies of validation and reliability, although some of them have not mentioned information concerning standardization in their respective manuals, as a sample of standardization and related variables. New studies are recommended, focusing on different tests or new analysis of the tests of intelligence. It is concluded that although the authors agree that all tests need studies to verify psychometric parameters and studies to obtain regional norms this action isn't spread yet in the Brazilian psychological assessment.

Key-words: psychometric parameters; intelligence; psychological tests

A construção de testes psicológicos vem sendo bastante questionada no Brasil atualmente. As Resoluções n^{os} 25/2001 e 02/2003 promulgadas pelo Conselho Federal de Psicologia propõem que os instrumentos passem por um processo de avaliação, de forma que gradualmente sejam mantidos no mercado profissional apenas os que possuem as condições técnicas para tal. Por condições técnicas entende-se a padronização com amostras nacionais e estudos de determinem os parâmetros psicométricos, ou seja, evidências de validade e de precisão.

Acredita-se que as resoluções antes mencionadas, embora tardiamente, e ainda levando em consideração que a avaliação não é uma atividade profissional recente, poderão organizar sobremaneira a distribuição de materiais psicológicos no mercado profissional, de forma que maus instrumentos não serão mais reeditados caso não superem as dificuldades técnicas apresentadas. Se por um lado, todo o processo desenvolvido pelo CFP causa certo desconforto, por outro, vem com o intuito de oferecer materiais mais adequados para uma avaliação psicológica com qualidade, que promova a Psicologia e seus profissionais.

A realidade internacional em muito se diferencia da brasileira, considerando que os padrões para

¹ Artigo recebido para publicação em 22/05/03; aceito em 06/08/03.

² Endereço p/ correspondência: Rua Alexandre Rodrigues Barbosa, 45, Itatiba, SP, Cep 13251-900, E-mail: ana.noronha@saofrancisco.edu.br

construção, revisão e adaptação de instrumentos de medida já se encontram devidamente estabelecidos e em uso há anos, tendo passado, inclusive, por revisões. Ainda nesta perspectiva dos padrões de excelência para os instrumentos, recentemente foi publicada uma revisão dos *Standards for Educational and Psychological Tests* (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999), uma referência clássica quanto aos parâmetros dos instrumentos utilizados na avaliação psicológica. Outro trabalho de referência é o *Guidelines on Adapting Tests e Guidelines on Test Use* da International Test Commission (ITC, 2001). Basicamente essa literatura indica que os instrumentos devem possuir uma fundamentação teórica, parâmetros psicométricos estabelecidos, um sistema de correção e interpretação dos escores, uma descrição clara dos procedimentos de aplicação e correção e um manual que contemple as informações citadas. Atualmente os instrumentos comercializados estão em análise para verificar quais atendem aos requisitos necessários.

No que se refere aos conceitos de validade e de precisão de testes psicológicos, vale destacar que eles são elementos indispensáveis para garantir a confiabilidade dos resultados das aplicações. Embora a importância da presença de estudos sobre a validade e precisão de instrumentos não seja novidade na história da Psicometria, é possível observar muitos testes que se abstém deste tipo de verificação.

Vários são os trabalhos teóricos de autores nacionais e internacionais que se destinam à discussão dos parâmetros psicométricos, considerando que um teste para ser confiável e recomendável deve apresentar os estudos de verificação de suas qualidades. Para aprofundamento do conceito, examinar as seguintes referências: Anastasi e Urbina (2000), Ancona-Lopez (1987), Cronbach (1996), Erthal (1987) e Pasquali (2001).

A validade tem sido entendida como a verificação do que o teste avalia e se ele realmente mede aquilo que se propõe (Pasquali, 2001). Como se vê, a ausência de estudos neste sentido não garante que os resultados obtidos sejam relativos ao construto que se estimou medir, o que em determinadas situações, se não em todas, pode inviabilizar a análise realizada.

Para a verificação da precisão, vale o mesmo pressuposto anterior, ou seja, considerando-a como a constância dos resultados dos sujeitos em diferentes situações de aplicação ou com aplicações advindas de testes similares (Anastasi & Urbina, 2000), a ausência de estudos desta natureza impede a verificação da manutenção dos resultados e da precisão do teste.

No que se refere às pesquisas recentes que visam analisar a qualidade dos instrumentos psicológicos, no Brasil, apenas recentemente começaram a aparecer trabalhos que avaliaram testes psicológicos aqui comercializados. Noronha (2001), em seu trabalho sobre a análise de coeficientes de 21 testes de inteligência, encontrou que dos instrumentos analisados aproximadamente 40% não possuía estudo de validade e/ou de precisão, quer fossem os desenvolvidos com os próprios instrumentos quer numa padronização brasileira. O estudo de validade mais presente foi o de construto, especialmente no que diz respeito à comparação dos resultados com outros testes. Já quanto à precisão, os coeficientes de consistência e de estabilidade foram os mais encontrados.

Outro trabalho desta natureza, mas com uma amostra abrangente, foi desenvolvido por Noronha, Primi e Alchieri (sub) com 146 testes psicológicos. Os resultados indicaram que 59,7% possuíam padronização brasileira, 48,6% estudos de precisão e 52,7% de validade, não necessariamente com amostras brasileiras. Tais achados corroboram os dados anteriores e reafirmam a necessidade de maior investimento nesta área de pesquisa.

Noronha e Vendramini (no prelo) identificaram os parâmetros psicométricos em 43 testes, sendo 22 de inteligência e 21 de personalidade. Os testes foram comparados quanto ao período de publicação no Brasil. Os dados revelaram que as décadas de 80 e 90 foram as de maior publicação e que os testes de inteligência apresentam mais estudos de padronização, validade e precisão do que os de personalidade, embora não tenha havido diferença significativa entre os grupos de testes (inteligência e personalidade).

Em perspectivas internacionais, estudos que evidenciem a qualidade dos instrumentos de avaliação são frequentes. Azar (1999) comenta que a revisão dos *Standards for Educational and Psychological*

Testing representam mais chances de maior crédito para os testes psicológicos. De acordo com a autora, os novos padrões já trazem questionamentos importantes até em relação a conceitos bastante difundidos, como é o caso da validade. McCarty (1999) evidencia a necessidade de se diferenciar o teste científico do não científico. E nesse sentido, ele elogia o trabalho desenvolvido pelo Committee on Psychological Tests and Assessment, cujo objetivo é melhor informar as pessoas a respeito do que é um teste e do seu uso adequado.

Outros estudos relativos à verificação de validade e/ou precisão de instrumentos foram realizados por Ballantyne e Sattler (1991), Hansen e Scullard (2002), Mendonza, Stafford e Stauffer (2000), Ryan e Ward (1999), Ward e Ryan (1996). Os trabalhos tratam da verificação de parâmetros psicométricos em diferentes instrumentos, alguns ainda não traduzidos para o português, como o Leisure Interest Questionnaire e o Reporter's Test.

Para Strauss, Spreen e Hunter (2000) a revisão e a pesquisa de testes são bastante comuns na Psicologia, particularmente na Neurociência. Embora esta asserção não seja condizente com a realidade brasileira, esta é uma meta a ser seguida. Em seu trabalho os autores abordam alguns problemas relativos à operacionalização da revisão e exemplificam três possíveis condições para revisões de testes: idade dos testes, idade dos participantes e mudanças no formato dos testes.

Como se vê, há diferenças notórias entre o desenvolvimento nacional e internacional no que se refere à evolução da área de avaliação psicológica, especialmente quanto ao estabelecimento dos padrões psicométricos, tão necessários quando se pensa na cientificidade de testes que visem auxiliar na compreensão de fenômenos psicológicos. Em contrapartida, a realidade brasileira vem dando sinais de recuperação da dignidade de seus instrumentos de avaliação, à medida em que movimentos de associação, de órgãos de classe e instituições de ensino estabeleceram metas para o crescimento da área como um todo.

Objetivo

O presente estudo objetiva contribuir com a análise de parâmetros psicométricos de instrumentos de avaliação da inteligência, especialmente no

que se refere à verificação da existência de padronização brasileira para instrumentos estrangeiros e da existência de estudos de validade e de precisão. Embora outros trabalhos já tenham iniciado esta tarefa e o Conselho Federal estabelecido metas em curto prazo para a verificação da qualidade de testes, acredita-se que outras pesquisas possam, de alguma forma, trazer contribuições.

Método

Instrumentos Consultados

Foram consultados 17 instrumentos de avaliação da inteligência, a saber: Bateria de Provas de Raciocínio BPR-5 (Primi & Almeida, 1998), Teste Coletivo de Inteligência para Adultos - Forma I.C.I.A. (Moraes, Andrade & Godoy, sem data), Matrizes Progressivas – escala avançada (Campos, 2002), Matrizes Progressivas – escala geral (Cunha, 2001), Matrizes Progressivas Coloridas (Angelini, Alves, Custódio & Duarte, 1999), BTAG (Braga, 1993), Teste Edites de Inteligência (Piovani & Piovani, sem data), TACOM A – B, TADIS 1 – 2, TADIM 1- 2 (Tonglet, 1999), TACOM C – D (Tonglet, 2002), TEMPLAM (Tonglet, 2000) e TRAP 1 (Tonglet, 2001)³.

Outra observação necessária refere-se aos instrumentos consultados. Sabe-se que eles não representam a totalidade dos comercializados no Brasil, no que diz respeito à avaliação da inteligência, mas em contrapartida os aqui referenciados representam mais de 50% dos disponíveis. A escolha destes em detrimento de outros justifica-se pelo fato de estes não terem sido ainda investigados.

Critérios Avaliados

Os elementos avaliados em cada instrumento foram: origem da produção (nacional ou internacional); data da publicação original; existência, ou não, de padronização brasileira e respectiva data; precisão, validade e respectivos estudos desenvolvidos.

Cada manual foi avaliado individualmente e os dados foram posteriormente analisados.

³ Vale destacar que as informações que constam dos parênteses referem-se às informações pertinentes à publicação nacional, ou seja, a autoria não é necessariamente a original, podendo estar relacionada à tradução, adaptação ou padronização.

Resultados

Dos instrumentos avaliados, observou-se que 70,6% (F=12) referem-se à produção nacional, enquanto 29,4% (F=5) à internacional. Tal fato merece atenção, considerando que embora os dados sejam relativos a uma pequena parcela dos instrumentos disponíveis no mercado brasileiro (em torno de 11%), há um notório crescimento da produção nacional dos instrumentos nas últimas duas décadas, de acordo com estudos já realizados como o de Noronha e Vendramini (no prelo) e por Noronha, Primi e Alchieri (sub). Em comum, as pesquisas mostraram que aproximadamente 25% dos manuais não traz as datas de construção, aproximadamente 20% é da década de 70, 25% foi elaborado nos anos 80 e 90 e o restante nos demais períodos, variando de 1920 a 1969.

Confirmando os dados anteriores, a Tabela 1 apresenta as datas de publicação original dos instrumentos analisados, a maior parte da década de 90.

Tabela 1: Data de publicação original dos instrumentos

	Frequência	%	% casos válidos	% cumulativa
1938	1	5,9	7,7	7,7
1947	1	5,9	7,7	15,4
1998	1	5,9	7,7	23,1
1999	7	41,2	53,8	76,9
2000	1	5,9	7,7	84,6
2001	2	11,8	15,4	100,0
Total	13	76,5	100,0	
Sem data	4	23,5		
Total geral	17	100,0		

No que se refere à existência de padronização brasileira para instrumentos não produzidos no Brasil, os dados revelaram que dos 5 estrangeiros, apenas 20% (F=1) possuía estudos com amostras brasileiras (Matrizes Progressivas Coloridas), suficientes para uma padronização nacional. Vale ressaltar que alguns manuais citaram a existência de trabalhos, no entanto não apresentaram os respectivos resultados.

Tal fato merece considerações, uma vez que não poucas vezes autores clássicos (Anastasi & Urbina, 2000; Cronbach, 1996) evidenciaram que testes são amostras de comportamento, cujos

resultados quando referenciados a normas devem revelar as características de sua amostra de padronização para que as comparações dos indivíduos testados sejam minimamente aproximadas. A ausência de estudos nacionais para os testes criados em regiões distintas não oferece elementos confiáveis de análise.

Considerando os parâmetros psicométricos dos instrumentos em questão, os resultados indicaram a presença de estudos de precisão em 94,1% deles (F=16). Em relação à validade 88,2% (F=15) trouxeram os respectivos trabalhos. Os dados são bastante animadores, embora não estejam revelando as médias nacionais já apresentadas em outras pesquisas, tais como as desenvolvidas por Noronha, Primi e Alchieri (sub) e Noronha e Vendramini (no prelo). A primeira envolveu uma análise abrangente de testes, sendo que 48,6% deles apresentavam estudos de precisão, 52,7% de validade e 41,8% dos instrumentos apresentavam concomitantemente verificações da validade e da precisão. Tais dados encontram-se discutidos na introdução deste trabalho. As informações sobre os parâmetros dos instrumentos encontram-se na Tabela 2.

Tabela 2: Parâmetros psicométricos dos instrumentos

Instrumentos	Precisão	Validade
Teste Edites de Inteligência	1	1
BPR-5	1	1
BTAG	0	0
CIA	1	0
TACOM C	1	1
TACOM D	1	1
TACOM A	1	1
TACOM B	1	1
TADIS 1	1	1
TADIS 2	1	1
TADIM 1	1	1
TADIM 2	1	1
TEMPLAM	1	1
TRAP 1	1	1
Matrizes escala avançada	1	1
Matrizes escala geral	0	0
Matrizes coloridas	1	1

No que se refere aos métodos utilizados para verificação da precisão nos testes constatou-se que a maioria fez uso das formas paralelas (52,9%, F=9), enquanto teste-reteste, consistência interna e Kuder-Richardson foram citados uma vez cada (5,9%). Já

na verificação da validade, 93,3% (F=14) olharam o construto e 6,7% (F=1), o conteúdo. Vale ressaltar que as porcentagens são relativas ao número de estudos apresentados, o que não necessariamente ocorreu em todos os instrumentos (Tabela 3). Os dados encontram-se em concordância com o trabalho desenvolvido por Noronha (2001). Talvez a

justificativa para a alta incidência de estudos de construto encontre-se na própria definição dele. Pasquali (2001) enfatiza que o construto é considerado como a forma mais fundamental de validade dos instrumentos, pois se constitui na possibilidade de verificar a hipótese da legitimidade da representação comportamental do traço latente.

Tabela 3: Métodos de precisão e estudos de validade dos instrumentos psicológicos

Instrumentos	Método de Precisão	Estudos de Validade
Teste Edites de Inteligência	Formas paralelas	Construto
BPR-5	Metades e consistência interna	Construto e critério
BTAG	-	-
CIA	Metades e Kuder-Richardson	-
TACOM C	Formas paralelas	Construto
TACOM D	Formas paralelas	Construto
TACOM A	Formas paralelas	Construto
TACOM B	Formas paralelas	Construto
TADIS 1	Formas paralelas	Construto
TADIS 2	Formas paralelas	Construto
TADIM 1	Formas paralelas	Construto
TADIM 2	Formas paralelas	Construto
TEMPLAM	Metades	Construto
TRAP 1	Metades	Construto
Matrizes escala avançada	Teste-reteste	Construto
Matrizes escala geral	-	-
Matrizes coloridas	Metades	Construto

Como já afirmado anteriormente, a verificação dos parâmetros psicométricos tem sido cada vez mais exigida quando da construção ou revisão de testes psicológicos. No que se refere à validade, em âmbito internacional ela vem sendo repensada e até redefinida pelos Standards for Educational and Psychological Testing. Azar (1999) aponta que não se fala mais em três tipos de validade (construto, conteúdo e critério), e sim, em tipos de evidência de validade. Embora tal discussão ainda seja incipiente em nosso meio acadêmico e científico, é importante que profissionais e pesquisadores da área fiquem atentos às atualizações necessárias.

Considerações Finais

Este trabalho objetivou discutir os parâmetros psicométricos de testes de inteligência já analisados sob outras perspectivas. Os resultados foram positivos, embora infelizmente eles não retratem uma realidade mais abrangente, considerando que a precariedade de instrumentos de avaliação ainda é presente no Brasil.

Os achados desta pesquisa apontam para urgentes modificações na área de avaliação psicológica. É imprescindível que os maus instrumentos sejam revistos, assim como também é preciso repensar a formação do profissional de Psicologia, de modo que este tenha uma bagagem teórica mais consistente e faça um uso ético e consciente dos instrumentos. No entanto, a formação deve ser tema de outros trabalhos. Este em especial, procurou revelar a situação de materiais que muitas vezes são usados sem as ressalvas necessárias, no que se refere à confiabilidade de seus resultados.

Aqui não se defende o uso de testes ou de quaisquer outras técnicas de medida, mas que a Psicologia, como ciência, disponha de recursos de avaliação que garantam a confiabilidade de seu trabalho e que, conseqüentemente, divulguem positivamente a profissão na sociedade. No caso dos testes, a existência de estudos com amostras nacionais, as verificações da validade e da precisão são consideradas as condições mínimas para um uso seguro (American Educational Research Association,

American Psychological Association & National Council on Measurement in Education, 1999; Anastasi & Urbina, 2000; Ancona-Lopez, 1987; Erthal, 1987).

Acredita-se que as resoluções ora promulgadas pelo Conselho Federal de Psicologia, assim como o movimento pelo qual passa a avaliação psicológica nacional, possam sanar ou ao menos minimizar as dificuldades apresentadas na docência, na pesquisa, na clínica, na educação e em tantas outras áreas de atuação do psicólogo brasileiro. Vale lembrar que a avaliação brasileira apresenta sinais de crescimento, tais como a criação de um Instituto Brasileiro de Avaliação Psicológica, o reconhecimento de um mestrado na área, o aumento de publicações científicas sobre o tema, dentre outros fatos que ilustram as conquistas.

Espera-se que a realidade citada por Strauss, Spreen e Hunter (2000), no sentido de que a revisão e a pesquisa de testes são bastante comuns na Psicologia, possa vigorar também proximamente na realidade nacional. Sugere-se que novos estudos sejam desenvolvidos com outras amostras de instrumentos.

Referências Bibliográficas

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. New York: American Educational Research Association.
- Anastasi, A. & Urbina, S. (2000). *Testagem Psicológica*. Porto Alegre: Artes Médicas.
- Ancona-Lopez, M. (1987). *Avaliação da Inteligência I*. São Paulo: E.P.U.
- Angelini, A; Alves, I.C.B; Custódio, E.M.& Duarte, J.L. (1999). *Matrizes Progressivas Coloridas*. São Paulo: CETEPP.
- Azar, B. (1999). Changes will improve quality of tests. *APA Monitor Online*, 30(11). Disponível em <http://www.apa.org/monitor>. Consulta feita em 16/12/99.
- Ballantyne, A. O. & Sattler, J. M. (1991). Validity and reliability of the Reporter's Test with normally achieving and learn children. *Psychological Assessment*, 3(1), 60-67.
- Braga, R. (1993). *BTAG*. Rio de Janeiro: CEPA.
- Campos, F. (2002). *Matrizes Progressivas –escala avançada*. Rio de Janeiro: CEPA.
- CFP – Conselho Federal de Psicologia (2001). Resolução nº 25/2001. Disponível em <http://www.pol.org.br>. Consulta feita em 04/12/2001.
- CFP – Conselho Federal de Psicologia (2003). Resolução no 02/2003. Disponível em <http://www.pol.org.br>. Consulta feita em 03/05/2003.
- Cronbach, J. (1996). *Fundamentos da Testagem Psicológica*. Porto Alegre: Artes Médicas, 6ª edição.
- Cunha, S. E. (2001). *Matrizes Progressivas – escala geral*. Rio de Janeiro: CEPA.
- Erthal, T.C. (1987). *Manual de Psicometria*. Porto Alegre: Artes Médicas.
- Hansen, J. C. & Scullard, M. G. (2002). Psychometric evidence for the Leisure Interest Questionnaire and analyses of structure of leisure interests. *Journal of Counseling Psychology*, 49(3), 331-341.
- International Test Comition – ITC (2001). Guidelines on Adapting Tests e Guidelines on Test Use. Disponível em <http://www.cop.es/tests>. Consulta feita em 04/12/2001.
- McCarty, R. (1999). Testy Issues. *APA Monitor Online*, 30(11). Disponível em <http://www.apa.org/monitor>. Consulta feita em 16/12/2002.
- Mendonza, J.L., Stafford, K.L. & Stauffer, J. M. (2000). Large-sample confidence intervals for validity and reliability coefficients. *Psychological Methods*, 5(3), 356-369.
- Moraes, R; Andrade, E.M. & Godoy, D. (s/d). *Teste Coletivo de Inteligência para Adultos – forma I*. Rio de Janeiro: CEPA.
- Noronha, A.P.P. (2001). Análise de coeficientes de testes de inteligência. *Psico*, 32(2), 73-86.
- Noronha, A.P.P., Primi, R. & Alchieri, J.C. (sub).

- Parâmetros psicométricos: uma análise de testes psicológicos comercializados no Brasil. *Revista Psicologia: Ciência e Profissão*.
- Noronha, A.P.P. & Vendramini, C.M.M. (no prelo). Parâmetros psicométricos: estudo comparativo entre testes de inteligência e de personalidade. *Psicologia: reflexão e crítica*, 16(1).
- Pasquali, L. (2001). *Técnicas do Exame Psicológico – TEP manual* (volume I). São Paulo: Casa do Psicólogo.
- Piovani, C. & Piovani, C.M. (s/d). *Teste Edités de Inteligência*. São Paulo: Edites.
- Primi, R. & Almeida, L.S. (1998). *BPR-5*. São Paulo: Casa do Psicólogo.
- Ryan, J.J. & Ward, L.C. (1999). Validity, reliability and standards errors of measurement for two seven-subtest of the Wechsler Adult Intelligence Scale-III. *Psychological Assessment*, 11(2), 207-211.
- Strauss, E., Spreen, O. & Hunter, M. (2000). Implications of test revisions for research. *Psychological Assessment*, 12(3), 237-244.
- Tonglet, E. C. (1999). *Bateria de Funções Mentais para Motorista – Testes de Atenção- BFM1*. São Paulo: Vetor.
- Tonglet, E. C. (2000). *Bateria de Funções Mentais para Motorista – Teste de Memória- BFM2*. São Paulo: Vetor.
- Tonglet, E. C. (2001). *Bateria de Funções Mentais para Motorista – Teste de Raciocínio Lógico – BFM3*. São Paulo: Vetor.
- Tonglet, E. C. (2002). *Bateria de Funções Mentais para Motorista – BFM4*. São Paulo: Vetor.
- Ward, L.C. & Ryan, J.J. (1996). Validity and time savings in the selection of short forms of the Wechsler Adult Scale-Revised. *Psychological Assessment*, 8(1), 69-72.