

Fundamentos e aplicabilidade dos Testes de Julgamento Situacional no contexto da Psicologia Organizacional

Fundamentals and applicability of the Situational Judgment Tests in the Organizational Psychology Context
Fundamento y aplicabilidad de los Tests de Juicio Situacional en el contexto de la Psicología Organizacional

Rodolfo Augusto Matteo AMBIEL¹
Maria Isabel de CAMPOS
Bárbara de Paula ALVES
Carla Priscila SILVA
Universidade São Francisco, Itatiba, SP, Brasil

Resumo Considerando a ausência de publicações sobre o assunto no Brasil, este artigo teórico tem como objetivo apresentar os fundamentos e pesquisas com os Testes de Julgamento Situacional, que são provas compostas por narrativas de situações que apresentam desafios realistas relacionados a tarefas de trabalho e alternativas que representam resoluções para estas situações. Este tipo de instrumento goza de prestígio na literatura internacional, por conta de seus bons resultados em termos de predição de comportamentos no trabalho. Neste artigo, discorre-se sobre sua aplicabilidade no campo da psicologia organizacional, mais especificamente na seleção de pessoal, bem como sobre suas propriedades psicométricas, principalmente no que diz respeito à influência das instruções na validade dos instrumentos utilizados. Em seguida, são apresentadas algumas pesquisas recentes que utilizaram Testes de Julgamento Situacional no contexto da psicologia organizacional. Apresentam-se também sugestões para a popularização dos Testes de Julgamento Situacional no Brasil e, por fim, diretrizes para pesquisas futuras.

Palavras-chave:

Psicologia organizacional; seleção de recursos humanos; testes de seleção; testes psicológicos; avaliação psicológica.

Abstract Considering the lack of publication on this subject in Brazil, the intent of this theoretical paper is to present the fundamentals and research concerning Situational Judgment Tests, which are tests composed of narratives of situations that introduce realistic challenges related to work tasks and alternatives representing resolutions to the given situation. This kind of assessment instrument enjoys prestige in the international literature, owing to its positive results in predicting work-related behaviors. In this paper, we discuss its applicability in the field of organizational psychology, specifically in personnel selection, as well as its psychometric properties, mainly regarding the influence of the instructions on the validity of the instruments. We then present some recent studies that have investigated aspects of Situational Judgment Tests in organizational psychology, as well as suggestions for popularizing Situational Judgment Tests in Brazil and guidelines for future research.

Keywords:

Organizational psychology; personnel selection; selection tests; psychological testing; psychological assessment.

¹ Endereço para correspondência: Universidade São Francisco, Setor de Apoio de Humanas, Rua Alexandre Rodrigues Barbosa, 45, Itatiba, SP, Brasil 13251-000. Telefone: (11) 4534.8000. Email: rodolfo.ambiel@saofrancisco.edu.br

Resumen Considerando la ausencia de publicaciones sobre este asunto en Brasil, este artículo teórico tiene por objetivo presentar los fundamentos e investigaciones de los tests de juicio situacional, que son pruebas compuestas por narrativas de situaciones que presentan desafíos realistas relacionados a tareas de trabajo y alternativas que representan soluciones para la situación. Este tipo de instrumento tiene buen prestigio en la literatura internacional, por cuenta de sus buenos resultados en relación a la predicción de comportamientos en el trabajo. En este artículo, se discute su aplicación en el campo de la psicología organizacional, más específicamente en la selección de personal, bien como sobre las propiedades psicométricas de los mismos, principalmente en lo que se refiere a la influencia de las instrucciones de validez de los instrumentos. En seguida, se presentan algunas investigaciones recientes que utilizaron Tests de Juicio Situacional en contexto de psicología organizacional. Se presentan también sugerencias para la popularización en Brasil y, finalmente, algunas directrices para investigaciones futuras.

Palabras-claves:
Psicología organizacional; selección de recursos humanos; tests de selección; tests psicológicos; evaluación psicológica.

Testes de Julgamento Situacional (TJS), em inglês *Situational Judgment Tests (SJT)*, são provas utilizadas para avaliar o julgamento de candidatos ou de funcionários a respeito da adequação de resoluções apresentadas para breves narrativas (cenários) que representam situações e desafios realistas relacionados ao trabalho (McDaniel & Nguyen, 2001). A utilização de TJS vem se tornando cada vez mais popular internacionalmente no contexto da psicologia organizacional (McDaniel & Nguyen, 2001; Ployhart & Ehrhart, 2003; Roth, Bobko, & Buster, 2013, Weekley & Ployhart, 2006b).

Apesar disso, publicações sobre o tema são quase inexistentes no Brasil. Uma busca realizada no Google Acadêmico em julho de 2014 com as palavras *Testes de Julgamento Situacional* retornou apenas duas publicações nacionais. A primeira é referente a uma tese de doutorado que aborda a construção e os estudos psicométricos da escala Ciclos de Adaptação de Carreira (Ambiel, 2013) e a segunda apenas menciona que esse tipo de avaliação é posicionado pela American Psychological Association - APA em seu *Handbook of Industrial and Organizational Psychology* como uma das possíveis estratégias de seleção de pessoas em ambientes organizacionais (Gondim, Borges-Andrade, & Bastos, 2010). Existe, ainda, um instrumento de avaliação de estilos de liderança situacional, que é a versão brasileira do *Leadership Judgment Indicator* (Lock, Wheeler, Burnard, & Cooper, 2005) publicado no Brasil como Indicador de Julgamento de Liderança (Rabelo, Alves, Ambiel, Pacanaro, & Leme, 2012). No estudo brasileiro, foram realizados os procedimentos de tradução e adaptação semântica dos itens e verificadas evidências de validade baseadas no conteúdo, na estrutura interna, que replicou a estrutura original de quatro fatores relativos aos estilos de liderança avaliados, além de evidências de validade de critério.

Embora a seleção de pessoas seja apenas uma das áreas de possível aplicação dos TJS, é a que tem maior presença na literatura internacional (McDaniel & Nguyen, 2001; Ployhart & Ehrhart, 2003; Weekley & Ployhart, 2006a). No Brasil, embora não haja publicações diretamente relacionadas a essa aplicação de TJS, uma forma estruturalmente similar de avaliação em processos de recrutamento e seleção foi encontrada em um modelo de entrevista por competências (Gondim & Queiroga, 2013).

No exterior, é grande o volume de publicações sobre o tema, abrangendo, além da seleção, estudos sobre avaliação da personalidade (Motowidlo & Beier, 2010) e de desempenho (Bledow & Frese, 2009; Hauenstein, Findlay, & McDonald, 2010). No entanto, muitas publicações indicam a necessidade de mais pesquisas na área, tanto no que concerne a investigações a respeito da validade (McDaniel, Whetzel, Hartman, Nguyen, & Grubb, 2006), quanto a aspectos teóricos ainda carentes de desenvolvimento (Brooks & Highhouse, 2006; Gessner & Klimonski, 2006; Ployhart & Ehrhart, 2003).

Tendo em vista a crescente popularidade internacional dos TJS entre profissionais da academia e da prática organizacional desde a década de 1990, a existência de diversas lacunas apontadas na literatura e a carência de produção brasileira a esse respeito, este artigo tem o objetivo de apresentar a técnica de avaliação dos TJS e o seu panorama na literatura internacional no período de 2009 a 2013, mostrando, assim, o estado da arte publicado sobre o tema no momento da realização da pesquisa. Objetiva-se, ainda, identificar lacunas e possibilidades para novas pesquisas.

Formatos dos Testes de Julgamento Situacional

Os TJS apresentam aos respondentes uma situação ou cenário (*stem item*) e uma lista de possíveis respostas ou resoluções (*response item*) para a situação dada (McDaniel & Nguyen, 2001). Em um TJS típico utilizado em um processo de seleção, por exemplo, o respondente depara-se com uma variedade de situações que poderá encontrar no exercício do trabalho, geralmente baseadas em incidentes críticos característicos à função. Espera-se que o respondente seja capaz de prever qual seria sua ação/reação à situação, de acordo com as alternativas disponíveis, ou que indique qual seria a melhor saída (Weekley & Ployart, 2006b), podendo ser apresentada por escrito ou por vídeos (Motowidlo & Beier, 2010).

De acordo com Bledow e Frese (2009), o julgamento do respondente a respeito de uma possível ação futura caracteriza uma das principais diferenças entre os TJS e as avaliações do tipo autorrelato. Brooks e Highhouse (2006) afirmam, porém, que o que realmente distingue os TJS de outros tipos de medidas é o seu potencial de alcançar não apenas a medida do conhecimento racional que compõe um julgamento e é passível de ser explicado, mas também a medida de um conhecimento intuitivo, tácito, de difícil descrição processual.

Retomando a questão do formato, um TJS é um conjunto de itens, composto por uma base, – a situação a ser julgada –, e por um conjunto de possíveis respostas (McDaniel & Nguyen, 2001). A forma como as respostas se apresentam ao testando depende de como o autor constrói o instrumento. O testando tanto pode ser instruído a fazer uma escolha forçada de uma das respostas possíveis, quanto ser orientado a pontuar todas elas indicando o nível de sua efetividade, utilizando uma escala do tipo Likert (Weekley & Ployart, 2006b).

Na Tabela 1, é apresentado um item hipotético como exemplo, formulado pelos autores deste artigo a partir das diretrizes fornecidas por McDaniel e Nguyen (2001) e de exemplos de TJS encontrados em Gessner e Klimoski (2006) e em Brooks e Highhouse (2006). Esse item hipotético poderia compor um instrumento de seleção para o exercício de uma função que abrangesse relacionamento com o cliente.

TABELA 1. Exemplo de item de um teste de julgamento situacional

Item	Conteúdo
Cenário (<i>stem item</i>)	Você está de férias e recebe um telefonema de um cliente que foi previamente informado sobre sua ausência. Ele se desculpa por incomodá-lo(a) nessa situação, mas explica que está enfrentando um problema que apenas você pode ajudar a resolver. O que você faria?
Itens de resposta (<i>response item</i>)	<ol style="list-style-type: none"> 1. Pediria ao cliente que aguardasse o seu retorno das férias, sem nem mesmo ouvir o problema. 2. Ouviria o problema e solicitaria ao cliente que entrasse em contato com seu chefe ou com quem tenha ficado em seu lugar na empresa durante suas férias. 3. Ouviria o problema e informaria ao cliente que iria repassar a situação à pessoa responsável por ela durante suas férias. 4. Ouviria o problema e se engajaria pessoalmente em solucioná-lo, mantendo contato direto com o cliente e informando-o sobre o andamento da solução.

A seleção da resposta dependerá, como indicam Gessner e Klimoski (2006), não apenas do julgamento do respondente a respeito da situação, mas também da sua percepção sobre a empresa por meio dos contatos anteriores ao longo do processo de recrutamento e da administração do teste. Ele poderá avaliar o que a empresa espera que seja feito, de acordo com o que entendeu a respeito da cultura da organização. A resposta, seu sentido e até mesmo sua validade psicométrica serão também altamente dependentes das instruções passadas pelo autor e pelo administrador do teste, uma vez que elas determinam sua aplicabilidade (Ployhart & Ehrhart, 2003).

A resposta depende da pergunta – A aplicabilidade também

Gessner e Klimoski (2006) consideram que há três tipos distintos de partes interessadas (*stakeholders*) nos TJS, em especial naqueles utilizados em processos de seleção, e que elas apresentam diferentes percepções a respeito desses testes, o que cria desafios especiais para o campo. Esses *stakeholders* são o autor, o administrador e o testando. Segundo os referidos autores, a literatura geralmente focaliza os interesses e desafios do autor do teste.

Por se tratar de um campo que ainda apresenta muitas questões em aberto, é natural que o foco se detenha nas diversas questões que o autor do teste deve considerar para que seus resultados sejam válidos, fidedignos e pertinentes aos interesses das outras duas partes interessadas. Uma das questões de alto interesse relaciona-se às instruções definidas pelo autor para todo o TJS ou para cada item dele. Conforme indicado por Ployhart e Ehrhart (2003), essa definição influenciará a seleção e a aplicação do teste pelo administrador. Afetará, também, a interpretação e o julgamento do respondente, refletindo na validade e na fidedignidade do teste.

Em um levantamento a respeito dos principais tipos de instruções encontrados na literatura, Ployhart e Ehrhart (2003) examinaram os efeitos das instruções sobre as propriedades psicométricas, especificamente sobre a validade de construto de um TJS. Concluíram que o uso de diferentes instruções pode levar a diferentes resultados para os mesmos itens de um TJS. Os diferentes tipos de instruções foram resumidamente agrupados sob as categorias: “O que você faria?” (*would do*) e “O que você deveria fazer?” (*should do*). A Tabela 2, adaptada de Ployhart e Ehrhart (2003), apresenta diferentes tipos de instruções possíveis para um item, agrupadas de acordo com a categoria em que se encaixam e com possíveis formas de pontuação.

TABELA 2. Tipos de instruções de um TJS

Tipo de instrução	Ação	Modo de pontuação
O que você faria?	O que você faria?	O respondente deve escolher apenas uma opção, aquela que represente a melhor forma com que ele agiria.
	Qual seria a resposta mais provável e qual a menos provável de você fazer?	O respondente deve escolher duas opções, sendo uma delas a mais provável de ser feita e a outra, a menos provável.
	Pontue cada uma das respostas de acordo com a probabilidade de você agir daquela forma perante essa situação.	Utilizando uma escala Likert, o respondente deve avaliar sua probabilidade de agir conforme cada opção.
O que você deveria fazer?	O que você deveria fazer?	O respondente deve escolher apenas uma opção, aquela que represente a forma como ele deve agir.
	Qual seria a ação mais efetiva e a menos efetiva?	O respondente deve escolher duas opções, sendo uma a que ele considera como a melhor opção e a outra, como a pior.
	Pontue a efetividade de cada resposta para essa situação.	Utilizando uma escala Likert, o respondente deve avaliar a efetividade de cada opção.

Pelo exemplo anterior e pelas informações da Tabela 2 pode-se notar que itens de julgamento situacional apresentam situações hipotéticas com as quais o respondente poderá se deparar e, a partir disso, é solicitado que avalie e/ou escolha respostas para o dilema. Conforme apontado na literatura (Brooks & Highhouse, 2006; Ployhart & Ehrhart, 2003), o uso de diferentes tipos de instruções, como “O que você faria” ou “O que você deveria fazer”, altera a natureza e a qualidade das predições que podem ser feitas a partir do que o respondente é solicitado a responder.

Brooks e Highhouse (2006) consideraram que versões do tipo “O que você faria?” requerem que o respondente preveja seu próprio comportamento e que versões do tipo “O que você deveria fazer?” requerem uma resposta do tipo dado x , preveja y . No primeiro tipo de instrução a avaliação ocorre, predominantemente, a partir da própria experiência do respondente. No segundo, a avaliação ocorre por meio de um conhecimento normativo, relacionado com o que se espera que a pessoa saiba sobre o procedimento ou situação em questão e, por isso, mais suscetível a deseabilidade social e com menor poder preditivo (Ployhart & Ehrhart, 2003). Assim, o teste deve instruir o respondente claramente a respeito do tipo de predição que se espera dele e essa responsabilidade é inerente ao seu autor.

Na prática, o esclarecimento pode ser dado também pelo administrador do teste, ou por algum material de apoio disponível ao testando. Um exemplo disso pode ser encontrado no livro de Metcalfe, Dev, Boursicot e Sales (2014), que tem o objetivo de preparar candidatos para o processo seletivo à formação em prática médica em Oxford, no Reino Unido. Parte desse processo abrange a aplicação de um TJS, que é posicionado como o principal componente do processo. Os tipos de questão e a maneira como devem ser respondidas são esclarecidos pelos autores, para que os candidatos se preparem com antecedência, pois, como referido, o uso de diferentes perguntas influencia a validade dos resultados obtidos no teste.

Metcalfe et al. (2014) apontam que, após apresentar o formato do TJS e os atributos a serem avaliados, os candidatos devem considerar que estarão sendo arguidos sobre “O que deveria ser feito”, e não sobre “O que eles fariam” em cada uma das situações. Informam, ainda, que a intenção é avaliar se o candidato sabe qual é a ação “correta” para cada cenário apresentado, e não se ele agiria corretamente caso trabalhasse como médico.

Além disso, Metcalfe et al. (2014) esclarecem também que todas as questões apresentadas no teste avaliarão nove tipos distintos de atributos profissionais identificados como necessários ao papel a que ora se candidatam: (a) comprometimento e profissionalismo, (b) forma de lidar com a pressão, (c) comunicação efetiva, (d) aprendizagem e desenvolvimento profissional (e) organização e planejamento, (f) foco no paciente, (g) resolução de problemas e tomada de decisão, (h) autoconsciência e *insight*, e (i) trabalho efetivo como parte de uma equipe.

Situações relacionadas à prática da aplicação dos TJS, como as orientações apontadas e a lista de atributos apresentados para o teste aplicado em Oxford (Metcalfe et al., 2014), são úteis para direcionar algumas das questões ainda polêmicas para os pesquisadores desse campo de estudos. Além de pontos relacionados ao aumento da popularidade desse tipo de instrumento, há também as importantes questões: o TJS é um construto ou um método? Se método, afinal, o que mede um TJS? (Brooks & Highhouse, 2006; Schmitt & Chan, 2006).

Por que a popularidade e o que medem os TJS?

Diversos autores, entre eles Bledow e Frese (2009), Gessner e Klimoski (2006), McDaniel e Nguyen (2001), Ployhart e Ehrhart (2003) e Weekley e Ployhart (2006b), apontam que a popularidade dos TJS vem crescendo há pouco mais de duas décadas, quando o tema foi retomado por Motowidlo, Dunnette e Carter (1990). Para Gessner e Klimoski (2006), essa popularidade deve-se ao potencial dos TJS como instrumentos de custo relativamente baixo e à facilidade com que podem ser aplicados.

Weekley e Ployhart (2006b) vão além desses benefícios básicos, apontando estudos que indicam que os TJS apresentam uma validade incremental acima e além dos preditores tradicionais, como habilidades cognitivas e traços de personalidade. Isso sugere, segundo os autores, que os TJS estão capturando algo único, relacionado ao desempenho de um indivíduo, que não é capturado por outros construtos tradicionais. Essa linha de pensamento leva ao debate sobre se os TJS representam um método ou um construto.

A polêmica parece existir com base na própria nomenclatura dos testes, uma vez que, conforme Brooks e Highhouse (2006), o nome *Teste de Julgamento Situacional*, em si, já posiciona o construto *juízo*. Possivelmente é o mesmo entendimento que leva McDaniel e Nguyen (2001) a definirem os TJS como avaliações projetadas para medir o julgamento em ambientes de trabalho.

Esse conceito, no entanto, não é consenso entre autores, visto que Ployhart e Ehrhart (2003) definem os TJS como um método popular de seleção, e Gessner e Klimoski (2006) os definem como uma simulação de baixa a moderada fidelidade, projetada para avaliar preferências por comportamentos apropriados em uma situação de trabalho, o que caracterizaria o julgamento do indivíduo. Porém, os autores chamam a atenção para o fato de que ainda é necessário desenvolver uma teoria sobre o que é um bom julgamento. Essa diversidade de definições, a carência de clareza sobre o construto julgamento e mesmo a redução relativa ao ambiente de aplicação tornam o campo de estudos ainda polêmico e carente de melhores especificações.

O que parece ser uma concordância na literatura é que os TJS são importantes devido a sua capacidade preditora do comportamento humano para a situação exposta (Hanson & Ramos, 1996; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Phillips, 1992, 1993; Weekley & Jones, 1997, 1999). Esse comportamento, entretanto, pode ou não estar relacionado a processos de seleção, ao ambiente de trabalho e, além disso, pode ou não ser aquele que seria assumido pelo indivíduo avaliado, conforme já visto no que tange às instruções de respostas. Dessa forma, para McDaniel e Nguyen (2001), assim como qualquer outro método de medida, os TJS podem ser projetados para medir uma variedade de construtos havendo, contudo, limites para o que um TJS pode ou não medir. Para esses autores, o julgamento não seria um construto, mas um processo requerido para resolver os itens desse tipo de instrumento, tal qual o autorrelato está para os inventários de personalidade, por exemplo.

Os aspectos abordados até aqui permitem alcançar um dos objetivos deste trabalho, o de apresentar a técnica de avaliação dos TJS. A seguir, serão descritas algumas pesquisas utilizando os TJS no contexto organizacional a partir de um levantamento do panorama de publicações sobre o tema no período de 2009 a 2013.

Pesquisas com TJS no período de 2009 a 2013

Em maio de 2014, foi realizada uma consulta no Portal de Periódicos da CAPES visando encontrar artigos publicados no período de 2009 a 2013. As consultas utilizaram três termos para o mapeamento, *Situational Judgment Tests*, *Situational Judgment Test* e *organizational*, combinados em dois pares distintos, que poderiam ocorrer em qualquer campo dos artigos. Foram encontradas 32 ocorrências a partir da combinação *Situational Judgment Tests* e *organizational*. Para a combinação com o primeiro termo no singular foram encontrados 17 artigos. Excluindo-se os termos repetidos, restaram 39 trabalhos.

Após esse refinamento e a leitura dos 39 artigos, observou-se que apenas 13 deles apresentavam pesquisas que tinham os TJS como foco principal do estudo. Os outros apenas citavam o termo. Desses 13, em 12 analisavam-se pesquisas empíricas e no último encontravam-se os resultados de uma metanálise. Dos 13 artigos, apesar da restrição do termo durante a busca, percebeu-se que seis eram relativos a outras áreas da psicologia, principalmente à área educacional. Por fim, foram identificados sete artigos de pesquisa, incluindo a abordagem de metanálise, com os TJS especificamente focados na área organizacional. A Tabela 3 apresenta um resumo dos achados relativos a esses sete artigos, que serão brevemente descritos a seguir.

TABELA 3. Artigos de pesquisa com os TJS especificamente focados na área organizacional publicados de 2009 a 2013

Autores e ano de publicação	Objetivo e conclusões da pesquisa
Bledow e Frese (2009)	Utilizar o TJS como método de avaliação da iniciativa pessoal no trabalho. Concluíram que o TJS é um método promissor para medir a iniciativa pessoal e que seu uso pode ser um meio de aumentar a validade das medidas organizacionais.
Salter e Highhouse (2009)	Entender se os TJS são instrumentos adequados para medir a inteligência prática (senso comum). Os autores sugerem que os TJS podem ser considerados um método eficiente de avaliação da inteligência prática e que podem ser facilmente adaptados para aplicação em diversos ambientes de trabalho.
Hauenstein et al. (2010)	Investigar a aplicabilidade de um TJS como instrumento de avaliação de treinamento. Os autores consideraram que os TJS são adequados para uso em contextos de treinamento, quando não há medidas específicas para avaliar a transferência de conhecimento alcançada.
Meijer, Born, Jaap e Molen (2010)	Demonstrar que é possível desenvolver um TJS para medir um construto específico (integridade). Os achados indicaram que o TJS foi capaz de avaliar integridade nos dois grupos participantes da pesquisa, um de maioria e um de minoria étnica.
Motowidlo e Beier (2010)	Realizar um estudo comparativo de validade de medidas entre os escores de um TJS desenvolvido para prever o desempenho profissional com pontuações relativas a traços de personalidade. Entenderam que o modelo teórico proposto estava correto e que, se os resultados puderem ser replicados, será possível afirmar que o conhecimento geral representado por traços de personalidade é tão importante quanto o conhecimento específico a respeito do trabalho na determinação do desempenho alcançado.
McDaniel, Psothka, Legree, Powell e Weekley (2011)	Verificar se os métodos que controlam a elevação e variação da pontuação do TJS melhoram a validade do item e se reduzem diferenças médias de pontuação em grupos étnicos compostos por brancos e negros. Os resultados apontaram que os métodos de controle melhoram a validade dos itens e reduzem significativamente as diferenças de escores entre os grupos.
Roth et al. (2013)	Verificar como construtos sobre habilidades cognitivas e sociais poderiam influenciar as padronizações de diferenças entre grupos étnicos, avaliadas por meio do índice estatístico <i>d</i> (tamanho do efeito da amostra). Os autores apontam resultados convergentes com a literatura e que fornecem estimativas de níveis de pontos para o TJS, para que os especialistas possam estimar níveis de <i>d</i> e os impactos adversos que podem ocorrer durante o desenvolvimento do sistema de seleção.

Salter e Highhouse (2009) realizaram uma pesquisa que visou entender se os TJS são instrumentos adequados para medir a inteligência prática, por eles denominada de senso comum. Realizaram uma revisão da literatura relevante em psicologia organizacional e industrial, considerando a história, a definição e o

desenvolvimento dos TJS. A pesquisa abrangeu também a revisão da literatura empírica e teórica no que diz respeito à utilização dos TJS para medir a eficácia na tomada de decisão gerencial.

A partir dessas revisões, Salter e Highhouse (2009) concluíram que os TJS têm sido apontados como preditores efetivos de comportamento em muitas situações organizacionais envolvendo ampla variedade de ambientes e posições de trabalho. Indicam que os TJS são utilizados para melhorar a predição de desempenho futuro no trabalho, apresentando resultados que vão além dos de outros preditores típicos. Em particular, quando comparados a outros métodos de seleção, como preditores do desempenho em vendas, os TJS apresentaram maior capacidade preditiva ($r = 0,34$) do que outros métodos amplamente utilizados, sendo eles testes de habilidade cognitiva ($r = 0,30$), pontuação de extroversão ($r = 0,28$), pontuação de orientação à realização ($r = 0,25$), entrevistas não estruturadas ($r = 0,20$) e grafologia ($r = 0,0$), ficando atrás apenas dos testes de habilidade de vendas ($r = 0,45$). Os autores indicam que TJS são relativamente fáceis de aplicar e de pontuar e que são bem recebidos por gerentes e candidatos em processos de seleção. Sugerem, ainda, que os TJS podem ser considerados um método eficiente de avaliar o senso comum (inteligência prática) e que podem ser facilmente adaptados para serem aplicados em diversos ambientes de trabalho.

O estudo de Bledow e Frese (2009) utilizou o TJS como método de avaliação da iniciativa pessoal no trabalho. Na fase inicial, de criação do teste, participaram 30 trabalhadores ocupantes de vários cargos. Na segunda fase, de validação de conteúdo, participaram 25 estudantes de psicologia organizacional, selecionados por serem especialistas no construto iniciativa pessoal. Por fim, na terceira fase, de validação, participaram 126 funcionários de bancos alemães, divididos entre supervisores e não supervisores. Foram utilizados TJS que refletissem cenários de trabalho com indicadores formativos do nível de iniciativa pessoal no trabalho, por meio dos quais os respondentes indicavam suas preferências a partir de simulações mentais. Os resultados da etapa de validação mostraram que os TJS de iniciativa pessoal tinham validade adequada. Suas medidas apresentaram correlações variando entre 0,29 e 0,48 com três variáveis critério: (a) pontuação de iniciativa pessoal, (b) pontuação de desempenho geral dadas pelos supervisores e (c) pontuação de iniciativa pessoal dada pelos próprios funcionários.

Segundo Bledow e Frese (2009), os TJS complementaram os resultados de predição de comportamento apresentados pelas escalas de autorrelato, tipo Likert. Baseados na hipótese de que preferências localizadas pela iniciativa pessoal, medidas em situações específicas, seriam preditoras do comportamento real, os autores encontraram que essas preferências mediam as relações entre autoeficácia generalizada, sentimento de responsabilidade e comportamento real. Assim, ao contrário do que ocorre com escalas do tipo Likert, os TJS, por não serem baseados em medidas descontextualizadas e generalizadas, mas em exemplos comportamentais específicos, podem reduzir a ambiguidade das respostas nos itens. Os autores concluíram que o TJS é um método promissor para medir a iniciativa pessoal e que seu uso pode ser um meio de aumentar a validade das medidas organizacionais. Indicam, no entanto, que mais pesquisas precisam ser realizadas, incluindo a mensuração de construtos independentes em uma mesma amostra. Isso ajudaria a esclarecer até que ponto há um fator metodológico geral para os TJS, que dificultaria a separação dos construtos avaliados.

Motowidlo e Beier (2010) realizaram um estudo comparativo de validade de medidas tomando por base uma teoria que afirma que o conhecimento procedural, na forma como é captado pelo TJS, pode ser entendido como uma união de conhecimento específico do trabalho e de traços implícitos, que são crenças implícitas sobre as relações entre as expressões de traços de personalidade e eficácia em situações de trabalho. Eles compararam os escores de chaves de pontuação de um TJS desenvolvido para prever o desempenho interpessoal e de solução de problemas em funções gerenciais, administrativas e profissionais com pontuações relativas a traços de personalidade considerados mais possíveis de estarem relacionados ao desempenho no trabalho, como amabilidade, extroversão e conscienciosidade. Para isso, utilizaram três amostras distintas, duas delas coletadas de estudantes de graduação e pós-graduação, que forneceram o denominado conhecimento de novatos. A terceira foi baseada em resultados de pesquisa anteriormente realizada por um dos autores, com profissionais que ocupavam posições administrativas e gerenciais em empresas de telecomunicações.

Os resultados encontrados por Motowidlo e Beier (2010) contribuíram para o fornecimento do denominado conhecimento de *experts* e de critérios de desempenho. A análise realizada com os dados das três amostras indicou que os escores obtidos pelos estudantes de graduação e de pós-graduação correlacionaram-se significativamente com os critérios de desempenho, mesmo considerando que nenhum dos grupos de estudantes tinha experiência de trabalho ou gerencial. Isso demonstrou que o modelo teórico proposto pelos autores

estava correto e que, se esses resultados puderem ser replicados, será possível afirmar que o conhecimento geral representado por traços de personalidade é tão importante quanto o conhecimento específico a respeito do trabalho na determinação do desempenho alcançado. Dessa forma, os autores apontam que o modelo proposto e os resultados sugerem que TJS podem ser utilizados mesmo para a seleção de candidatos novatos, que não têm experiência profissional relevante ou conhecimento do trabalho específico, desde que os itens considerem o conhecimento relacionado ao domínio em questão, mas desenvolvido por meio de experiências gerais de vida.

Hauenstein et al. (2010) investigaram a aplicabilidade de um TJS como instrumento de avaliação de treinamento, diferenciando esse contexto daquele em que é mais comumente utilizado, o de seleção de pessoas. Os autores descrevem o processo de desenvolvimento e a aplicação de um teste utilizado para avaliar militares norte-americanos depois de receberem um treinamento. Após serem treinados, os militares passam a exercer a função de conselheiros sobre oportunidades iguais (*equal opportunity advisors*), uma vez que o treinamento em questão visa habilitar os treinandos a colaborar com outros militares, aconselhando-os sobre como se comportar frente a situações de diversidade e de favorecimento à igualdade de oportunidades na diversidade.

Duas coletas distintas foram realizadas, a primeira com o grupo treinado em 2005, que contou com 46 militares oriundos do exército, da marinha e da força aérea, e a segunda com o grupo treinado em 2006, contando com 55 militares também provenientes das três distintas forças. As duas coletas foram realizadas com versões diferentes do instrumento TJS e, por esse motivo, o foco dos pesquisadores ateu-se aos 20 itens comuns às duas versões. Avaliando os resultados obtidos, Hauenstein et al. (2010) consideraram que os TJS são adequados para uso em contextos de treinamento, quando não há medidas específicas de comportamento e de desempenho para avaliar a transferência de conhecimento alcançada. Consideraram, também, que, mesmo que existam critérios de medida disponíveis para as avaliações, os TJS podem ser mais úteis que essas medidas, pois podem ser afinados para questões específicas à formação aplicada. Os autores apontam, ainda, que os TJS não se limitam a avaliar a melhoria do formando e que apresentam outros usos potenciais como diagnóstico de problemas com o sistema de treinamento, controle de qualidade dos resultados do treinamento e facilitação do desenvolvimento do treinando.

A pesquisa de Meijer et al. (2010) teve o objetivo de demonstrar que é possível desenvolver um TJS para medir um construto específico. Foi selecionado o construto integridade a ser medido pelo TJS e a validade desta medida foi verificada por meio de um estudo realizado com dados de uma amostra multiétnica. Sua composição foi de 203 candidatos em uma posição na academia de polícia da Holanda. Foram utilizados testes de habilidade cognitiva, um TJS desenvolvido pelos autores com o objetivo de medir a integridade, uma entrevista construída sobre o modelo HEXACO, que visa medir honestidade e humildade e um questionário denominado “*How I Think*”- *HIT*, que visa medir distorções cognitivas. O TJS foi desenvolvido com base em cenários, sendo apresentado para cada um deles um curto vídeo, relacionado aos principais problemas de integridade a que os policiais estão suscetíveis. No TJS, um narrador introduz cada cenário e a cada item a cena congela em um ponto importante. O candidato deve, então, apresentar as respostas relacionadas com a situação em questão.

Os resultados encontrados por Meijer et al. (2010) apontaram evidências de validade convergente entre os testes, com correlações de moderada a alta entre o TJS e a entrevista que mensurava honestidade e humildade, e entre o TJS e o questionário *HIT*. Apontaram, também, evidências de validade discriminante entre o TJS e o teste de habilidade cognitiva. Esses achados sustentam as duas hipóteses testadas: (a) que as pontuações do TJS para integridade se correlacionam com pontuações de outras dimensões relacionadas a ela e (b) que não haveria boa correlação do TJS com as habilidades cognitivas. Os achados indicam, ainda, que o TJS foi capaz de avaliar integridade em ambos os grupos participantes, o de maioria e o de minoria étnica.

McDaniel et al. (2011) tiveram o objetivo de compreender as estratégias para evitar pontuações extremas nos TJS. Os autores verificaram se os métodos que controlam a elevação e a variação da pontuação do TJS podem melhorar a validade do item e se podem reduzir as diferenças médias de pontuação em grupos étnicos compostos por brancos e negros. Foram utilizados diversos TJS em quatro estudos distintos. Nos estudos 1 e 2 foram utilizados TJS com cenários diferentes, sendo o estudo 1 composto de 20 cenários com 136 opções de respostas em uma escala Likert de 1 a 9. O estudo 2 foi composto por 31 cenários com 192 opções de respostas em uma escala de 6 pontos. Para os estudos 3 e 4 foram utilizados dois TJS com 10 cenários, um TJS em cada estudo, onde os respondentes deveriam escolher a melhor resposta entre duas opções.

Os resultados obtidos por McDaniel et al. (2011) apontaram que os ajustes nos métodos de pontuação dos TJS que controlam a elevação e a variação da pontuação melhoram a validade dos itens e reduzem significativamente as diferenças de escores entre os diferentes grupos. Além disso, esses ajustes podem eliminar a elevação dos escores associados com estratégias de treinamento para evitar as pontuações extremas. A validade de item é mostrada pela relação em forma de U com os itens significativos, ou seja, com altas e baixas pontuações interligadas podendo seguir uma direção como se formassem uma letra U, pois os itens de pontuação próxima ao ponto médio possuem menor variabilidade. Os estudos 1 e 4 foram consistentes em encontrar uma relação em forma de letra U entre itens significativos e validade do item. Os achados permitiram a retirada de itens sem perda da validade.

Também sob o contexto de diferenças de resultados entre brancos e negros, Roth et al. (2013) objetivaram padronizar as possíveis diferenças entre grupos étnicos e realizar uma análise de validade de instrumentos do tipo TJS, verificando como construtos sobre habilidades cognitivas e habilidades sociais poderiam influenciar as padronizações de diferenças entre grupos, avaliadas por meio do índice estatístico d (tamanho do efeito da amostra). A amostra foi composta por 1236 trabalhadores divididos em quatro grupos distintos segundo a profissão: contadores, técnicos em contabilidade, policiais de patrulhamento e funcionários do departamento de compras. Para cada profissão foi aplicado um TJS com questões específicas de seus ambientes de trabalho e outro teste que avaliasse as demais habilidades correlacionadas à profissão específica (habilidades cognitivas e sociais).

Os achados de Roth et al. (2013) foram convergentes com a literatura no que diz respeito à capacidade de avaliação dos TJS, isto é, indicam que esse é um método e pode ser designado para avaliar um número diferente de categorias de construto. Os construtos alvo tenderam à associação com diferentes valores (ou variações de valores) de d , confirmando as hipóteses de que os resultados obtidos por um TJS apresentam altos valores de d , quando associados a medidas de habilidades cognitivas, e apresentam baixos valores de d , quando associados a medidas de habilidades sociais. Os autores apontam que os resultados convergem com a literatura e que ajudam profissionais da área organizacional no desenvolvimento de testes, observando que o nível real dos resultados dos candidatos terá relação com o método e instrumento de seleção e que estes devem permitir comparações precisas. Assim, os resultados fornecem estimativas de níveis de pontos para o TJS, para que os especialistas possam ter um pouco mais de facilidade em estimar níveis de d e os impactos adversos que podem ocorrer durante o desenvolvimento do sistema de seleção.

DISCUSSÃO

Este artigo cumpriu com o objetivo de apresentar os fundamentos dos testes de julgamento situacional, trazendo à tona diversos debates existentes no contexto estrangeiro sobre seu formato e aplicabilidade em atuações relacionadas à psicologia organizacional, para além dos processos seletivos. Ainda, permitiu ao leitor conhecer algumas pesquisas recentes que têm sido conduzidas com os TJS, a fim de exemplificar suas aplicações, bem como traçar, ainda que de forma breve, um panorama dos avanços científicos sobre o tema. Por fim, este artigo cumpre também o objetivo de enunciar algumas lacunas que podem servir para fomentar pesquisas futuras no Brasil, onde o tema ainda é praticamente ignorado pelos pesquisadores.

Considerando que o tema é recorrente na literatura internacional, com evidências bastante convincentes acerca de sua utilidade e poder preditivo, e que a atuação de psicólogos brasileiros no campo da psicologia organizacional é bastante intensa, as informações contidas neste artigo indicam que a falta de interesse ou o desconhecimento por parte de pesquisadores e docentes brasileiros das áreas de avaliação psicológica e psicologia organizacional tem impedido que o conhecimento a respeito dos TJS chegue até profissionais e estudantes tanto nos níveis de graduação, quanto nos de pós-graduação. A falta de literatura em português também pode ser um fator que dificulta a ampliação do conhecimento, especialmente para os alunos de graduação. Nesta direção, esta pode ser uma importante contribuição do presente artigo, escrito com a intenção de apresentar de forma didática a técnica e um recorte da literatura a seu respeito.

Para instigar o estudo e o uso dos TJS no Brasil, pode-se afirmar que um primeiro passo é despertar o interesse de pesquisadores e estudantes de mestrado e doutorado para traduzir e adaptar instrumentos e modelos já consagrados em estudos estrangeiros e também para construir novos instrumentos para a realidade brasileira. Em seguida, faz-se importante disponibilizar instrumentos com boas qualidades psicométricas, especialmente com índices conhecidos de validade preditiva e incremental para os profissionais envolvidos

com a prática da psicologia organizacional, a fim de torná-los familiarizados com os alcances e as limitações da técnica. Vale lembrar que, conforme referido, há um instrumento já disponível para o Brasil nesse formato (Rabelo et al., 2012), mas que conta apenas com o manual técnico publicado, sem outros estudos disponibilizados, por exemplo, em artigos, o que limita sua acessibilidade. Junto a isso, a inclusão do tema em disciplinas de psicologia organizacional e avaliação psicológica, em cursos de graduação e pós-graduação, também pode ajudar a popularizar os TJS.

Futuras pesquisas devem levar em conta a realidade do mercado de trabalho brasileiro e as especificidades dos contextos nos quais os instrumentos poderão ser aplicados em avaliações. As questões relacionadas ao tipo de instrução, discutidas anteriormente, parecem ser essenciais, uma vez que suas implicações em termos de validade das interpretações dos instrumentos parecem ser importantes. Outro tópico importante refere-se à avaliação de construtos diversos por meio dos TJS, compreendendo que os mesmos constituem-se como métodos de avaliação. Testar a possibilidade de avaliar personalidade, autoeficácia, adaptabilidade de carreira, entre outras tantas variáveis, parece ser um assunto vasto, tanto no sentido de se construir os cenários e itens quanto em relação à utilização de diferentes instruções e métodos de resposta e na comparação com outros instrumentos que avaliem os mesmos construtos, mas por meio de instrumentos de autorrelato. Além disso, a expansão dos estudos para outras áreas, para além da psicologia organizacional, também parece algo promissor.

Por fim, é importante ressaltar que este artigo não teve a intenção de apresentar uma revisão sistemática da literatura sobre os TJS, mas apresentar aos leitores brasileiros os fundamentos desta técnica, bem como exemplificar sua aplicabilidade ao apresentar um recorte, pontual e limitado, da literatura sobre o assunto. Espera-se, de forma pretensiosa, que a partir desta contribuição, os testes de julgamento situacional possam ser popularizados, a fim de prover avaliações com maior poder preditivo para melhores tomadas de decisão no contexto da psicologia organizacional no Brasil.

REFERÊNCIAS

- Ambiel, R. A. (2013). *Construção e estudos psicométricos da Escala Ciclos de Adaptação de Carreira* (Tese de doutorado). Recuperado de <http://www.usf.edu.br/busca/?q=Constru%C3%A7%C3%A3o+e+Estudos+Psicom%C3%A9tricos+da+Escala+Ciclos+de+Adapta%C3%A7%C3%A3o+de+Carreira.&x=13&y=23>
- Bledow, R., & Frese, M. (2009). A Situational Judgment Test of personal initiative and its relationship to performance. *Personnel Psychology, 62*(2), 229–258. doi: 10.1111/j.1744-6570.2009.01137.x
- Brooks, M. E., & Highhouse, S. E. (2006). Can good judgment be measured? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational Judgment Tests: Theory, measurement, and application* (pp. 39-55). New Jersey: Lawrence Erlbaum Associates.
- Gessner, T. L., & Klimoski, R. J. (2006). Making sense of situations. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational Judgment Tests: Theory, measurement, and application* (pp. 13-38). New Jersey: Lawrence Erlbaum Associates.
- Gondim, S. M. G., & Queiroga, F. (2013). Recrutamento e seleção de pessoas. In L. O. Borges & L. Mourão (Orgs.), *O trabalho e as organizações: Atuações a partir da psicologia* (pp. 376-405). Porto Alegre: Artmed.
- Gondim, S. M. G., Borges-Andrade, J. E., & Bastos, A. V. B. (2010). Psicologia do Trabalho e das Organizações: Produção científica e desafios metodológicos. *Psicologia em Pesquisa, 4*(2), 84-99.
- Hanson, M. A., & Ramos, R. A. (1996). Situational Judgment Tests. In R. S. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 119-124). Westport: Quorum Books.
- Hauenstein, N. M. A., Findlay, R. A., & McDonald, D. P. (2010). Using Situational Judgment Tests to assess training effectiveness: Lessons learned evaluating military equal opportunity advisor trainees. *Military Psychology, 22*(3), 262-281. doi: 10.1080/08995605.2010.492679
- Lock M., Wheeler R., Burnard N., Cooper C. (2005). *Leadership Judgment Indicator*. Oxford: Hogrefe.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of Situational Judgment Tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730-740. doi: 10.1037/0021-9010.86.4.730
- McDaniel, M. A., & Nguyen, N. T. (2001). Situational Judgment Tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*(1/2), 103-113. doi: 10.1037/0021-9010.86.4.730
- McDaniel, M. A., Whetzel, D. L., Hartman, N. S., Nguyen, N., & Grubb, W. L. (2006). Situational Judgment Tests: Validity and an integrative model. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational Judgment Tests: Theory, measurement, and application* (pp. 183-204). New Jersey: Lawrence Erlbaum Associates.

- McDaniel, M. A., Psocka, J., Legree, P. J., Powell, A., & Weekley, J. A. (2011). Toward an understanding of Situational Judgment item validity and group differences. *Journal of Applied Psychology, 96*(2), 327-336. doi: 10.1037/a0021983
- Meijer, L. A. L., Born, M. P., Jaap, Z., & Molen, H. T. (2010). Construct-driven development of a video-based Situational Judgment Test for integrity a study in a multi-ethnic police setting. *Original Articles and Reviews, 15*(3), 229-236. doi: 10.1027/1016-9040/a000027
- Metcalf D., Dev H., Boursicot K., & Sales D. (2014). *Situational Judgement Test: Oxford assess and progress*. New York: Oxford University Press.
- Motowidlo, S. J., Dunnette, M.D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640-647. doi: 10.1037/0021-9010.75.6.640
- Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a Situational Judgment Test. *Journal of Applied Psychology, 95*(2), 321-333. doi: 10.1037/a0017975
- Phillips, J. F. (1992). Predicting sales skills. *Journal of Business and Psychology, 7*(2), 151-160. doi: 10.1007/BF01013925
- Phillips, J. F. (1993). Predicting negotiation skills. *Journal of Business and Psychology, 7*(4), 403-411. doi: 10.1007/BF01013754
- Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment, 11*(1), 1-16. doi: 10.1111/1468-2389.00222
- Rabelo I. S., Alves G. A. S., Ambiel R. A. M., Pacanaro S. V., & Leme I. F. A. S. (2012). *LJI - Indicador de Julgamento de Liderança – Manual*. São Paulo: Casa do Psicólogo.
- Roth, P. L., Bobko, P., & Buster, M. A. (2013). Situational Judgment Tests: The influence and importance of applicant status and targeted constructs on estimates of black–white subgroup differences. *Journal of Occupational and Organizational Psychology, 86*(3), 394-409. doi: 10.1111/joop.12013
- Salter, N. P., & Highhouse, S. (2009). Assessing managers' common sense using Situational Judgment Tests. *Management Decision, 47*(3), 392-398. doi: 10.1108/00251740910946660
- Schmitt N., & Chan D. (2006). Situational Judgment Tests: Method or construct? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp.135-155). New Jersey: Lawrence Erlbaum Associates.
- Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50*(1), 25-49. doi: 10.1111/j.1744-6570.1997.tb00899.x
- Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*(3), 679-700. doi: 10.1111/j.1744-6570.1999.tb00176.x
- Weekley, J. A., & Ployhart, R. E. (2006a). *Situational Judgment Tests: Theory, measurement, and application*. New Jersey: Lawrence Erlbaum Associates.
- Weekley, J. A., & Ployhart, R. E. (2006b). An Introduction to Situational Judgment Testing. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational Judgment Tests: Theory, measurement, and application* (pp. 1-10). New Jersey: Lawrence Erlbaum Associates.