

Editorial

A centralidade da justiça social na interpretação de testes psicológicos

Evandro Morais Peixoto , Ricardo Primi 
Universidade São Francisco, Campinas/SP, Brasil

A avaliação psicológica é uma prática bastante sensível da psicologia, pois seus resultados impactam diretamente a vida das pessoas e a tomada de decisões profissionais em diferentes esferas sociais. No Brasil, um país marcado por profundas desigualdades históricas, econômicas e raciais, refletir sobre a justiça na avaliação psicológica é uma necessidade ética e científica recorrente. Nesse sentido, um número especial da *Revista Avaliação Psicológica*, dedicado ao tema, representa uma oportunidade valiosa para fomentar discussões sobre as dimensões sociais e culturais envolvidas nos processos de avaliação psicológica e construção de instrumentos, promovendo uma prática mais equitativa e contextualizada, bem como uma consciência profissional que alinham a sofisticação técnica com consciência social.

A avaliação psicológica tem sido historicamente apresentada como uma prática científica neutra, por ser guiada por parâmetros objetivos e tecnicamente fundamentados, implicando objetividade, imparcialidade e ausência de valores subjetivos. Entretanto, precisamos nos deter mais profundamente nesse aspecto, já que a ideia de neutralidade pode, potencialmente, estar refletindo uma construção ideológica, isto é, uma justificativa aparentemente robusta, científica, de processos mais complexos sociais e históricos que refletiam intensões de diferentes naturezas no uso de instrumentos psicológicos e da avaliação psicológica, como a manutenção de certas hierarquias sociais e a justificativa de exclusão de acesso a recursos a certos grupos sociais. Nesse sentido, ao longo da história, especialmente nos campos da psicologia e da educação, os testes psicológicos foram utilizados para justificar desigualdades, rotular populações e excluir grupos sociais das oportunidades de desenvolvimento pleno, como evidenciado pelo vínculo direto entre os primórdios da mensuração psicológica e o movimento eugenista (neste número, ver Bishop & Camiré, 2025).

Mais do que denunciar esse “legado”, este editorial propõe discutir a centralidade da tomada de decisão dos pesquisadores e pesquisadoras no estabelecimento de normas interpretativas para os testes psicológicos. Haja vista que reflexões sobre justiça social no campo da avaliação psicológica têm se concentrado majoritariamente na validade dos instrumentos, como demonstram os avanços representados pela proposta de validade orientada à justiça e ao antirracismo (Randall et al., 2023). Entretanto, uma lacuna importante ainda pode ser observada na aplicação dessas discussões críticas ao processo de normatização dos testes, ou seja, à definição dos parâmetros estatísticos que orientam a interpretação dos escores de um teste com base em uma população de referência.

Normatizar é estabelecer o que será considerado “esperado”, “adequado” ou “desviante” a partir de um grupo normativo, e assim oferecer os parâmetros para atribuição de sentidos psicológico aos escores do teste psicológico. Processo que envolve decisões com implicações éticas e políticas profundas. Este editorial defende que tais decisões também devem ser guiadas por compromissos com a justiça social, especialmente em contextos marcados por desigualdades estruturais. Para tanto, será brevemente apresentada a proposta da Teoria Quantitativa Crítica das Raças (QuantCrit) (Castillo & Strunk, 2025) e discutida a validade orientada à justiça (Randall et al., 2023), como perspectivas que podem contribuir para a prática avaliativa baseada na justiça social.

QuantCrit: repensando os métodos quantitativos

QuantCrit, ou Teoria Quantitativa Crítica das Raças, é uma abordagem que surge da integração entre a Teoria Crítica das Raças (*Critical Race Theory* – CRT) e os métodos quantitativos. Ela se fundamenta na premissa de que o racismo é uma característica estrutural e permanente da sociedade, afetando diretamente o contexto científico por meio da maneira como os dados são coletados, analisados e interpretados. Os cinco pilares centrais do QuantCrit incluem: 1. A centralidade do racismo; 2. A negação da neutralidade dos números; 3. A compreensão de que categorias analíticas são construções sociais; 4. A valorização da voz e da experiência dos grupos marginalizados; e 5. O comprometimento

com a justiça social. Diferentemente de outros paradigmas críticos que utilizam os métodos quantitativos apenas como ferramenta de descrição das desigualdades, QuantCrit propõe uma transformação radical dos pressupostos que estruturam esses métodos, exigindo reflexividade crítica sobre os usos da estatística e sobre as implicações éticas da pesquisa quantitativa. Assim, não se trata apenas de usar dados para evidenciar disparidades, mas de pensar os métodos quantitativos para servir à equidade racial. Aplicada à mensuração e à avaliação psicológica, essa perspectiva convida a uma reflexão sobre como os testes, suas construções, processos de validação e estabelecimentos de normas interpretativas podem contribuir para a manutenção de desigualdades estruturais.

Validade dos testes como compromisso com a justiça

A proposta de validade orientada à justiça e ao antirracismo (*Justice-Oriented and Antiracist Validity* – JAV), apresentada por Randall et al. (2023), amplia a tradicional abordagem baseada em argumentos de interpretação e uso dos testes. Essa proposta enfatiza que o processo de validação deve considerar os efeitos das avaliações sobre os grupos marginalizados desde o momento da concepção do teste até a interpretação dos seus resultados. A validade orientada à justiça parte da premissa de que os modelos tradicionais de validade são insuficientes se desconsiderarem as estruturas de opressão que moldam as experiências dos grupos sociais. Ao incorporar os princípios da Teoria Crítica das Raças e das abordagens antirracistas, a validade orientada à justiça amplia o escopo da validade tradicional ao exigir que as suposições subjacentes aos testes sejam examinadas criticamente, considerando as histórias de marginalização dos participantes e os impactos potenciais das decisões baseadas nos testes. Assim, a proposta amplia o foco da validade como comprovação de evidência empírica incorporando a validade como compromisso com a equidade e a transformação social.

Nessa perspectiva, o processo de validação não se limita a demonstrar que um teste mede o que se propõe a medir, mas também que esse processo não reforça a inferiorização de grupos minorizados. Para tanto, os efeitos sociais e políticos das interpretações precisam ser considerados e explicitados.

Normas interpretativas e a perpetuação da desigualdade

Um exemplo concreto dessa problematização refere-se à definição de normas interpretativas para testes de inteligência cristalizada, como o Exame Nacional do Ensino Médio (ENEM), uma das principais vias de acesso ao ensino superior no Brasil. É amplamente documentada a diferença substancial de desempenho médio no ENEM entre estudantes oriundos de escolas públicas e privadas, favorecendo sistematicamente estes últimos e, conseqüentemente, ampliando suas chances de aprovação nos processos seletivos das Instituições de Ensino Superior. Jaloto e Primi (2021) evidenciaram uma vantagem de quase meio desvio-padrão ($r=0,44$) a favor dos alunos da rede privada.

Tais discrepâncias intensificam o debate em torno da equidade e da justiça nos processos seletivos. Um dos argumentos centrais é que essas diferenças refletem desigualdades estruturais de natureza histórica, racial e socioeconômica, o que fundamenta a proposta de adoção de normas diferenciadas para os grupos, como forma de política afirmativa.

O racional psicométrico associado à equidade está ancorado na análise de funcionamento diferencial dos itens (DIF) ou, de forma mais ampla, na avaliação da invariância da medida em relação aos grupos de interesse. O objetivo do DIF é verificar se os itens de um teste operam psicometricamente de forma equivalente entre os grupos. Para isso, comparam-se indivíduos de diferentes grupos (por exemplo, oriundos de escolas públicas e privadas) com o mesmo nível de proficiência global no ENEM. Caso tenham a mesma probabilidade de acerto em um item específico, considera-se que o item possui funcionamento equivalente. No entanto, se indivíduos de um grupo apresentam maior probabilidade de acerto, mesmo com igual proficiência geral, isso caracteriza um viés sistemático do item, favorecendo esse grupo.

A presença de itens com DIF sistemático é, do ponto de vista técnico, a única justificativa considerada “neutra” para a adoção de normas diferenciadas, uma vez que a diferença entre os grupos não refletiria variações reais no construto, mas sim artefatos do instrumento. Em uma perspectiva estritamente psicométrica, apenas neste cenário seria recomendada a utilização de tabelas normativas separadas por grupo.

Entretanto, é possível que existam diferenças globais entre os grupos que não sejam decorrentes de DIF, ou seja, que os itens funcionem igualmente, mas os grupos apresentem médias distintas no traço latente. Nesse caso, o raciocínio técnico tradicional não endossaria a adoção de normas distintas, pois entenderia tais diferenças como “reais” no construto avaliado.

É precisamente nesse ponto que se torna necessário aprofundar a análise da complexidade inerente ao tema, tanto do ponto de vista técnico quanto social, a fim de embasar de forma mais informada as distintas posições em debate. Diferenças globais de proficiência podem ser resultado de desigualdades sociais no acesso a oportunidades de aprendizagem, sejam familiares (por exemplo, nível de escolaridade dos pais e disponibilidade de recursos culturais e educacionais), sejam escolares (qualidade e infraestrutura das instituições de ensino público). Utilizar os escores do ENEM sem qualquer forma de correção ou compensação implica, na prática, perpetuar a distribuição desigual de oportunidades educacionais, favorecendo sistematicamente um grupo em detrimento de outro em relação às oportunidades educacionais futuras.

Do ponto de vista técnico, a principal justificativa para o uso do ENEM em processos seletivos é sua validade preditiva, isto é, o grau em que o desempenho no exame indica o potencial de sucesso acadêmico no ensino superior. No entanto, estudos sobre essa validade são escassos tanto para o ENEM quanto para outras provas de processos seletivos adotados no Brasil. Ferreira-Rodrigues (2015), por exemplo, estimou uma correlação de aproximadamente $r=0,50$ entre o ENEM e o ENADE. Ainda assim, pouco se sabe sobre a magnitude dessa validade preditiva quando analisada separadamente por origem escolar. Esse ponto é crucial: se a validade for semelhante para alunos de escolas públicas e privadas, então a adoção de normas distintas pode ser tecnicamente justificável como uma medida de equidade, pois permitiria selecionar, dentro de cada grupo, os indivíduos com maior probabilidade de êxito no ensino superior, mesmo que os grupos apresentem desempenhos médios diferentes.

A interpretação dominante no senso comum tende a compreender as diferenças de desempenho entre grupos como reflexo direto e absoluto de diferenças de capacidade, incluindo a capacidade de aprender no futuro. Nessa lógica, estudantes que obtêm notas mais baixas no ENEM seriam, por definição, menos aptos ao sucesso acadêmico, independentemente de sua trajetória ou contexto. Dentro desse raciocínio, a proposta de utilizar normas diferenciadas entre grupos é frequentemente percebida como uma forma de “injustiça”, uma vez que implicaria aplicar injustamente pesos distintos para os dois grupos.

No entanto, essa visão simplifica de maneira excessiva, e problemática, a complexa rede de relações causais que influencia o desempenho em testes educacionais. Ignora-se, por exemplo, que o desempenho observado é o resultado de múltiplos fatores interativos, entre eles o acesso desigual a recursos educacionais, as condições socioeconômicas, a qualidade do ensino, e até mesmo fatores psicossociais como expectativas de professores e estigmas sociais. Esses fatores não apenas afetam o que foi aprendido, mas também moldam o quanto foi possível aprender, ou sequer foi uma possibilidade de aprendizado, ao longo da trajetória educacional.

Em outras palavras, quando grupos apresentam diferenças sistemáticas de desempenho, não se pode inferir automaticamente que essas diferenças refletem desigualdades “naturais” de capacidade. O desempenho em uma prova como o ENEM mede não apenas a proficiência atual, mas também os efeitos acumulados de oportunidades (ou da falta delas). Ao ignorar essa complexidade, o discurso meritocrático acaba por naturalizar as desigualdades, mascarando as condições estruturais que limitam o potencial de certos grupos e, assim, reforçando ciclos de exclusão sob o pretexto de neutralidade.

É interessante observar iniciativas recentes no campo da inteligência artificial aplicadas à promoção da equidade, como a desenvolvida pelo Google, que disponibilizou uma ferramenta interativa para visualização das decisões de concessão de crédito baseada em modelos de *machine learning*. Essa ferramenta implementa o conceito de equidade de oportunidades, distinguindo entre variáveis que devem ou não influenciar decisões automatizadas com base em princípios de justiça.

A proposta, acessível em <https://research.google.com/bigpicture/attacking-discrimination-in-ml/>, ilustra como algoritmos podem reproduzir ou mitigar desigualdades históricas dependendo dos critérios adotados para a modelagem e da forma como as variáveis sensíveis (como raça, gênero ou origem social) são tratadas. Ao simular diferentes cenários, a ferramenta evidencia que decisões aparentemente neutras podem, na prática, acentuar discriminações existentes, reforçando a importância de incorporar concepções de justiça distributiva e de equidade procedimental no desenvolvimento de sistemas de decisão automatizados.

Essa abordagem fornece um paralelo relevante para o debate em psicometria e avaliação educacional: assim como em sistemas algorítmicos, decisões baseadas em testes psicológicos e educacionais também podem perpetuar desigualdades se não forem sensíveis aos contextos históricos e sociais que moldam o desempenho. Ao estabelecer uma única norma, baseada em toda a amostra, ignoram-se as diferenças estruturais que afetam os grupos sociais de forma distinta. As normas interpretativas são referências estatísticas que servem como base para interpretar os escores dos sujeitos em relação a uma população de referência. Contudo, quando essa população é tratada como homogênea, desconsiderando as desigualdades históricas, raciais e socioeconômicas, corre-se o risco de mascarar ou reproduzir injustiças.

Sob a perspectiva do QuantCrit, a decisão por utilizar normas unificadas reflete uma crença na neutralidade dos números (princípio 2), ignorando que as condições de aprendizagem, acesso a recursos e vivências são moldadas por estruturas de opressão racializadas (princípio 1). Além disso, desconsidera-se a importância da voz e da experiência dos grupos marginalizados (princípio 4), os quais poderiam indicar que uma interpretação “igual para todos” não é, de fato, equitativa. Portanto, a adoção de normas diferenciadas pode ser justificada teoricamente como uma forma de reconhecer e responder às desigualdades materiais e simbólicas entre os grupos, atuando de forma orientada à justiça social (princípio 5), corre-se o risco de estigmatizar estudantes de escolas públicas que, em condições desiguais de acesso ao ensino e oportunidades de aprendizagem obtêm escores mais baixos. Isso não reflete um déficit individual, mas sim desigualdades sociais. Portanto, a opção por normas distintas pode representar um ato de justiça interpretativa, reconhecendo que é injusto comparar sujeitos com oportunidades tão diferentes com a mesma régua.

Retomando os conceitos técnicos de validade e considerando o exemplo do ENEM, é pertinente introduzir o conceito de validade consequencial, conforme proposto por Messick (1989). Esse conceito enfatiza que a validade de um teste não se limita apenas à quão bem ele mede um construto, mas também inclui a análise crítica das consequências sociais, tanto intencionais quanto não intencionais, decorrentes da interpretação e uso dos escores. Messick argumenta que essas consequências são parte integrante da validade, especialmente quando envolvem questões de viés, equidade e justiça distributiva.

Aplicando esse conceito ao contexto do ENEM, o propósito fundamental da seleção é identificar estudantes com potencial para se desenvolver no ensino superior. Se a validade preditiva do ENEM for equivalente entre subgrupos (por exemplo, alunos de escolas públicas e privadas), mas um grupo tiver maior probabilidade de acesso devido a vantagens acumuladas por oportunidades educacionais anteriores, ignorar essas relações pode levar a consequências indesejadas. Especificamente, isso pode resultar em um acesso desproporcional ao ensino superior para o grupo com desempenho historicamente favorecido, perpetuando desigualdades sociais sob a suposição de que escores mais altos indicam maior potencial acadêmico.

Essa interpretação simplista desconsidera a complexidade das relações causais entre variáveis sociais e educacionais que influenciam o desempenho nos testes. Ao não reconhecer que fatores como qualidade da educação básica, contexto socioeconômico e acesso a recursos educacionais afetam significativamente os escores, corre-se o risco de validar decisões que reforçam injustiças estruturais. Portanto, a análise das consequências do uso dos escores do ENEM, à luz da validade consequencial e orientada à justiça, é essencial para garantir que as decisões baseadas nesses escores promovam equidade e justiça social, alinhando-se aos princípios éticos fundamentais da avaliação psicológica e educacional.

Essa perspectiva exige dos pesquisadores um posicionamento bem definido, o de romper com a ilusão de neutralidade e reconhecer que as escolhas metodológicas moldam a forma como compreendemos as realidades sociais. É preciso superar a tradição de simplesmente comparar grupos para evidenciar diferenças, prática que historicamente tem servido para justificar desigualdades e naturalizar hierarquias sociais. Em contraponto, abordagens como o modelo QuantCrit e a validade orientada à justiça propõem o oposto, o uso dos métodos e dados quantitativos para combater as injustiças, reconhecendo que toda decisão, por mais que seja orientada por dados “objetivos”, envolve algum grau de subjetividade e, portanto, de valores e crenças. Essa posição, favorece que o uso dos testes ocorra de fato para a compreensão ampliada e justa dos indivíduos.

Considerações finais: para uma avaliação comprometida com a equidade

A psicologia brasileira, especialmente em sua interface com as políticas públicas, tem uma longa trajetória de compromisso com os direitos humanos e com a promoção da equidade. A avaliação psicológica, como prática central dessa ciência, não pode se furtar a essa responsabilidade. A adoção de perspectivas críticas para a mensuração, como a QuantCrit e os modelos de validade antirracistas nos desafia a assumir uma nova postura na produção e interpretação dos dados quantitativos.

Ao definir os critérios normativos de um teste, pesquisadores tomam decisões que impactam diretamente a forma como os indivíduos serão compreendidos, avaliados e, muitas vezes, classificados em contextos decisivos como o acesso a oportunidades educacionais, sociais e profissionais. A escolha por um determinado tipo de norma pode, portanto, contribuir para a perpetuação ou a mitigação das desigualdades. Quando normas únicas são utilizadas para comparar grupos que vivem sob condições sociais desiguais, há o risco de reforçar estigmas e alimentar narrativas deficitárias sobre determinados segmentos populacionais.

Por outro lado, adotar normas que reconhecem essas desigualdades estruturais pode ser um passo ético e necessário em direção à justiça social. Nesse sentido, os impactos éticos dessas decisões envolvem refletir sobre quem se beneficia e quem é prejudicado por determinadas interpretações, e como a avaliação psicológica pode servir como instrumento de equidade ou de opressão. Este editorial, propõe que os pesquisadores devem assumir uma postura ativa e comprometida com a equidade reconhecendo os contextos de vida das pessoas que compõem os grupos normativos.

Referências

- Bishop, E. & Camiré, M. (2025). A Conceptual Looking Back(for)wards to Foster Justice in Psychometrics. *Avaliação Psicológica*, 24 (n. esp.), e25458. <http://dx.doi.org/10.15689/ap.2025.24.e25458>
- Castillo, W., & Strunk, K. K. (2025). *How to QuantCrit: Applying critical race theory to quantitative data in education*. Editora Routledge.
- Ferreira-Rodrigues, C. F. (2015). *Estudos com o ENEM a partir de uma abordagem psicométrica da inteligência* [Tese de doutorado, Universidade São Francisco].
- Jaloto, A., & Primi, R. (2021). Fatores socioeconômicos associados ao desempenho no Enem. Em: V. Weller Enem e Gaokao: repercussões no ensino médio e na educação superior. *Em Aberto*, 34(112), 125-141. INEP. <https://doi.org/10.24109/2176-6673.emaberto.34i112.5002>

- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). ACE & NCME.
- Randall, J. (2025). The House that Hate Built: Fixing the Mess We Made. *Avaliação Psicológica*, 24(n. esp.), e25632. <http://dx.doi.org/10.15689/ap.2025.24.e25632>
- Randall, J., Slomp, D., Poe, M., & Oliveri, E. (2023). Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. In *Twin Pandemics* (pp. 78-86). Routledge.

Como citar este artigo

Peixoto, E. M., & Primi, R. (2025). A centralidade da justiça social na interpretação de testes psicológicos [Editorial]. *Avaliação Psicológica*, 24, nº especial 1, e26109. <http://doi.org/10.15689/ap.2025.24.e26109>