

Avaliação da Escala *Likert* dos Itens do CSAI-2 em Atletas

Daniel Bartholomeu^{*,a}, José Maria Montiel^b & Afonso Antonio Machado^c

^aFundação Instituto de Ensino para Osasco, Osasco, Brasil, ^bFundação Instituto de Ensino para Osasco, Osasco, Brasil & ^cUniversidade Estadual Paulista, Rio Claro, Brasil

RESUMO

Este artigo investigou evidências de validade por estrutura interna para o Competitive State Anxiety Inventory (CSAI-2) em atletas pelo modelo de Rasch. Foram investigados 172 atletas de cinco modalidades esportivas distintas, com idades entre 14 e 58 anos com média de idade de 21 anos (DP=5,99). O instrumento avalia ansiedade em situação de competição e possui 27 itens em formato likert com quatro níveis possíveis de avaliação para assinalar a frequência de ocorrência. A análise das categorias de resposta dos itens revelou a pertinência de quatro níveis para a avaliação. Os resultados sugerem que uma reformulação de uma das categorias de análise do CSAI-2 deveria ser feita para possibilitar uma melhor compreensão dos itens.

Palavras-chave: CSAI-2; psicologia do esporte; análise de Rasch; avaliação psicológica; psicodiagnóstico.

ABSTRACT

CSAI's-2 likert rating scale assessment in athletes

This article aimed to investigate internal structure validity evidences to CSAI-2 in athletes by means of Rasch model. 172 athletes, from five different sports modality, aged from 14 to 58 years old (mean age 21 years, DP=5,99) were studied. The scale is composed by 27 items assessed in a four point likert scale based on the occurrence frequency of the behaviors. The item response category analysis revealed the adequacy of three category levels in the items assessment. The results suggest a reformulation of one of the CSAI's-2 category response should be done to facilitate item comprehension.

Keywords: sport psychology; Rasch analysis; psychological assessment; psychodiagnostic.

Uma escala de avaliação em itens que apresentem várias categorias de resposta (como as escalas do tipo Likert, por exemplo) serve, basicamente para duas funções. Em primeira instância, permite ao pesquisador enfatizar aspectos relevantes das variáveis em questão como, por exemplo, de se observar a frequência ou a intensidade das mesmas nos indivíduos quando perguntas com respostas dicotômicas não fornecem uma quantidade de informação suficiente sobre o constructo em questão. Uma segunda utilidade diz respeito ao formato das respostas permitir, em alguns casos, optar por uma categoria intermediária. Apesar disso, caso os itens sejam ambíguos, não apresentem definições claras, ou quando há uma categoria neutra (no meio da escala de resposta), ou ainda um número

excessivo delas, um ruído pode ser introduzido no padrão de respostas obtido por um respondente, afetando os resultados do teste e sua validade (Smith, Wakely, de Kruif & Swartz, 2003).

No caso específico do *Competitive State Anxiety Inventory* (CSAI), foco desta pesquisa, desde sua criação não foram identificados estudos que tratassem a quantidade de categorias necessárias para avaliação dos itens. Neste sentido, como medida de avaliação, o CSAI foi construído por Martens, Burton, Rivkin e Simons (1980) com o objetivo de avaliar a ansiedade estado de atletas em situação de competição. Desde seu desenvolvimento, tem sido administrado a atletas de diferentes modalidades esportivas como volei, boxe, futebol, dentre outros. Os estudos fatoriais deste ins-

* Endereço para correspondência: Daniel Bartholomeu – d_bartholomeu@yahoo.com.br

trumento, via de regra, têm separado os fatores de ansiedade cognitiva e somática e seus índices de precisão na versão original variaram de 0,94 a 0,97.

A reação emocional da ansiedade varia de acordo com a intensidade de suas flutuações temporais. Assim, serenidade e calma fisiológica e subjetiva indicam a ausência de respostas de ansiedade. Níveis moderados são caracterizados por apreensão, nervosismo, preocupação e tensão. Já níveis muito elevados envolvem sentimentos intensos de medo e pensamentos catastróficos, assim como excitação fisiológica. Nesse contexto, o termo “ansiedade estado” compreende níveis de ansiedade experienciados momentaneamente, com caráter transitório (Spielberger, 1966; Smoll & Wiechman, 1999).

Em uma revisão de seus itens, surgiu o CSAI-2, sendo acrescentados indicadores como medo de dano físico e ansiedade generalizada, alguns deles modificados de outros instrumentos de ansiedade-estado (Deffenbacher, 1980). Após a análise de juízes e inúmeras análises estatísticas, dos 102 itens iniciais restaram 27 na versão final. A análise fatorial deste instrumento indicou três fatores, sendo um de ansiedade somática. O fator cognitivo foi separado em dois componentes, um deles indicando autoconfiança positiva e um outro fator negativo de ansiedade cognitiva (Martens et al., 1990; Burton, 1989), sendo que cada item é sempre avaliado em 4 níveis possíveis que indicam a intensidade dos sintomas descritos para os atletas.

Brandão (2007) esclarece que o CSAI-2 tem sido um instrumento bastante utilizado na avaliação de ansiedade de atletas no Brasil, não se sabe se o número de categorias de interpretação dos itens é de fato adequado. Além disso, há que se questionar se o formato de apresentação dos itens do teste é compreendido da mesma forma por pessoas de diferentes culturas, modalidades esportivas, dentre outras questões, sendo útil sua análise para estudar o funcionamento das categorias de resposta dos itens em grupos diferentes.

Em relação às escalas politômicas, especificamente no caso do CSAI-2, cada número consecutivo da escala representa uma quantidade maior de ansiedade de competição, dependendo do item, já que há itens que indicam autoconfiança. Nesses casos, o aumento dos pontos da escala indica uma maior incidência desse aspecto. Há que se considerar que, invertendo-se os valores nesses itens, pode se referir à sintomas ansio-

gênicos. Isso implica que o aumento dos níveis de resposta corresponderia a níveis mais elevados de ansiedade ou autoconfiança. Conforme uma pessoa se movimentava no *continuum* dessas variáveis (quanto ao seu nível), cada ponto sucessivo torna-se, portanto, a resposta mais provável (pressupondo que os indivíduos conseguiram diferenciar entre todos os níveis da escala).

O modelo de teoria de resposta ao item supõe que a resposta a um item de um teste depende por um lado da habilidade das pessoas e por outro de características dos itens como o nível de dificuldade, discriminação e probabilidade de acerto ao acaso. Esses são os três parâmetros deste tipo de análise e os modelos derivados dele podem empregar 1, 2 ou 3 deles. Dentre os vários modelos de teoria de resposta ao item disponíveis, os indicadores da escala do CSAI-2, foram investigados quanto à sua adequação ao modelo de análise de itens proposto por Rasch. Esse sistema considera somente o parâmetro de dificuldade dos itens e habilidade das pessoas como função para determinar a probabilidade de acerto a um item. Esse modelo é o mais popular dentro da proposta da teoria de resposta ao item devido a sua grande simplicidade matemática (Muñiz, 1990).

O modelo de Rasch pode ser utilizado no processo de otimização do número de pontos e categorias dos itens de um teste sem a necessidade de administração de versões diferentes de uma mesma escala. Esse modelo pressupõe a aditividade do dado, definida como unidades de medida (*logits*) que têm o mesmo tamanho no contínuo (dados intervalares), caso o dado ajuste-se ao modelo. Essas medidas, por serem intervalares podem ser utilizadas em análises estatísticas paramétricas, dependendo, todavia, do quanto o dado se ajusta ao modelo de Rasch. Assim, esses parâmetros são estimados e utilizados na determinação dos padrões de resposta esperados para cada item. O ajuste é derivado de uma comparação desses com os padrões observados. Tal avaliação forneceria evidências de validade. Por sua vez, os erros padrão associados com a calibração de cada item e a estimação da habilidade das pessoas são utilizados no cálculo da precisão nesse modelo. Esses erros podem ser utilizados para descrever o intervalo de confiança em que a verdadeira dificuldade dos itens e habilidade das pessoas se situa (Wright & Stone, 1988).

Algumas formas de se observar a quantidade de categorias de resposta adequada aos itens de um teste

podem ser propostas. Primeiramente, pode-se fazer um exame das médias de habilidade dos respondentes observadas para cada categoria. Caso as médias dessa medida aumentem em razão das categorias dos itens da escala, implicaria que uma maior ansiedade está associada com categorias de resposta mais elevadas. Se uma desordem nessa progressão é verificada indicaria que um ponto da escala não representa uma maior quantidade de ansiedade em comparação com o nível anterior. Isso pode ocorrer quando números excessivos de pontos são dispostos nos itens de um instrumento; quando o sentido dos rótulos de cada um deles é ambíguo; ou quando categorias neutras são fornecidas.

Pode-se ainda observar os parâmetros dos limiares para determinar quais categorias não são efetivas na mensuração da variável sob investigação. Limiares ordenados implicam que enquanto uma pessoa se movimenta ao longo do contínuo da ansiedade, cada categoria se torna a resposta mais provável. A desordem nas respostas ocorre pelos mesmos motivos que anteriormente citado e pode ser mais bem observada por meio do gráfico de curvas de probabilidade.

Finalmente, uma análise do *outfit* pode detectar usos aleatórios de categorias de análise. O *outfit* avalia concordâncias inesperadas às categorias dos itens nas regiões extremas (com altas e baixas probabilidades de resposta aos itens) da curva característica dos itens (que descreve a função da probabilidade de acerto aos itens em razão do nível de dificuldade dos itens e habilidade das pessoas no caso do modelo de Rasch). Esse tipo de problema pode ser decorrente de baixas frequências para uma dada categoria, da percepção de seu significado ou da sua relação com outras categorias adjacentes (Linacre, 1997, 1999). Algumas possíveis soluções para esses desajustes podem ser unir categorias adjacentes, modificar a localização das categorias adjacentes, ou tratar as respostas como faltantes, caso o nome não seja adequado ou não compartilhe o mesmo traço que as demais (Linacre, 1997, 1999). Vale destacar que o diagnóstico deve ser feito levando-se em conta esses três critérios, além da inspeção visual do gráfico de probabilidades de concordância com as categorias da escala dos itens.

Todavia, embora essa análise forneça algumas diretrizes de quais categorias apresentam problemas potenciais, deve ser ressaltado que a decisão final por unir ou deletar uma dada categoria deve ser feita não

somente baseado no critério estatístico mas em suposições previstas na variável sob investigação. Além disso, a otimização de escalas depende da amostra estudada e deve ser testada novamente com uma nova amostra da mesma população (Smith, Wakely, de Kruif & Swartz, 2003).

Bond e Fox (2001), Lopez (1996) e Wright e Masters (1982) concordam que no processo de otimização de escalas, quando estão sendo comparadas várias formas de categorizações para a mesma escala politômica, além do diagnóstico das categorias, outros indicadores como os índices de precisão, separação e validade por meio das medidas de *infit* e *outfit* dos itens podem ser observados, fornecendo um dado adicional sobre o funcionamento das diferentes formas de categorização dos itens. Todos esses procedimentos foram adotados no presente estudo e serão analisados para cada formato de categorização. Tendo em vistas esses aspectos e as justificativas ora postas sobre o funcionamento das categorias do CSAI-2 no Brasil, o objetivo do presente trabalho foi analisar o funcionamento das categorias de resposta aos itens do CSAI-2 buscando uma estrutura que forneça bons dados de validade e precisão para essa escala em uma amostra de atletas.

MÉTODO

Participantes

Participaram da pesquisa 172 atletas de ambos os sexos, sendo 61,6% do sexo masculino. As idades variaram de 14 a 58 anos com média de 21 (DP=5,99). O nível de escolaridade também foi variado, indo do ensino fundamental (4,7%) ao nível de doutorado (0,6%), sendo que a maior parte dos sujeitos (62,8%) apresentou nível universitário incompleto. Ainda, 58,7% dos participantes trabalhavam e praticavam esporte, não se tratando de profissionais, embora a maior parte tenha uma rotina intensa de treinamentos, uma vez que treinavam, em sua maioria, de três a cinco dias na semana (78%), assim como de duas a três horas por dia (76,2%). Em relação à disputa de títulos esportivos, 90,7% relatou já ter competido alguma vez na vida. A pesquisa foi feita em diversos clubes e equipes do interior paulista e as modalidades esportivas foram as seguintes: Basquete (14,5%), Futebol (28,5%), Handebol (22,7%), Jiu-jitsu (13,4%), e Voleibol (20,9%).

Instrumento

Competitive State Anxiety Inventory (CSAI-2)

Utilizou-se a versão revisada com 27 itens que descrevem condutas de ansiedade física; cognitiva, indicando basicamente preocupação por situações de competição; e autoconfiança. Todos os comportamentos são solicitados a serem avaliados em uma escala de quatro pontos, variando de nada (1), alguma coisa (2), moderado (3) e muito (4), indicando a intensidade com que o participante apresenta algum dos aspectos descritos nos itens, circulando a alternativa desejada. Não há limite de tempo para a execução do teste.

Após os estudos iniciais do CSAI-2 feito por Martens e colaboradores (1990) inúmeros autores (Moraes, 2007; Moraes, Lobo & Lima, 2001) utilizaram esse instrumento para a avaliação da ansiedade-estado em situação de competição em várias modalidades esportivas. Moraes (2007) ressalta que na atualidade esse teste é um dos mais utilizados na mensuração da ansiedade em atletas no exterior e no Brasil. No Brasil, embora não tenham sido produzidas muitas pesquisas com o CSAI-2, algumas merecem destaque como é o caso do trabalho de Moraes, Lobo e Lima (2001) que descreve a avaliação de 275 atletas entre 12 e 14 anos que participaram dos Jogos da Esperança por meio do CSAI-2 48 horas antes da competição. Os autores não evidenciaram diferenças significativas entre os sexos no componente cognitivo. Em contrapartida, o componente somático foi mais característico dos atletas do sexo masculino. A mesma tendência foi observada em relação à autoconfiança. Também Moraes (1987) identificou uma maior pontuação de autoconfiança em atletas judocas homens que em mulheres. Esse fato foi explicado pelo autor com base na influência que a pressão familiar exerceria sobre a prática do judô, já que se acredita ser um “esporte masculino”.

Procedimento

A aplicação do instrumento foi coletiva e realizada sempre antes dos treinamentos esportivos em cada uma das modalidades por um psicólogo ou um professor de educação física devidamente treinados. Somente participaram da pesquisa os sujeitos que autorizaram ou foram autorizados pelos seus pais (por meio de assinatura do termo de consentimento livre e esclarecido, no caso dos atletas menores de idade). A aplicação não demorou mais que 10 minutos e aplicou-se o instrumento, em todas as ocasiões, em salas previa-

mente cedidas pelos clubes, com vistas a assegurar a padronização dos procedimentos. O projeto foi aprovado no Comitê de ética da Anhanguera Educacional sob o número 1241/2011.

RESULTADOS

Primeiramente, os indicadores do CSAI-2 para atletas em seu formato de resposta original (quatro níveis) foram investigados quanto à sua adequação ao modelo de análise de itens de Rasch. Para isso, a unidimensionalidade deve ser demonstrada como princípio básico para que o modelo possa operar (Bond & Fox, 2001). Uma das análises possíveis para demonstrar a unidimensionalidade de um teste é a análise de componentes principais de resíduos pelo modelo de Rasch. Alguns estudos de simulação sugeriram que fatores com *eigenvalues* menores que 1,4 poderiam ser considerados aleatórios (Smith & Miao, 1994). Quando isso ocorre, segue-se à seleção de itens com maiores saturações fatoriais nas dimensões positivas e negativas do primeiro fator extraído. Esses itens foram resubmetidos ao modelo de Rasch separadamente e correlacionadas às pontuações do fator positivo e negativo. O critério para se considerar a presença de outro fator é baseado na magnitude do coeficiente de correlação entre essas pontuações obtidas que segundo Linacre (2009) deve ser o menos próximo possível de 1 e para Sisto e colaboradores (2006) não deve ser superior a 0,51. Em outros termos, correlações acima deste patamar sugerem a possibilidade de unidimensionalidade no teste (Linacre, 2009; Sisto, Rueda & Bartholomeu, 2006).

O primeiro fator explicou 8,6 unidades de variância de 38,8. Considerando-se que o *eigenvalue* é superior a 1,4 procedeu-se à análise fatorial dada a possibilidade de um outro fator (Smith & Miao, 1994). Os resultados do primeiro contraste estão na Tabela 1. Por esses dados, percebe-se que os itens com cargas fatoriais acima de 0,5 são os maiores em cada um dos dois fatores. Assim, selecionaram-se os itens com maior saturação no contraste negativo e positivo, que foram submetidos novamente ao modelo de Rasch para o novo cálculo das pontuações totais dos sujeitos nesses dois fatores que foram posteriormente correlacionadas pela prova de *Pearson* com coeficiente de 0,62 ($p=0,000$). Este coeficiente sugere que os dois conjuntos de itens cujos resíduos (diferenças do dado observado e teórico pelo modelo de Rasch) estavam mais discrepantes, compartilham variância suficiente

(acima de 0,51) para serem considerados provenientes de uma única dimensão. Além disso, a maior parte dos itens apresentou *infit* e *outfit* (*mean squares*) próximas

ou abaixo do valor 1, considerado por Linacre (2009) como uma intensificação da primeira dimensão e não uma segunda.

Tabela 1

Estatísticas do Fator 1, Extraído por Análise de Componentes Principais dos Resíduos, em Relação aos Itens do CSAI-2

Contraste	Carga	Infit (Mnsq)	Outfit(Mnsq)	Measure	Item
1	0,78	1,02	0,96	-0,64	CSAI12
1	0,76	1,04	1,05	-0,09	CSAI14
1	0,74	1,02	1,07	-0,78	CSAI24
1	0,72	1,00	1,55	-0,11	CSAI3
1	0,71	0,98	1,30	-0,92	CSAI15
1	0,70	1,46	1,07	-0,20	CSAI6
1	0,68	1,00	1,02	-1,03	CSAI9
1	0,64	0,89	1,01	-0,92	CSAI18
1	0,64	1,03	1,03	-0,10	CSAI21
2	-0,57	0,83	0,81	0,09	CSAI22
2	-0,56	0,86	0,85	-0,03	CSAI2
2	-0,54	0,81	0,81	0,35	CSAI8
2	-0,54	0,83	0,81	0,02	CSAI16
2	-0,53	0,76	0,75	0,12	CSAI17
2	-0,50	0,86	0,83	0,58	CSAI5

Analizada a unidimensionalidade dos dados do CSAI-2 (pressuposto básico para se estabelecer uma análise pelo modelo de Rasch), a precisão fornecida por esse modelo para os itens forneceu um índice de 0,97 para os itens e 0,71 para as pessoas, indicativo de uma alta precisão tanto para os itens como para as pessoas. O erro médio de medida foi de 0,21 (DP=0,01) para os itens e de 0,08 (DP=0,01). A precisão das pessoas indica a estabilidade das respostas das pessoas que poderíamos esperar se essa amostra fosse submetida a outro grupo de itens semelhantes que avaliam o mesmo construto. A precisão é melhorada por um erro pequeno nas estimativas das habilidades das pessoas, como foi o caso. A precisão dos itens informa a estabilidade dos itens se os mesmos indicadores fossem aplicados a outra amostra com níveis de habilidades similares (Bond & Fox, 2001). Essas informações constam da [Tabela 2](#).

Nesse ponto cabe esclarecer o parâmetro de dificuldade dos itens em termos de ansiedade. Assim, quanto maior a dificuldade de um item, menos concordância é manifestada naquele particular, sendo uma situação ou uma faceta da ansiedade de competição mais incomum, sugerindo, caso o item não apresente significado invertido, que pessoas mais ansiosas manifestariam maior concordância. Por sua vez, caso uma pessoa apresente alta concordância com itens que avaliam níveis de ansiedade em competição maiores, seria sugestivo de que essa apresenta uma maior incidência de sintomas físicos e cognitivos de ansiedade e baixa autoconfiança, sendo também elevada a probabilidade de apresentar alta concordância com itens que mensuram níveis inferiores de ansiedade de competição (itens mais fáceis).

Tabela 2
Estatísticas do Modelo de Rasch dos Itens e das Pessoas (N=172)

Parâmetros	Itens				Pessoas			
	Dificuldade	Infit	Outfit	Erro	Habilidade	Infit	Outfit	Erro
Média	0,00	1,00	1,01	0,08	-0,03	1,00	1,01	0,21
DP	0,54	0,21	0,24	0,01	0,42	0,42	0,46	0,01
Máximo	1,05	1,54	1,63	0,10	1,10	2,72	3,13	0,25
Mínimo	-0,96	0,71	0,71	0,08	-1,11	0,18	0,19	0,20

Em relação ao ajuste ao modelo de Rasch duas medidas devem ser observadas, o *infit*, que informa sobre discrepâncias do dado observado e teórico na região em que a probabilidade de acerto ao item é próxima a 50%, em outros termos, na região central da curva característica dos itens; e o *outfit*, que corresponde a acertos inesperados nas regiões extremas da curva, com altas e baixas probabilidades de acerto aos itens. Esse ajuste é dado tanto para os itens quanto para as pessoas e Bond e Fox (2001) consideram que o patamar 1,00 seria um padrão esperado em cada uma dessas medidas. Wright e Linacre (1989) consideram valores entre 0,6 e 1,4 como indicativos de bom ajuste para escalas politômicas de autorrelato. Ressaltam, todavia, que esses valores podem variar em razão do propósito para o qual a escala foi construída. Entretanto, há autores que consideram como bom ajuste, estimativas dentro do intervalo de 0,70-1,30 (Sisto, Rueda & Bartholomeu, 2006), parâmetro esse, adotado na presente pesquisa, com vistas a restringir mais o critério de ajuste, obtendo-se um padrão mais conciso. Também Linacre (2002) esclarece que medidas até 1,50 seriam consideradas o limite máximo para aceitação de um item.

Os itens forneceram média de *infit* de 1,00 (DP=0,21), indicando um bom ajuste. Analisando mais detidamente esses dados, observa-se que a variação de *infit* esteve dentro do intervalo de 1,54-0,71, sugerindo que nem todos os itens estiveram ajustados ao modelo. Dos itens, 3,70% (1 item) obtiveram valores de *infit* acima de 1,50; outros 7,40% (2 itens) acima de 1,30. Nenhum item apresentou este índice abaixo de 0,70. Assim, 11% dos itens evidenciaram desajustes de *infit*.

Quanto ao *outfit*, a média foi 1,01 (DP=0,24), indicando que a maioria dos itens esteve dentro dos parâmetros esperados. O intervalo de variação nessa medida foi de 1,63-0,71, indicativo de que houve itens com desajustes. De fato, 14,80% evidenciaram pro-

blemas, sendo cerca de 7,40% (2 itens) acima de 1,50, 7,40% (2 itens) acima de 1,30, e nenhum abaixo de 0,70. Todavia, sugerem que um olhar mais atento deva ser dado a esses itens. Houve uma maior quantidade de discrepâncias importantes na medida de *outfit* que de *infit* (patamar 1,50) e esses itens devem ser analisados quanto à sua compreensão ou pertinência à escala, já que ocorreram altas concordâncias a itens inesperadas em razão da habilidade (ansiedade) dos sujeitos. Talvez a análise da redação destes itens, já que se trata de uma adaptação para a realidade brasileira seja útil, eliminando-se problemas de compreensão e de adequação da língua. Tal fato convida a novas pesquisas.

Em relação ao ajuste das pessoas ao modelo de Rasch, em termos de *infit* e *outfit*, de forma geral, esteve em níveis esperados, já que suas médias foram 1,00 (DP=0,42) e 1,01 (DP=0,46) respectivamente. Analisando mais detidamente esse aspecto, evidenciou-se um intervalo de variação para o *infit* de 2,72-0,18 e de *outfit* de 3,13-0,19 sendo que, 6 pessoas (3,73%) apresentaram índices de *infit* acima de 1,30 e 10 pessoas (5,88%), acima desse patamar em *outfit*. Considerando o critério de Linacre (2002), 17 sujeitos (10%) superaram o critério de 1,50 em *infit* e outros 16 (9,41%) em *outfit*. No critério de 0,70, 27 pessoas (15,88%) e outras 17 (10%) apresentaram problemas no *infit* e *outfit* respectivamente. Em suma, o total de discrepâncias encontrado no *infit* foi de 29,61% e no *outfit*, 25,29%, que coloca quase um quarto dos sujeitos com problemas na solução dos itens, devendo-se considerar em que medida os atletas compreenderam a tarefa solicitada.

A ansiedade em competição dos atletas teve uma média de -0,03 (DP=0,42), com um intervalo de variação de theta de -1,11-1,10. Por sua vez, a média de dificuldade dos itens (nível de ansiedade avaliado pelos itens) em logit, foi 0,00 (DP=0,54) e teve seu intervalo de -0,96 até 1,05. Esse dado indica que os

itens mensuraram cada um dos níveis de ansiedade, havendo atletas suficientes para prover informação sobre cada um dos indicadores (o que também foi confirmado pelo exame do mapa de itens). Essa variabilidade não foi grande, e a maior parte dos itens esteve no intervalo de -1 e 1 avaliando a ansiedade das pessoas com maior precisão (uma vez que ocorreu maior variabilidade de informação nesse intervalo).

Investigou-se também se a quantidade de categorias apresentada nos itens do CSAI-2 seria representativa do traço latente sob investigação, nesse caso a an-

siedade de competição. Nesse contexto, conforme uma pessoa avança no contínuo de habilidade, cada um dos pontos subsequentes da escala torna-se a resposta mais provável (Linacre, 1999; Smith, Wakely, de Kruif & Swartz, 2003).

Os métodos utilizados para se delimitar a quantidade de categorias de análise necessária para representar adequadamente a ansiedade nesse caso foram os já indicados na introdução, conforme as sugestões de Linacre (1997; 1999). Os detalhes dessa análise estão na Tabela 3 e na Figura 1.

Tabela 3

Médias das Medidas nas Categorias, Limiares, Erro Padrão dos Limiares e Medidas de Infit e Outfit para a Escala do CSAI-2 (N=172)

Rótulo da Categoria	Número	Infit Mnsq	Outfit Mnsq	Medida da Categoria	Medida de estrutura	Erro padrão
Nada	1	0,95	1,00	-1,83	----	----
Alguma coisa	2	1,09	1,15	-0,53	-0,29	0,04
Moderado	3	0,89	0,87	0,47	-0,24	0,03
Muito	4	1,07	1,03	1,91	0,53	0,04

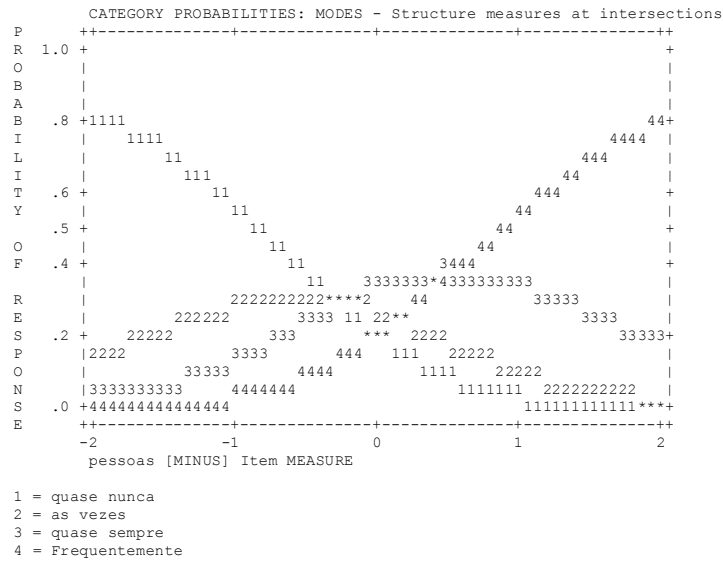


Figura 1. Curvas de probabilidade de categorias para a escala do CSAI-2 (N=172).

A análise da progressão da medida Rasch indicou um aumento da habilidade das pessoas em razão das categorias de resposta. Por sua vez, a medida de *Outfit*, sugeriu bons ajustes em todas as categorias (Figura 1). Finalmente, analisando-se os limiares da categoria “alguma coisa” (categoria 2), verificou-se uma descontinuidade. Dentre as soluções propostas para esse

aspecto, optou-se por combinar as categorias de resposta, nesse caso, “alguma coisa” e “moderado”, que apresentaram limiares bastante próximos. Os resultados dessa análise estão na Tabela 4 e Figura 2. Esses dados revelam dados apropriados para a estrutura de três categorias de análise.

Tabela 4

Médias das Medidas nas Categorias, Limiares, Erro Padrão dos Limiares e Medidas de Infit e Outfit para a Escala de Três Níveis do CSAI-2 (N=172)

Rótulo da Categoria	Número	Infit Mnsq	Outfit Mnsq	Medida da Categoria	Medida de estrutura	Erro padrão
Nada	1	0,97	1,08	-1,60	---	---
Moderado	3	0,86	0,90	0,23	-1,15	0,04
Muito	4	1,03	1,00	2,26	1,15	0,04

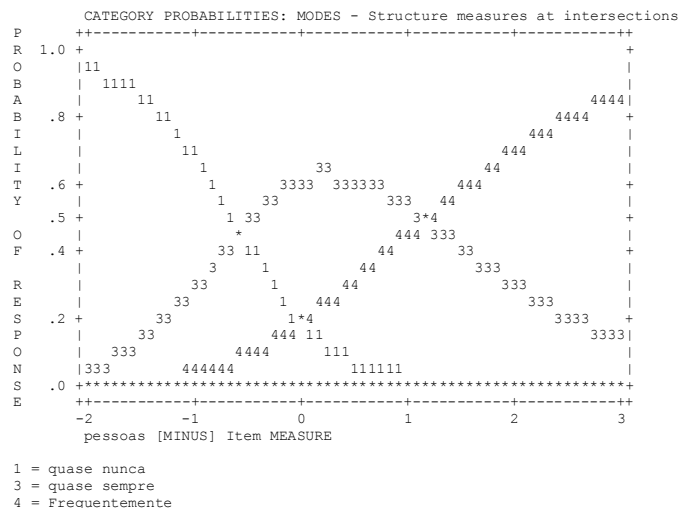


Figura 2. Curvas de probabilidade de categorias para a escala de três níveis do CSAI-2 (N=172).

Investigou-se o ajuste dos itens ao modelo de Rasch novamente para averiguar se as evidências de precisão e validade dos itens foram melhoradas com esse novo formato de respostas. De fato, observou-se uma piora nesses resultados conforme pode ser evi-

denciado na Tabela 5. Para os itens, a precisão foi 0,97, e para as pessoas 0,75, sendo ligeiramente superior em relação à estrutura de 4 níveis de análise, enquanto para os itens o valor foi o mesmo.

Tabela 5

Parâmetros de Ajuste de Itens e das Pessoas com a Versão de Quatro Níveis de Análise (N=172)

Parâmetros	Itens				Pessoas			
	Dificuldade	Infit	Outfit	Erro	Habilidade	Infit	Outfit	Erro
Média	0,00	1,02	1,02	0,09	0,06	0,97	1,02	0,22
DP	0,56	0,28	0,33	0,02	0,44	0,41	0,53	0,02
Máximo	0,96	1,62	1,81	0,13	1,11	2,61	3,70	0,30
Mínimo	-1,11	0,65	0,61	0,07	-1,04	0,24	0,32	0,19

A dificuldade dos itens, indicativa do nível de ansiedade avaliado pelos mesmos, indicou média semelhante nas duas estruturas, embora o valor máximo tenha sido menor e o mínimo maior, avaliando níveis de ansiedade mais baixos do que o outro formato. Analisando as médias e valores máximos e mínimos

de *infit* e *outfit* dos itens, observou-se que a média de ambas as medidas foi maior com essa estrutura em detrimento da outra, embora indique ainda que os itens, de forma geral, estiveram ajustados ao modelo. Todavia, os limites máximos e mínimo também foram aumentados, sugerindo um número maior de desajustes.

Na medida de *infit*, sete itens demonstraram desajustes (26%) sendo 2 acima de 1,5 (7,4%), quatro acima de 1,3 (14,8%) e um item abaixo de 0,7 (3,7%). Já no *outfit*, o total de desajustes foi nove itens (33,3%), sendo quatro acima de 1,5 (14,8%), dois acima de 1,3 (7,4%) e três abaixo de 0,7 (11,11%).

Em relação às pessoas, a ansiedade média foi superior nessa estrutura, apresentando limites máximo e mínimo bastante próximos, avaliando os mesmos níveis de habilidade basicamente. Em relação ao ajuste das pessoas, o *infit* apresentou média inferior do que na estrutura inicial, assim como o valor máximo menor e mínimo maior, sugerindo um menor número de pessoas ajustadas nessa medida assim como no *outfit* que apresentou média igual, mas valor máximo maior e mínimo menor. Mais especificamente 15 pessoas apresentaram valores de *outfit* acima de 1,5 (8,8%), oito pessoas acima de 1,3 (4,7%) e 26 abaixo de 0,7 (15,3%), somando um total de 49 pessoas com desajustes nessa medida (27,3%). Em relação ao *infit*, 13 pessoas evidenciaram valores acima de 1,5 (7,6%), sete acima de 1,3 (4,1%) e 26 abaixo de 0,7 (15,3%). O total de pessoas com problemas de *infit* foi 46 (27%).

Contrariamente à estrutura original, os pequenos desajustes ocorridos nas medidas de ansiedade de competição dos atletas e dificuldade dos itens (níveis de ansiedade medidos por eles) fizeram que nem todos os itens tivessem pessoas suficientes para prover boa informação, já que avaliam níveis mais baixos de ansiedade em que não houve muitos sujeitos. Também ocorreram alguns casos de pessoas com ansiedade muito alta que os itens não foram capazes de captar. A maior parte dos itens manteve-se no intervalo de -1 e 1, indicativo de maior precisão. Aparentemente, houve uma alteração no nível de dificuldade dos itens, e principalmente de habilidade estimada das pessoas ao se alterar a estrutura dos itens o que pode ter ocorrido pelo aumento de variabilidade de respostas em uma categoria do contínuo de avaliação dos itens (escala Likert) (Bond & Fox, 2001).

DISCUSSÃO

Este estudo foi proposto pensando-se no fato de que o CSAI-2, um dos instrumentos mais utilizados no Brasil, na atualidade, não apresenta estudos de análise de itens com uma amostra brasileira (Moraes, 2007). Reforçando essa necessidade, Brandão (2007) e Moraes (2007) mencionaram que a maior parte dos instrumentos de avaliação psicológica de atletas no

país não apresenta estudos de validade suficientes. Além disso, partiu-se da premissa que itens avaliados em uma dada escala, possivelmente não sejam semelhantes em culturas diferentes, grupos etários e outras características

Desse modo, este trabalho se deteve na investigação das qualidades dos itens desse instrumento por meio do modelo de Rasch. Mais especificamente, procurou-se otimizar as categorias de resposta dos itens de tal forma que os dados de validade e precisão fossem satisfatórios.

Em um primeiro momento foi analisado o ajuste dos itens e das pessoas à esse modelo em sua estrutura original, ou seja, com quatro níveis de análise para os itens, quais sejam, nada (1), alguma coisa (2), moderado (3) e muito (4), indicando a intensidade com que os participantes apresentavam algum dos sintomas descritos. Essa análise demonstrou que poucos itens apresentaram desajustes nas medidas de *infit* e *outfit*, indicativo de acertos ou erros inesperados em razão da dificuldade dos itens. Quanto às pessoas, aproximadamente um quarto delas na amostra em questão evidenciaram problemas na solução dos itens com um padrão de respostas não esperado, considerando os níveis de ansiedade desses sujeitos. Esses dados sugerem que pessoas com alta ansiedade, às quais seria esperado respostas elevadas em itens que avaliam níveis baixos de ansiedade, não apresentaram esse padrão e vice-versa, caracterizando os desajustes.

Embora o nível de ansiedade dos atletas tenha apresentado uma variação um pouco maior que a variação de dificuldade dos itens (nível de ansiedade medido por cada um deles), suas médias foram coincidentes e houve sujeitos suficientes em cada um dos indicadores para fornecer informação sobre eles em todos os níveis de dificuldade analisados pelos itens. Essa análise de itens por esse modelo não havia sido feita com o CSAI-2 por nenhum autor na literatura.

Por fim, estudou-se a quantidade de categorias necessárias para a avaliação da ansiedade nos itens. A categoria “alguma coisa” não demonstrou uma boa discriminação da ansiedade das pessoas nos itens do instrumento, sendo agregada à categoria “moderado”. Desse modo, demonstrou-se que três categorias de análise [nada (1), moderado (2), e muito (3)] seriam mais adequados para a avaliação dos itens pelos atletas brasileiros pela análise das categorias dos itens. Possivelmente, a categoria “alguma coisa” possa não ter sido bem interpretada pelos respondentes, já que

seu rótulo não fornece uma indicação precisa da intensidade que se deseja expressar. Seria interessante que novas pesquisas investiguem as possíveis causas desse fato.

Apesar da estrutura dos itens de três níveis de resposta ter demonstrado melhores propriedades estatísticas com uma progressão das médias de ansiedade e dos limiares melhores, assim como valores de *outfit* e o gráfico de probabilidades apresentar curvas mais discriminativas de cada uma das categorias, Bond e Fox (2001) Lopez (1996) e Wright e Masters (1982) ressaltam que ao serem comparadas várias formas de avaliação dos itens, deve-se demonstrar uma melhora nos índices de precisão e validade por meio das medidas de ajuste, *infit*, *outfit* e o funcionamento diferencial dos itens. Em outros termos, a nova estrutura de níveis dos itens deveria evidenciar um melhor funcionamento dos mesmos como uma redução de erros e vieses na medida e melhora nas evidências de validade.

De fato, a nova estrutura de três itens a quantidade de desajustes ao modelo de Rasch (tomados pelas medidas de *infit* e *outfit*) tanto para as pessoas como para os itens foi superior, sugerindo uma quantidade maior de respostas inesperadas aos indicadores. Em outras palavras, houve pessoas muito ansiosas que seria esperado alta concordância com itens que avaliam altos índices de ansiedade, o que não foi observado em inúmeros casos, assim como pessoas com baixa incidência de ansiedade que manifestaram alta concordância em itens que avaliavam altos níveis dessa variável.

Associado a isso, observou-se que, nessa estrutura, não houve pessoas suficientes para prover boa informação sobre alguns itens, assim como pessoas muito ansiosas que os indicadores não foram capazes de captar. Esse fato não ocorreu na estrutura de quatro níveis.

De fato, a alteração dos níveis de análise de quatro para três, apesar de melhorar as qualidades das escalas politômicas, não evidenciou grandes melhoras nos níveis de precisão e piorou os indicadores de validade, tomados pelo ajuste ao modelo Rasch. Isso pode ter ocorrido pela possibilidade de se deixar uma categoria mediana que não força o atleta a se decidir sobre o quanto o sintoma está presente, podendo optar pelo meio. Apesar dessa sugestão, é algo que convida a novas pesquisas, já que não foi analisada a resposta efetiva dos atletas após a recategorização. Assim, poder-se-ia manter os quatro níveis de análise, clarifi-

cando, entretanto o significado da categoria “alguma coisa” que não indica a intensidade desejada e pode causar ambiguidade no momento de responder aos itens. Essa categoria foi a que mais apresentou problemas.

Novos estudos deveriam ser feitos ainda testando-se a estrutura de três níveis em busca de uma melhora em suas propriedades estatísticas com amostras maiores ou com outras modalidades esportivas, embora o mais adequado pareça ser uma reformulação no título das categorias com quatro níveis de resposta, como sugerido no presente estudo. Além disso, seria interessante investigar a estrutura fatorial desse instrumento com as quatro categorias de interpretação dos itens (que demonstrou melhor ajuste ao modelo de Rasch e menos indicadores com funcionamento diferencial) excluindo-se os indicadores com muitos problemas (já indicados) ou separando-se duas escalas em razão do sexo dos atletas, já que existem indicadores com esse tipo de viés. Deve-se ter cautela, todavia, com essa solução. O funcionamento diferencial sugere que nas repostas aos itens com esse tipo de problema, outras variáveis que não o traço latente (nesse caso a ansiedade-estado) estariam exercendo efeito, características essas, peculiares ao grupo favorecido. Nesses termos, separando-se as escalas, por um lado, assegurar-se-ia o controle por inclusão dessa característica, tornando-a constante; por outro lado, o que deve ser medido pelo instrumento é o traço (nesse caso, a ansiedade), supondo unidimensionalidade que seria perdida, caso estivessem presentes características dessa natureza. O funcionamento diferencial dos indicadores por modalidade esportiva também seria outro aspecto interessante a ser investigado.

Por fim, espera-se que este estudo possa incitar o desenvolvimento de pesquisas com a ansiedade de atletas no Brasil, já que os indicadores do CSAI-2 mostraram características psicométricas iniciais adequadas que permitem seu uso em avaliações de atletas, podendo ser uma ferramenta bastante útil para esse fim.

REFERÊNCIAS

- American Psychological Association, American Educational Research Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Andriola, W. B. (2001). Descrição dos principais métodos para detectar o funcionamento diferencial dos itens (DIF). *Psicologia: Reflexão e Crítica*, 14, 643-652.

- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. London: Lawrence Erlbaum Associates.
- Brandão, M. R. F. (2007). A psicologia do exercício e do esporte e seus desafios para o milênio. In M. R. F. Brandão & A. A. Machado (Eds.), *Coleção psicologia do esporte e do exercício: Teoria e aplicação* (pp. 143-157). São Paulo: Atheneu.
- Burton, D. (1989). Winning isn't everything: Examining the impact of performance goals on collegiate swimmers' cognitions and performance. *The Sport Psychologist, 15*, 105-132.
- Deffenbacher, L. I. (1980). Worry and emotionality in test anxiety. In I. G. Samson (Ed.), *Test anxiety: Theory, research and implications* (pp. 111-128). Hillsdale, NJ: Erlbaum.
- Draba, R. E. (1977). The identification and interpretation of item bias. *Rasch Measurement Transactions, MESA Memorandum, 25*. Retirado de <http://www.rasch.org/rmt/rmt122m.htm>
- Gould, D., Petlichkoff, L., & Weinberg, R. S. (1984). Antecedents of, temporal changes and relationships between CSAI-2 sub-components. *Journal of Sport Psychology, 5*, 289-304.
- Jensen, A. R. (1980). *Bias in mental testing*. New York, NY: Free Press.
- Jones, J. G., & Cale, A. (1989). Precompetition temporal patterning of anxiety and self-confidence in males and females. *Journal of Sport Behavior, 12*, 183-195.
- Krane, V., & Williams, J. (1994). Cognitive anxiety, somatic anxiety, and confidence in track and field athletes: Impact of gender, competitive level and task characteristics. *Journal of Sport Psychology, 25*, 203-217.
- Linacre, J. M. (1997). *Guidelines for rating scales*. Retirado de <http://mesa.spc.uchicago.edu/rn2.htm>.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 2*, 103-122.
- Linacre, J. M., (2002). What do Infit and Outfit, mean-squared and standardized mean? *Rasch Measurement Transactions, 16*, 878. Retirado de <http://209.238.26.90/rmt/rmt82a.htm>
- Linacre, J. M., (2009). *Winsteps Rasch measurement computer program*. Chicago, IL: Winsteps.
- Lopez, W. A. (1996). *The resolution of ambiguity: An example from reading instruction* (Tese de doutorado). Universidade de Chicago, Chicago.
- Martens, R., Burton, I. T., Rivkin, F., & Simon, J. (1980). Reliability and validity of the Competitive State Anxiety Inventory (CSAI-2). In C. L. I. Nadeau, W. C. Halliwell, K. M. Newell, & G. C. Roberts (Eds.), *Psychology of motor behavior and sport* (pp. 91-99). Champaign, IL: Human Kinetics.
- Martens, R., Vealey, K-S., & Burton, A. (1990). *Competitive anxiety in sports*. Champaign, IL: Human Kinetics.
- Martinelli, S. C., & Bartholomeu, D. (2007). Escala de motivação acadêmica: Uma medida de motivação extrínseca e intrínseca. *Avaliação Psicológica, 6*, 21-32.
- Merbitz, C., Morris, J., & Grip, J. C. (1989). Ordinal scales and foundations of misinference. *Archives of Physical Medicine and Rehabilitation, 70*, 308-312.
- Moraes, L. C. (1987). *The relationship of anxiety and performance of Brazilian judokas* (Tese de doutorado). Universidade de Michigan, Michigan.
- Moraes, L. C., Lobo, I. L. B., & Lima, O. M. S. (2001). Análise da ansiedade pré-competitiva de atletas de diferentes esportes. In E. S. Garcia & L. C. Moraes (Eds.), *Olimpíada colegial jogos da esperança 2001* (pp. 23-32). Poços de Caldas, MG: Resultados.
- Moraes, L. C. C. A. (2007). Emoções no esporte e na atividade física. In M. R. F. Brandão & A. A. Machado (Eds.), *Coleção Psicologia do esporte e do exercício: Teoria e aplicação* (pp. 125-142). São Paulo: Atheneu.
- Pajares, F., Hartley, J., & Valiante, G. (2000). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development, 33*, 214-221.
- Passer, M. W. (1983). Fear of failure, fear of evaluation, perceived competitive-trait-anxious children. *Journal of Sport Psychology, 5*, 172-188.
- Sisto, F. F. (2006). O funcionamento diferencial do item. *Psico-USF, 11*, 35-43.
- Sisto, F. F., Rueda, F. J. M., & Bartholomeu, D. (2006). Estudo sobre a unidimensionalidade do teste Matrizes Progressivas Coloridas de Raven. *Psicologia: Reflexão e Crítica, 19*, 66-73.
- Smith, E. V., Wakely, M. B., De Kruif, R. E. L., & Swartz, C. W. (2003). Optimizing rating scales for self-efficacy (and other) research. *Educational and Psychological Measurement, 63*, 369-391.
- Smith, R. M., & Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (pp. 132-154). Norwood NJ: Ablex.
- Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing and Health, 13*, 227-236.
- Willis, J. D. (1982). Three scales to measure competition-related motives in sport. *Journal of Sport Psychology, 4*, 338-153.
- Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal: Measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation, 70*, 857-860.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: MESA.
- Wright, B. D., & Stone, M. H. (1988). *Validity in Rasch measurement: Research memorandum 54*. Chicago, IL: MESA.

Recebido: 04/05/2012
 Última revisão: 27/12/2012
 Aceito: 26/02/2013

