



Evidências de Validade de uma Prova de Letramento Estatístico: um estudo com estudantes universitários de cursos tecnológicos

Evidences of Validity of a Statistics Literacy Test: a study with college students of technological courses

Verônica Yumi Kataoka*

Claudette Maria Medeiros Vendramini**

Claudia Borim da Silva***

Maria Helena Palma de Oliveira****

Resumo

O estudo buscou evidências de validade numa prova de letramento estatístico, com seis questões, aplicada a 236 universitários de seis cursos tecnológicos, recém-concluintes da disciplina Estatística. As respostas foram analisadas de acordo com categorias apresentadas por Watson e Callingham, construídas com base na interação da Taxonomia SOLO e estágios de conhecimento de contexto. A classificação das respostas foi feita com o auxílio do mapa de itens do modelo de créditos parciais de Rasch. Propriedades psicométricas das questões indicam evidências de validade da prova para avaliar o

* Doutora em Estatística e Experimentação Agropecuária pela Universidade Federal de Lavras (UFLA). Professora da Universidade Bandeirante de São Paulo (UNIBAN), Programa de Pós-Graduação em Educação Matemática, São Paulo, SP, Brasil. Endereço para correspondência: Av Braz Leme, 3029, Santana, CEP: 02022-011, São Paulo, SP, Brasil. *E-mail*: veronicayumi@terra.com.br

** Doutora em Educação pela Universidade Estadual de Campinas (UNICAMP). Professora da Universidade São Francisco (USF), Programa de Pós-Graduação em Psicologia, Itatiba, SP, Brasil. Endereço para correspondência: Rua Alexandre Rodrigues Barbosa, 45, Centro, CEP: 13251-900. Itatiba, SP, Brasil. *E-mail*: cvendramini@uol.com.br

*** Doutora em Educação Matemática pela Pontifícia Universidade Católica de São Paulo (PUC-SP). Professora da Universidade São Judas Tadeu (USJT), São Paulo, SP, Brasil. Endereço para correspondência: Rua Taquari, 546, Mooca, CEP: 03166-000, São Paulo, SP, Brasil. *E-mail*: dasilvm@uol.com.br

**** Doutora em Psicologia pela Universidade de São Paulo (USP). Professora da Universidade Bandeirante de São Paulo (UNIBAN), Programa de Pós-Graduação em Educação Matemática, São Paulo, SP, Brasil. Endereço para correspondência: Av Braz Leme, 3029, Santana, CEP: 02022-011. São Paulo, SP, Brasil. *E-mail*: mhelenapalma@terra.com.br

construto *letramento estatístico*, apresentando consistência interna satisfatória e valores de *infit* e *outfit* adequados. Foi comparada, também, a ordem de dificuldade das categorias de respostas obtidas no estudo de Watson e Callingham com os determinados neste trabalho. Os resultados indicam que mesmo num contexto de aplicação diferente (*background* estatístico, faixa etária), a escala de letramento estatístico pode ser utilizada para ambos os grupos.

Palavras-chave: Escala de letramento estatístico Taxonomia. SOLO Estágios de conhecimento de contexto (*Tiers*). Modelo de Rasch.

Abstract

The study was aimed at testing the validity of a statistics literacy test composed of six questions, applied to 236 college students from six technological programs who had recently completed a Statistics course. Answers were analyzed using categories presented by Watson and Callingham, based on interaction of SOLO Taxonomy and Hierarchy of statistical literacy (*Tiers*). Answers were classified using the item map of the Rasch model for partial credits. Psychometric properties of the questions provide evidence of the validity of the test to assess the “statistic literacy” construct, presenting satisfactory internal consistency and adequate *infit* and *outfit* values. The order of difficulty of the answer categories was also compared to that reported in the Watson and Callingham study. Results suggest that even in a different context of application (statistics background, age), the statistics literacy scale can be used for both groups.

Keywords: Statistics literacy scale. SOLO Taxonomy. Hierarchy of statistical literacy (*Tiers*). Rasch model.

1 Introdução

Para Soares (2005), o letramento diz respeito à utilização e à apropriação social que a pessoa faz da leitura e da escrita em seu contexto social. Corresponde à capacidade de um grupo social ou de uma pessoa compreender a utilidade prática do domínio da leitura e da escrita para tomar decisões com eficácia na resolução de situações cotidianas, na vida familiar, em sua comunidade e ambiente de trabalho. A pessoa letrada transforma seu estado cognitivo, social, cultural, linguístico e econômico, assim como transforma seu modo de avaliar os fatos e seu entorno. Além disso, uma pessoa letrada desenvolve modos de pensar e de falar diferentes, que associados à ampliação do vocabulário podem levar a uma mudança significativa na vida em sociedade.

De modo mais específico, o termo letramento estatístico foi definido pela primeira vez por Wallman (1993, p.1) como sendo a “[...] competência

para compreender e avaliar criticamente resultados estatísticos que permeiam nossas vidas diárias junto à capacidade para reconhecer a contribuição que o pensamento estatístico pode trazer para as decisões públicas e privadas, profissionais e pessoais.”

Essa competência estatística, como apoio ao exercício da cidadania, também está presente na definição de letramento estatístico de Gal (2002), que propõe um modelo pensando em leitores ou *consumidores de informações*, isto é, quando as pessoas assistem à televisão, leem jornal, analisam anúncios para comprar, examinam um debate político ou uma propaganda eleitoral para fazer suas escolhas baseadas em informações estatísticas.

O modelo de Gal (2002) é composto de dois componentes: o cognitivo e o de disposição, representados por cinco e dois elementos, respectivamente. Os cinco elementos do componente cognitivo são: o letramento; o conhecimento estatístico; o conhecimento matemático; o conhecimento de contexto e a competência de elaborar questões críticas. Os dois elementos do componente de disposição são descritos como atitudes e crenças que moldam a visão de mundo das pessoas e postura crítica, que nada mais é do que a propensão para um comportamento questionador diante de informações estatísticas (GAL, 2002).

Mais especificamente, sobre o conhecimento de conceitos e procedimentos básicos de Estatística e Probabilidade Gal (2002) explica que esse aspecto não pode ser discutido em termos absolutos, mas depende do nível de letramento estatístico esperado pelos cidadãos. Gal (2002) propõe cinco tópicos que devem ser abordados: a) conhecimento dos motivos e das maneiras pelas quais a coleta de dados aconteceu; b) familiaridade com termos e ideias básicas relacionadas à Estatística Descritiva; c) familiaridade com termos e ideias básicas relacionadas às apresentações gráficas e tabulares; d) compreensão das noções básicas de Probabilidade; e) conhecimento de como as conclusões e inferências estatísticas são obtidas.

Para formar adultos estatisticamente letrados, Watson (2003) diz que é necessário que os alunos, ainda na escola, possam vivenciar alguns conceitos estatísticos e probabilísticos. No caso da Austrália, por exemplo, a partir da inserção de conteúdos estatísticos na educação básica (AUSTRALIAN EDUCATION COUNCIL [AEC], 1991), muitas ações foram desenvolvidas para promover e avaliar o letramento estatístico dos alunos. Dentre essas ações, destaca-se a pesquisa longitudinal (de 1993 a 2000) de Watson e Callingham (2003), que construíram e validaram um instrumento com 80 questões para definir o nível de letramento estatístico das categorias de respostas (itens) dos alunos, culminando com o desenvolvimento da escala de letramento estatístico - *scale*

of statistical literacy. Essa pesquisa foi aplicada nos anos de 1993, 1995, 1997 e 2000¹, perfazendo um total de 3.852 alunos da escola básica na Austrália, com faixa etária correspondente aos alunos do 3º e do 5º ao 9º anos do ensino brasileiro.

Considerando que as questões desse instrumento remetem ao uso social da Estatística, seis delas foram aplicadas a estudantes de seis cursos tecnológicos de uma universidade particular de São Paulo, como parte do projeto de pesquisa que tem como objetivo verificar os processos de autorregulação da aprendizagem de Estatística desses estudantes e sua relação com o nível de letramento estatístico.

Nesse contexto, esse artigo tem como objetivos: avaliar a consistência interna e a validade do construto letramento estatístico de uma prova de Estatística elaborada com seis questões do instrumento de Watson e Callingham (2003) para estudantes de cursos tecnológicos de São Paulo e comparar os índices de dificuldade das categorias de respostas (itens) obtidos nesse estudo com a pesquisa da Austrália, por meio da Teoria de Resposta ao Item (TRI).

2 Referencial teórico

Para atingir aos objetivos propostos nesse trabalho, o processo de validação do instrumento foi semelhante ao utilizado por Watson e Callingham (2003), que teve início com a categorização das respostas dos alunos por meio da utilização da taxonomia SOLO (BIGGS; COLLIS, 1991), associada aos estágios do conhecimento de contexto² (WATSON, 1997). A partir da classificação das respostas, foram determinados os níveis de letramento estatístico por meio de uma análise qualitativa, mas apoiada nos resultados obtidos com o uso da Teoria de Resposta ao Item – TRI, mais especificamente, pelo modelo de Rasch (RASCH, 1980).

2.1 Taxonomia SOLO

O modelo hierárquico SOLO - *Structure of Observed Learning Outcomes* é utilizado para categorizar as respostas do aluno a um conjunto de

¹ Para cada ano escolar foram selecionadas 24 questões e aplicadas repetidamente nos anos de 1993, 1995 e 1997, sendo que 8 dessas questões eram comuns a todos os anos escolares, e 16, a pelo menos dois anos escolares. Já no ano de 2000, foram 20 questões comuns a todos os anos escolares, e 4, a pelo menos dois anos escolares (essas 4 questões foram aplicadas também nos anos de 1993, 1995 e 1997).

² O termo *hierarchy of statistical literacy*, definido por Watson (1997), foi traduzido pelas autoras desse trabalho como estágio de conhecimento de contexto. Não foi feita uma tradução literal para que o termo não fosse confundido com níveis de letramento estatístico.

tarefas pré-estabelecidas, de acordo com a complexidade estrutural e o número de conceitos exigidos nas mesmas. No processo, avalia-se o desempenho do aluno para uma determinada tarefa, em vez de avaliar a estrutura cognitiva desse indivíduo (BIGGS; COLLIS, 1982). Isto é, o fundamento básico desse modelo é a suposição de que o entendimento cognitivo do aluno não é o mesmo para toda tarefa, pois depende do conteúdo e do contexto envolvidos; sendo assim, a análise não foca a estrutura cognitiva desse indivíduo e, sim, o entendimento específico de uma tarefa, o que difere do proposto por Piaget (PANIZZON, PEGG; MCGEE, 2004).

Segundo Panizzon, Pegg e Mcgee (2004), o SOLO tem duas características importantes: a primeira foca na natureza ou na abstração das respostas e é referida como modo de pensamento, que é o tipo de funcionamento intelectual requerido para direcionar adequadamente um estímulo particular. A segunda característica depende de uma competência individual para manipular, com sofisticação crescente, exemplos relevantes. Esta característica refere-se aos níveis de resposta que ocorrem dentro dos modos de aprendizagem, fornecendo uma descrição hierárquica da natureza da estrutura das respostas.

Biggs e Collis (1991) definiram cinco modos de pensamento: sensomotor, icônico, concreto simbólico, formal e pós-formal. O sensomotor é o modo em que habilidades motoras são requeridas, no caso, das crianças. Na vida adulta, este modo está associado com habilidades requeridas em vários esportes. O icônico é o modo em que a pessoa é capaz de internalizar ações em forma de imagens que representam objetos e eventos; para o adulto, este modo de pensamento é requerido na apreciação de arte, música e permite a forma de conhecimento referida como intuitiva (PANIZZON, PEGG; MCGEE, 2004). No modo concreto simbólico, a pessoa utiliza um elemento simbólico na aprendizagem, tal como linguagem escrita ou sistema de números; o pensamento, neste modo, depende de uma referência do mundo real. No modo formal, a pessoa é capaz de considerar conceitos que são abstratos, não está restrita à referência de um mundo real e trabalha com princípios e teorias. Finalmente, no modo pós-formal, a pessoa é capaz de questionar ou mudar a estrutura fundamental de teorias ou disciplinas (BIGGS; COLLIS, 1991).

Quanto aos níveis de resposta, Biggs e Collis (1991) definiram cinco estágios:

- Pré-estrutural: O aluno comete erros de entendimento fundamental, apresentando respostas não significativas. Este nível indica a transição do modo de pensamento anterior para o atual.
- Uniestrutural: O aluno apresenta diversas conclusões, que podem ser

corretas, mas que não são coerentes entre si, foca apenas um aspecto relevante.

- **Multiestrutural:** o aluno mostra entendimento, discute o conteúdo de forma significativa, tem conhecimento de uma quantidade razoável de conteúdos com várias respostas relevantes, mas não coordenadas, podendo apresentar algumas inconsistências.
- **Relacional:** o aluno integra as partes dentro de um todo de forma coerente, isto é, apresenta uma conclusão capaz de relacionar todos os aspectos relevantes, evidenciando uma coerência global. Contudo, a conclusão final pode não servir para todos os contextos.
- **Abstrato estendido:** entendimento metacognitivo, o aluno generaliza além das informações fornecidas, utiliza-se de diversos princípios para formar um novo conjunto de assuntos. Este nível indica a transição para o modo de pensamento subsequente.

De acordo com Watson (2006), a maior parte da aprendizagem que ocorre nos anos escolares requer o modo concreto simbólico. Segundo essa mesma autora, pelo tempo de permanência dos alunos na escola, e a complexidade dos conceitos no modo concreto simbólico, é provável que a aprendizagem ocorra nesse modo de pensamento em dois ou até mais ciclos: uniestrutural, multiestrutural, relacional, dependendo de como as ideias são construídas, consolidadas e, então, aplicadas em outros contextos. Por exemplo, no primeiro ciclo, o aluno pode pensar na definição de média como *normal, total ou um algoritmo*, e, então, *um algoritmo que resulta num valor que representa um conjunto de dados*. Uma vez que esse conceito é consolidado como ferramenta, pode ser usado como elemento no contexto mais complexo; talvez na tomada de decisões sobre as diferenças entre dois ou mais conjuntos de dados; com isso, torna-se parte de um outro ciclo, mas ainda dentro do modo concreto simbólico. (WATSON, 2006, p.15).

2.2 Estágios do conhecimento de contexto

O outro modelo utilizado por Watson e Callingham (2003), na categorização das respostas dos alunos, foi o dos estágios do conhecimento de contexto proposto por Watson (1997):

As competências requeridas para interpretar as informações estatísticas apresentadas à sociedade pela mídia podem ser representadas em três estágios hierárquicos: (a) conhecimento básico das terminologias de Probabilidade e

Estatística, (b) entendimento da linguagem e conceitos de Probabilidade e Estatística quando as informações estão inseridas num contexto mais amplo de discussão social (c) atitude questionadora, utiliza-se de conceitos mais sofisticados para contradizer afirmações sem base Estatística apropriada. Estas competências representam um pensamento mais sofisticado e são consistentes com os modelos de aprendizagem da psicologia do desenvolvimento, como por exemplo a taxonomia SOLO de Biggs e Collis (1982, 1991). (WATSON, 1997, p.1).

De acordo com Watson (1997), os três estágios hierárquicos apresentam como principais características:

- Estágio 1 - Entendendo a terminologia básica. As competências estão relacionadas diretamente com tópicos do currículo; são geralmente ensinadas de modo convencional, incluindo porcentagem, mediana, média, probabilidades, gráficos, medidas de dispersão etc. Podem ser ensinadas sem fazer referência aos assuntos sociais.
- Estágio 2 – Incorporando a linguagem e os conceitos num contexto mais amplo. Depois que os alunos trabalham com alguns conceitos estatísticos básicos e são expostos às informações transmitidas pela mídia, torna-se importante saber ler e interpretar os relatórios escritos, retirar conclusões e tomar decisões, ao invés de apenas executar cálculos.
- Estágio 3 – Questionando as afirmações. Nesse estágio, os alunos já têm a confiança necessária para questionar o que leem na mídia. Algumas vezes, as informações são transmitidas sem uma base estatística adequada, seja por descuido ou por intencionalidade, os alunos devem estar preparados para questionar constantemente as conclusões obtidas, utilizando-se dessas informações. Esse processo de questionamento do contexto está coerente com o componente de disposição de letramento estatístico definido por Gal (2002).

2.3 Níveis de letramento estatístico

Watson e Callingham (2003) identificaram que, para categorizar as respostas dos alunos ao instrumento de letramento estatístico, a taxonomia SOLO não era suficiente, uma vez que muitas questões envolviam, além dos aspectos cognitivos, o entendimento do contexto. Essa constatação levou à necessidade

do uso conjunto com os estágios do conhecimento de contexto. É, portanto, a interação da complexidade estrutural e apropriação estatística em termos dos estágios que determinam a classificação hierárquica das respostas (categorias).

Um exemplo dessa relação é apresentado por Watson (2004) quando analisa quatro tarefas (WATSON; MORITZ, 2000) envolvendo o conteúdo de amostragem. Na primeira tarefa, que corresponde ao estágio 1 do conhecimento de contexto, solicita-se apenas a definição do termo *amostra*, sendo que os níveis de resposta podem ir do pré-estrutural (exemplo, *algo muito quente*) ao relacional (*pequena parte representando o todo*). A segunda tarefa está relacionada com a compra de um carro (Toyota ou Honda) e enquadra-se no estágio 2 do conhecimento do contexto, em que os níveis de resposta podem variar de uniestrutural (*comprar um Toyota baseado na experiência de três amigos*) a relacional (*comprar um Honda baseado na pesquisa com 800 pessoas*). Nas terceira e quarta tarefas, são apresentados os resultados, respectivamente, de uma pesquisa voluntária de rádio sobre a legalização da maconha e uma reportagem de um jornal com uma amostra não representativa (somente Chicago) do acesso de estudantes a armas nos Estados Unidos. Essas duas tarefas correspondem ao estágio 3 do conhecimento de contexto, e os níveis de resposta podem ir do pré-estrutural (*não deve haver armas; maconha não deve ser considerada um crime*) ao abstrato estendido (*Chicago não representa os Estados Unidos; é uma pesquisa voluntária somente com os ouvintes desse rádio*).

Na classificação hierárquica das respostas, Watson e Callingham (2003) utilizaram a análise de Rasch para estabelecer os logitos dos índices de dificuldade dos itens (cada categoria de resposta de uma questão está sendo denominada de item) e das habilidades dos alunos e, por conseguinte, para construir o mapa de itens (itens e alunos colocados em ordem crescente de logito numa mesma escala). Em seguida, analisaram qualitativamente a posição dos itens no mapa, estabelecendo seis agrupamentos (pontos de corte, em inglês *standard setting*) de acordo com os conteúdos, habilidades e contexto exigidos em cada tarefa. Esses agrupamentos foram denominados de níveis de letramento estatístico: idiossincrático, informal, inconsistente, consistente e não crítico, crítico e matematicamente crítico; definindo, assim, a escala de letramento estatístico (Figura 1).

Segundo Watson (2003), no primeiro nível, idiossincrático, as respostas sugerem um engajamento limitado com o contexto e com o uso da terminologia e conceitos estatísticos: a) resultados devidos ao acaso são compreendidos como favoritismo (de cores ou números); b) não há compreensão de amostra não

representativa; c) não há engajamento em tarefas com medidas de tendência central; d) a leitura das tabelas restringe-se a valores isolados.

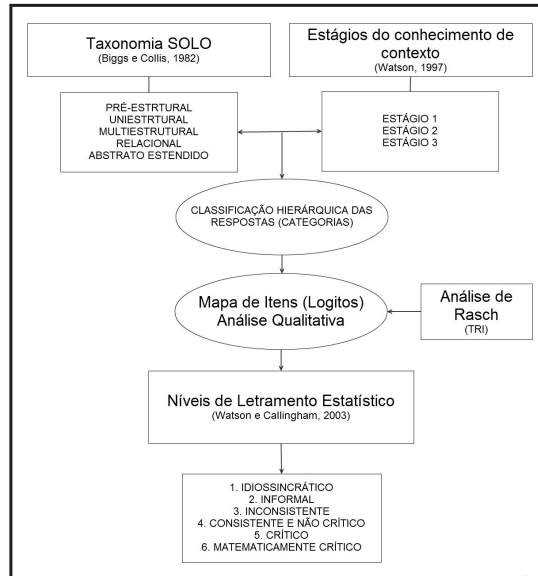


Figura 1 - Esquema para a determinação dos níveis de Letramento Estatístico da Watson e Callingham (2003).

No nível informal, as respostas sugerem um engajamento simples ou informal com o contexto, refletindo crenças intuitivas e não estatísticas no uso de elementos simples para terminologias e definições complexas, como ocorre nas etapas de construção de gráficos e tabelas e nos cálculos de chance. Para a justificativa de predições ao acaso, provavelmente, as respostas são baseadas em: *alguma coisa pode acontecer*; e a variação em tentativas repetidas, pois é improvável de acontecer espontaneamente.

As respostas no nível inconsistente são mais engajadas com o contexto do que nos níveis anteriores, frequentemente dependem da maneira como a tarefa é apresentada. As conclusões são mais apropriadas, mas ainda sem justificativa consistente; são usadas ideias estatísticas mais qualitativas do que quantitativas. A compreensão do contexto começa a surgir no nível de entendimento consistente e não crítico, no entanto não se questionam afirmações de natureza suspeita. São utilizados nas respostas: os múltiplos aspectos da terminologia; o conceito de variação apenas no contexto de chance e as competências estatísticas em problemas envolvendo média, probabilidade simples e características gráficas.

A compreensão e análise do contexto aparecem no nível de entendimento crítico, desde que não requeiram um raciocínio matemático sofisticado. Nesse nível, é utilizada a terminologia de maneira adequada e as definições são apresentadas de forma integrada; por exemplo, amostra e variação. No nível mais avançado de letramento estatístico – entendimento matemático crítico (WATSON, 2003) – já se observa nas respostas a habilidade de questionar criticamente os resultados em qualquer contexto. Informações veiculadas na mídia a partir de amostra não representativa são identificadas, bem como é questionada a presença de valores discrepantes em um conjunto de dados pequeno; o que leva a eliminá-los no cálculo da média.

Existe uma relação entre os níveis de letramento estatístico e os estágios do conhecimento de contexto, por exemplo, as habilidades matemáticas e estatísticas que são expressas em diferentes níveis usam a terminologia de conceitos estatísticos que são essenciais no Estágio 1. Já a aplicação da terminologia na interpretação do contexto, objetivo do Estágio 2, aparece do nível inconsistente em diante, e pensar criticamente sobre informações e métodos inadequados, objetivo do Estágio 3, aparece do nível crítico em diante (WATSON; CALLINGHAM, 2003; WATSON, 2004).

Vale salientar que Watson e Callingham (2003) também utilizaram a análise de Rasch para validar esse construto.

2.4 Validade do construto *letramento estatístico*

A validade pode ser compreendida como uma verificação direta da possibilidade do instrumento satisfazer o seu objetivo (PASQUALI, 2003). São diversas as técnicas empregadas para a determinação do índice de validade de um instrumento, porém, este estudo terá como base a evidência de validade de construto. Segundo Dias e Vendramini (2008), a validade de construto de um instrumento é a extensão em que se pode dizer que o instrumento mede um construto teórico ou um traço.

Para verificar se existem evidências de validade de um instrumento, uma das técnicas de estatística multivariada que pode ser utilizada é a Teoria de Resposta ao Item (TRI), que se baseia em modelos matemáticos nos quais as estimativas dos traços dependem das respostas dos sujeitos e das propriedades dos itens avaliados em uma mesma escala métrica (EMBRETSON; REISE, 2000). Uma das vantagens dessa técnica é a possibilidade de equalização de traços de indivíduos, pertencentes ou não à mesma população, e submetidos a diferentes instrumentos de medida; o que permite a comparação de seus desempenhos.

Um dos modelos da TRI é o de créditos parciais de Rasch (RASCH, 1980; MASTERS, 1982). Esse modelo considera o escalonamento hierárquico das categorias de resposta dos itens e a interação entre pessoas e itens para estimar a probabilidade de cada pessoa responder a cada item. Utiliza uma mesma escala de medida para representar o traço medido da pessoa e a dificuldade do item, produzindo um conjunto de escores que definem a posição de cada item e de cada pessoa nessa escala de medida. A unidade de medida é o logito, isto é, o logaritmo da chance de sucesso $p/(1-p)$.

3 Método

Participaram dessa etapa do projeto 236 alunos (54,3% do gênero masculino) de graduação de seis cursos tecnológicos: Logística Empresarial, Gestão Bancária, Segurança Empresarial, Gestão de Seguros e Previdência, Gestão de Negócios e Finanças e Gestão de Pessoas, de uma universidade particular instalada na Grande São Paulo, que concluíram uma disciplina de Estatística em 2007. A coleta de dados aconteceu no 1º semestre de 2008. A média de idade dos participantes era de 28,50 anos e o desvio padrão igual a 7,21 anos.

A prova de letramento estatístico continha oito questões, sendo que nesse estudo estão sendo avaliadas apenas seis questões que foram selecionadas, traduzidas e adaptadas do instrumento da Watson e Callingham (2003). Nessas questões, foram abordados os seguintes conteúdos: medidas de tendência central, probabilidade, variabilidade e leitura e interpretação de tabela de dupla entrada. As respostas foram analisadas de acordo com as categorias apresentadas pelas pesquisadoras australianas, ressaltando, mais uma vez, que as mesmas utilizaram conjuntamente a taxonomia SOLO e os estágios de conhecimento do contexto.

Após a etapa de classificação das respostas, foi utilizado o modelo de créditos parciais de Rasch (RASCH, 1980; MASTERS, 1982) com o objetivo de avaliar a consistência da prova e testar a validade do construto *letramento estatístico*.

Para avaliar a consistência da prova, analisou-se, inicialmente, a qualidade das questões que compõem a prova para saber se todas devem fazer parte da prova ou se algumas devem ser excluídas por ter propriedades psicométricas ruins, que não se ajustem ao modelo de Rasch. Para essa análise, foram utilizados os índices de dificuldade, medidas de ajuste ao modelo: *Infit* e

*Outfit*³ e a correlação entre a questão e a medida de Rasch.

Para testar a validade do construto *letramento estatístico* foram investigados, também, os resultados da análise dos componentes principais, baseada no resíduo do modelo de Rasch da prova. Para um bom ajuste do modelo de Rasch, espera-se encontrar uma variância explicada pelo modelo acima de 60% e variâncias, não explicadas pelos contrastes, inferiores a 5%. Nessa análise de resíduos, o que se deseja é encontrar o número mínimo de contrastes que expliquem o máximo de variância possível.

Para avaliar o comportamento das questões da prova, comparou-se a ordem de dificuldade das categorias de respostas (itens) obtidas no estudo de Watson e Callingham (2003) com os determinados neste trabalho. Grandes alterações na ordem de dificuldade dos itens podem indicar que a escala de letramento estatístico não está adequada para ser utilizada entre populações com contextos diferentes.

Para todas as análises estatísticas realizadas neste trabalho foi utilizado o *software Winsteps® Rasch Measurement* (LINACRE, 2009).

4 Teste estatístico

A seguir, são apresentadas a descrição das questões do teste e as categorias utilizadas para classificar as respostas dos alunos, baseado no estudo de Watson e Callingham (2003).

4.1 Descrição das questões

Na primeira questão (Q.1), foi apresentado o resultado de uma pesquisa, que exigia do aluno a interpretação do conceito de média, bem como a inversão do seu algoritmo (Figura 2).

Para obter o número médio de crianças por família em uma cidade, um professor contou o número total de crianças na cidade. Então, ele dividiu por 50, que é o número total de famílias. O número médio de crianças por família foi 2,2. Em relação a esta situação, assinale qual dessas afirmativas é verdadeira:

- a) Metade das famílias da cidade tem mais do que duas crianças.

³ O índice *infit* informa sobre discrepâncias do dado teórico esperado e empírico em regiões centrais da Curva Característica do Item (CCI), indica padrões de respostas inesperados para pessoas com níveis de habilidades próximos à dificuldade do item. O índice *outfit* informa sobre esse tipo de diferença nas extremidades da CCI, acusa a presença de padrões inesperados de respostas de pessoas com níveis de habilidade muito diferentes da dificuldade do item.

- b) Mais famílias na cidade têm 3 crianças do que tem 2 crianças.
- c) Há um total de 110 crianças na cidade.
- d) Na cidade, há 2,2 crianças para cada adulto.
- e) Número mais comum de crianças na família é 2.
- f) Nenhuma das afirmativas é correta.

Figura 2 - Primeira questão do teste estatístico

A segunda questão (Q.2) abordava o cálculo de medidas de tendência central com a presença de um valor discrepante. O contexto apresentado para o aluno trazia pesos (em gramas) de um mesmo objeto medido por 9 alunos com uma mesma balança. Desses 9 alunos, 8 encontraram peso variando entre 6 e 6,3 e apenas 1 aluno determinou um peso de 15,3. Então, os alunos tinham que responder as seguintes questões: *qual medida de tendência central você usaria para representar estas medidas? Por que você escolheu esta medida de tendência central? Calcule a medida que você escolheu.*

A terceira questão (Q.3) envolve um problema de probabilidade simples, com eventos não equiprováveis, em que o aluno precisava escolher uma interpretação mais apropriada para a situação apresentada (Figura 3).

Um vidro de remédio tem o seguinte aviso: *Para aplicações na pele, há uma chance de 15% de aparecerem erupções. Se aparecerem as erupções, procure seu médico.* O que significa esta informação?

- a) Não use este medicamento em sua pele – Há uma boa chance de aparecer erupções.
- b) Para aplicações na pele, aplique somente 15% da dose recomendada.
- c) Se você tiver erupções, provavelmente envolverá somente 15% de sua pele.
- d) Aproximadamente 15 dentre 100 pessoas que usaram este remédio tiveram erupções.
- e) Há uma pequena chance de aparecerem erupções com o uso deste medicamento.

Figura 3 - Terceira questão do teste estatístico

Na quarta questão (Q.4), o aluno analisava a seguinte situação: *imagine que você lançou um dado 60 vezes. Na tabela abaixo, preencha o número de vezes que você acha que cada número do dado vai aparecer.* Para responder a questão, o aluno deveria levar em consideração a variabilidade inerente à experimentação aleatória. Segundo Batanero (2001), os resultados da experimentação aleatória devem ser usados com cautela para que não ocorra a extensão indevida da *Lei dos grandes números*; nesse estudo credita-se à existência de uma *Lei de pequenos números*, justamente devido à sensibilidade do tamanho da amostra. A quinta questão (Q.5), assim como a Q.3 abordava

problema de probabilidade simples, com eventos não equiprováveis, solicitando explicitamente o evento com maior probabilidade.

Uma sala de aula de Economia tem 13 alunos casados e 16 solteiros. O nome de cada aluno foi escrito num pedaço de papel e colocados num estojo. O professor misturou todos os papéis e, sem olhar, pegou um papel. Considerando que todos os papéis tenham o mesmo tamanho, é mais provável que o professor tenha selecionado que tipo de aluno? Explique sua resposta

a) um aluno casado ou b) um aluno solteiro ou

c) são ambos, aluno casado e aluno solteiro, igualmente prováveis de serem selecionados

Figura 4 - Quinta questão do teste estatístico

A sexta questão tinha como objetivo explorar a leitura e interpretação de uma tabela de dupla entrada, condição importante para a compreensão de diferentes espaços amostrais e, por conseguinte, a probabilidade condicional. Foi apresentada uma tabela de dupla entrada com informações sobre uma academia de ginástica com o tipo de esporte praticado (alongamento, futebol, tênis e natação) de acordo com o gênero (aluna, aluno). Foram feitas cinco perguntas que foram analisadas separadamente: a) quantas alunas jogam tênis? b) qual é o esporte mais popular para as alunas? c) qual é o esporte mais popular para os alunos? d) considerando que esta academia só trabalha com estes esportes, quantas alunas esta academia tem? e) considerando que esta academia só trabalha com esses esportes, qual o número total de alunos desta academia? Para efeito das análises, essas cinco perguntas foram consideradas, em ordem, como sendo as questões Q.6, Q.7, Q.8, Q. 9 e Q.10.

4.2 As categorias

A quantidade de categorias utilizadas para classificar as respostas dos alunos variou de acordo com a questão, por exemplo, para a quarta questão foram necessárias cinco categorias, já para a sexta questão, apenas duas (Tabela 1).

Tabela 1- Descrição das categorias por questão baseada em Watson e Callingham (2003)

Q.	Descrição da Categoria
1	0. Resposta em branco. 1. Escolhe a alternativa <i>a</i> – interpretação incorreta do conceito de média ou <i>b</i> – aproximação para cima do valor da média. 2. Escolhe a alternativa <i>d</i> – interpretação informal do conceito de média ou <i>e</i> – aproximação para baixo do valor da média ou <i>f</i> – considera que as respostas não são totalmente apropriadas. 3. Escolhe a alternativa <i>c</i> – interpreta corretamente o conceito e usa o algoritmo de forma inversa.
2	0. Resposta em branco ou medida inadequada. 1. Escolhe uma das 3 medidas (moda, mediana ou média), mas com cálculo ou justificativa errada ou sem resposta do valor calculado, ou ainda o valor correto de uma das medidas, mas com justificativa inconsistente. 2. Determina o valor correto da moda ou da mediana, apresenta a definição ou conceito, mas não utiliza o nome da medida; ou calcula a média, mas sem retirar o valor de 15,3. 3. Remove o valor de 15,3 e calcula a média ou determina o valor correto do da moda ou mediana, apresenta a definição ou conceito e utiliza o nome da medida.
3	0. Resposta em branco ou incorreta - alternativa <i>b</i> ou <i>c</i> . 1. Respostas considerando apenas a interpretação verbal de chance – alternativa <i>a</i> ou <i>e</i> . 0. Respostas considerando só uma interpretação numérica para os 15% de chance ou a interpretação numérica e verbal – somente a alternativa <i>d</i> ou alternativas <i>d e a</i> ou <i>d e e</i> .
4	0. Resposta em branco ou com soma diferente de 60 ou com algum valor maior que 21. 1. Soma igual a 60 mais com valores altos e/ou baixos (exemplo: 10, 10, 20, 0, 10, 10). 2. Respostas envolvendo o conceito da probabilidade clássica (10, 10,10, 10, 10, 10) ou múltiplos de 5 (5, 10, 15, 5, 10, 15). 3. Respostas refletindo variação, mas com uma diferença entre os valores de no máximo 1 unidade. 4. Respostas refletindo variação, mas com uma diferença entre os valores (exemplo, de 2 unidades).
5	0. Resposta em branco ou não soube responder ou com valores incorretos. 1. Escolhe alternativa <i>a</i> ou <i>b</i> ou <i>c</i> , mas justifica usando termos como sorte, deve ser qualquer coisa ou o professor é de certo gênero e deu preferência na escolha. 2. Escolhe alternativa <i>c</i> , justifica com termos como sorteio, mesma chance ou diferença pequena. 3. Escolhe a alternativa <i>b</i> , justifica com termos como 13 versus 16 ou mais sorteios. 4. Escolhe a alternativa <i>b</i> , justifica com termo como probabilidade maior ou apresenta os cálculos
6*	0. Resposta em branco ou não soube responder. 1. Resposta com valor correto.
8	0. Resposta em branco ou não soube responder. 1. Resposta indicando apenas um dos esportes – Futebol ou Natação. 2. Resposta indicando dois esportes - Futebol e Natação.
10	0. Resposta em branco ou não soube responder. 1. Resposta com algum tipo de erro de cálculo, mas dentro de um intervalo aceitável. 2. Resposta correta.

Ressalta-se que as respostas das questões 7 e 9 foram categorizadas da mesma forma que a questão 6.

5 Resultados e discussão

Com as respostas categorizadas, foi utilizado o modelo de créditos parciais de Rasch para a avaliação dos dados. Em uma primeira análise, a questão 4 foi excluída por apresentar índice de *Infit* igual a 1,57, que está fora dos limites aceitáveis (entre 0,5 e 1,5). Após a eliminação dessa questão, os índices de dificuldade e de ajuste foram recalculados. Uma segunda extração, da questão

10, foi necessária por apresentar índices de *Outfit* fora do limite estabelecido como aceitável com valor de 1,70.

Refeitas as análises, os resultados indicaram que os índices de dificuldade das 8 questões restantes variaram de -1,55 a 0,75 ($M = 0,00$; $DP = 1,12$) e os índices de desajuste ao modelo de Rasch *Infit* e/ou *Outfit* ficaram dentro dos limites aceitáveis (Tabela 2). A exceção foi apenas para a questão 2, que apresentou índice de dificuldade de 2,46 e *Outfit* igual a 1,79. Mesmo com esses índices, essa questão não foi retirada por ser considerada importante para as pesquisas na área, uma vez que a mesma envolve as medidas de tendência central.

Tabela 2 - Propriedades psicométricas das questões do teste estatístico por ordem decrescente de índice de dificuldade

Questão	Índice de dificuldade	<i>Infit</i>	<i>Outfit</i>
1	-0,37	1,52	1,35
2	2,46	1,13	1,79
3	-0,16	0,83	0,74
5	0,75	0,94	0,91
6	-1,55	0,65	0,49
7	-0,79	0,83	0,87
8	-0,37	0,93	1,01
9	0,03	1,01	1,11

Os índices de confiabilidade de separação das questões e das pessoas foram considerados altos, indicando que as repostas dadas pelos alunos às questões da prova realmente refletem a amplitude de dificuldade ao longo da variável que permitirá descrever o construto *letramento estatístico*; além disso, que as questões do teste são apropriadas à amostra de estudantes que participaram da pesquisa (Tabela 3).

Tabela 3 - Índices de confiabilidade e medidas de ajuste dos 8 itens da prova

Índice	Valores
Confiabilidade dos pontos de separação das questões	0,98
Confiabilidade dos pontos de separação das pessoas	0,79
Média quadrática do <i>INFIT</i> das questões	0,98 (DP=0,24)
Média quadrática do <i>INFIT</i> das pessoas	1,03 (DP=0,67)
Alfa de Cronbach	0,82

Ainda, tomando como base os resultados da Tabela 3, observa-se que a média quadrática do *Infit*, tanto das pessoas como das questões, ficaram próximos de 1, o que, de acordo com Linacre (2009), indica que não existe dependência dos dados e tampouco presença de *outliers*. Já o Alfa de Cronbach é alto, o

que, segundo esse mesmo autor, mostra que existe uma alta probabilidade de que pessoas (ou itens) estimadas com altas medidas, de fato têm mais altas medidas do que pessoas (ou itens) estimadas com baixas medidas. Esses resultados sugerem que as questões do teste formam uma escala unidimensional hierárquica, e que o instrumento apresenta uma boa consistência interna.

A variância explicada pelo modelo foi de 56,7%, podendo ser considerada aceitável, por ser um valor próximo do ideal que é acima de 60%. Apenas a variância não explicada pelo primeiro contraste teve autovalor um pouco acima de 2,0 (Tabela A, Anexo). Esses resultados, associados aos já apresentados, reforçam a constatação de que a escala é predominante unidimensional, e, por conseguinte, que as questões do teste estão medindo de fato o construto *letramento estatístico* no que se refere aos conteúdos abordados nesse estudo.

Foi possível analisar, também, os índices de dificuldade de cada item (Tabela 4). Como explicado anteriormente, cada categoria de resposta foi denominada de item e está sendo representada pelo número da questão seguida do valor da categoria, por exemplo, 1.2, que corresponde a questão 1 e resposta classificada como categoria 2.

Tabela 4 - Item por ordem de dificuldade (de acordo com o logito) e a porcentagem de respostas

Aumento da dificuldade			Aumento da dificuldade		
Item	Logit	% resposta	Item	Logit	% resposta
6.0	-2,37	22	2.3	3,53	3
1.0	-2,34	18	5.4	2,34	10
5.0	-2,28	22	2.2	1,56	21
3.0	-2,17	24	3.2	1,48	46
8.0	-2,17	25	1.3	1,44	42
7.0	-1,61	31	5.3	1,40	33
9.0	-0,91	42	8.2	1,21	58
2.0	-0,52	58	9.1	1,18	58
1.1	-0,27	11	7.1	1,13	69
5.1	0,02	21	6.1	1,07	78
1.2	0,47	28	2.1	0,94	19
3.1	0,48	30	8.1	0,78	17
5.2	0,65	15			

Verifica-se pelos resultados da Tabela 4 que o item com índice de dificuldade mais baixa (sem considerar a categoria 0) foi o 1.1, e o índice de maior dificuldade e menor porcentagem de respostas foi o 2.3.

Avaliando cada questão individualmente:

- Na primeira questão, 42% das respostas foram classificadas na

categoria 3. Esse percentual ainda pode ser considerado baixo, dado que era esperado que os alunos já compreendessem o algoritmo da média e soubessem aplicar a inversão da fórmula. Ressalta-se que não era necessário interpretar esse conceito para optar pela alternativa correta.

- Os itens da segunda questão apresentaram os maiores índices de dificuldade, sendo que 58% das respostas foram classificadas na categoria 0 e apenas 3% na categoria 3. Nessa questão, o aluno teria que interpretar e calcular uma das medidas de tendência central; perceber a presença de um valor discrepante e calcular a média nesse contexto. Esses resultados parecem indicar que a abordagem desses conteúdos no nível universitário está restrita à manipulação de algoritmos, sem ênfase sobre a leitura e a interpretação crítica, nem tampouco sobre o envolvimento de um contexto, que são características do nível de letramento estatístico, considerado como matematicamente crítico. Corroborando essa suposição, no estudo de Kataoka et al. (no prelo), apenas 4% dos 1343 alunos de 5 universidades brasileiras responderam corretamente essa mesma questão, salientando apenas que nesse estudo as respostas eram de múltipla escolha.
- Na terceira questão, que se referia à interpretação do conceito de chance, pode-se dizer que os resultados foram satisfatórios, uma vez que 46% das respostas foram classificadas na categoria 2 e 30% dos alunos na categoria 1, sendo que, para esta última categoria, o aluno pelo menos ponderou que não deveria usar o medicamento, pois havia chance de ocorrer erupções (alternativa a) ou considerou que 15% era uma pequena chance (alternativa e). Kataoka et al. (no prelo) identificaram, para essa mesma questão, 67% das respostas classificadas na categoria 2 e 17% na categoria 1. Esses resultados eram esperados, já que em tese, os alunos devem concluir o ensino médio com um conhecimento básico de Probabilidade.
- Na questão anterior, salientou-se que, em teoria, era esperado dos alunos um conhecimento básico em Probabilidade, justamente porque na quinta questão, que também se refere a esse tópico, apenas 33% das respostas foram classificadas na categoria 3 e 10% na categoria 4. Diferentemente da terceira questão, essa exigia não apenas a interpretação de uma sentença, mas envolvia, também, o cálculo de probabilidade simples para embasar a tomada de decisão.
- As questões de 6 a 9, como dito, referem-se à leitura e interpretação

de uma tabela de dupla entrada. Nesse caso, os alunos, de forma geral, não apresentaram dificuldades, quando se leva em consideração os percentuais de respostas nas maiores categorias, isto é, 78%, 69% e 58% na categoria 1, respectivamente, para a sexta, sétima e nona questão; na oitava questão, 58% das respostas foram classificadas na categoria 2.

Outra análise, realizada nesse estudo, foi a comparação da ordem de dificuldade dos itens obtidos no estudo de Watson e Callingham (2003) com a determinada neste trabalho, com exceção do item 1.3 que não consta do estudo das pesquisadoras australianas (Tabela 5).

Tabela 5 - Comparação entre Austrália e Brasil da ordem de dificuldade dos itens associados ao Nível de Letramento estatístico (NL) obtido por Watson e Callingham (2003)

	Austrália*	Brasil		Austrália	Brasil
	Aumento da dificuldade	7.1 (NL1)		6.1	Aumento da dificuldade
6.1(NL1)		1.1	2.3 (NL6)	5.4	
5.1 (NL1)		5.1	2.2 (NL4)	2.2	
3.1 (NL1)		3.1	2.1 (NL3)	2.1	
8.1 (NL2)		8.1	1.2 (NL3)	3.2	
5.2 (NL2)		7.1	3.2 (NL3)	5.3	
9.1 (NL2)		1.2	1.1 (NL3)	5.2	
8.2 (NL2)		9.1	5.3 (NL3)	8.2	

No trabalho de Watson e Callingham (2003) as questões tinham, respectivamente, as seguintes denominações: AV12; AVG1; CH11; HAT8; TBL1, TBL2, TBL3 e TBL4

Observa-se, pelos resultados da Tabela 5, que não há grandes diferenças, com exceção dos itens 1.1 e 1.2. Watson e Callingham identificaram, no estudo de 2003, que os limites de ajuste para os itens da questão 1 estavam fora do padrão aceitável, podendo ser essa uma explicação teórica para essa diferença na ordem de dificuldade dos itens. Fazendo também uma relação da ordem de dificuldade com o nível de letramento estatístico (NL) estabelecido no estudo da Austrália, observa-se que a maioria dos itens do Brasil está no mesmo nível ou pelo menos num nível acima ou abaixo. Por exemplo, no nível idiossincrático (NL1), os itens 3.1, 5.1 e 6.1 estão no mesmo patamar, já o item 7.1, no Brasil está no patamar do nível informal (NL2). Esses resultados parecem indicar que, mesmo num contexto de aplicação diferente, no Brasil, com universitários de cursos tecnológicos e, na Austrália, com estudantes do ensino fundamental II, a escala de letramento estatístico pode ser utilizada para ambos os grupos.

Resultados similares foram obtidos por Callingham (2008) num estudo

comparativo com estudantes de Hong Kong do ensino médio e com estudantes da Austrália do ensino fundamental II, em que foram utilizadas 15 questões comuns. Uma das conclusões desse estudo é que a escala de letramento estatístico forneceu medidas de validade e confiabilidade em ambos os contextos, indicando que, com um instrumento robusto disponível, trabalhos futuros podem focar contextos sociais e culturais diferentes na área de Educação Estatística.

6 Considerações finais

A prova, da forma que foi proposta inicialmente, não apresentava evidências de validade, mas após a exclusão de duas questões, a consistência interna pode ser considerada satisfatória e os valores de *infit* e *outfit* adequados. Entretanto, como a variância explicada ficou um pouco abaixo de 60%, novos estudos devem ser realizados, visando uma revisão das questões, a busca por outras evidências de validade, e, por conseguinte, o aperfeiçoamento dessa prova de Estatística que pode auxiliar na identificação dos níveis de letramento estatístico das categorias de respostas dos estudantes universitários de cursos tecnológicos.

No processo de aperfeiçoamento dessa prova, já estão sendo incorporadas outras questões do instrumento de Watson e Callingham (2003) com intuito de aplicar a um maior número de respondentes e de adequar a metodologia para que fique similar à utilizada pelas pesquisadoras australianas. Esse processo pode dar maior efetividade às comparações dos resultados do Brasil com os da Austrália, ou com os de Hong Kong (CALLINGHAM, 2008), ou ainda, com os de qualquer outro país que realize pesquisa semelhante.

Espera-se que os resultados desta pesquisa contribuam para o melhor entendimento das dificuldades de estudantes universitários na solução de problemas estatísticos, como, por exemplo, os identificados nesse trabalho, quais sejam: leitura e interpretação crítica de informações envolvendo medidas de tendência central e conceitos probabilísticos.

Referências

AUSTRALIAN EDUCATION COUNCIL. **A national statement on mathematics for Australian schools**. Carlton, Vic.: Author, 1991.

BATANERO, C. **Didáctica de la Estadística**. Granada: Grupo de Investigación en Educación Estadística (Dept. de Didáctica de la Matemática - Universidade Granada). 2001. Disponível em: <http://www.ugr.es/~batanero/publicaciones_index.htm>. Acesso em: 10 jun. 2010.

BIGGS, J.; COLLIS, K. **Evaluating the Quality of Learning: the SOLO Taxonomy**. New York: Academic Press, 1982.

BIGGS, J.; COLLIS, K. Multimodal learning and the quality of intelligent behaviour. In: ROWE, H. (Ed.). **Intelligence, Reconceptualization and Measurement**. New Jersey: Laurence Erlbaum Assoc. 1991. p. 57-76.

CALLINGHAM, R. Using Rasch Measurement to Identify Cross- cultural aspects of Statistical Literacy. In: AARE INTERNATIONAL EDUCATION RESEARCH CONFERENCE, 2008, Brisbane. Melbourne, **Proceedings...** Brisbane. Melbourne: Australian Association for Research in Education, 2008. p.1-10. (JEFFERY, P.L. (Ed.)). Disponível em: <<http://www.aare.edu.au/08pap/cal08481.pdf>> Acesso em: 20 abr. 2010.

DIAS, A. S.; VENDRAMINI, C. M. M. Análise Fatorial com informação completa de uma prova de compreensão em leitura em estatística. **Revista Semestral da Associação Brasileira de Psicologia Escolar e Educacional (ABRAPEE)**, São Paulo, SP, v. 12, n. 2, p. 357-367, jul./dez. 2008. Disponível em: <http://www.abrapee.psc.br/sumarios.htm> Acesso em: 20 abr. 2010.

EMBRETSON, S. E.; REISE, S. P. **Item response theory for psychologists**. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, 2000.

GAL, I. Adult's Statistical Literacy: Meanings, Components, Responsibilities. **International Statistical Review**, Malden, Massachusetts, v. 70, n. 1, p. 1-25, Apr. 2002. <http://www.stat.auckland.ac.nz/~iase/publications/isr/02.Gal.pdf>. Acesso em: 20 abr. 2010.

KATAOKA, V. Y.; SILVA, C. B.; VENDRAMINI, C.; CAZORLA, I. Using Rasch Partial Credit Model to analyses the response of Brazilian undergraduate students to a statistic questionnaire. In: CONFERENCE OF EUROPEAN RESEARCH IN MATHEMATICS EDUCATION, 7th, 2011, Rzeszów, Polonia. **Proceedings...** (no prelo).

LINACRE, J. M. **A user's guide to WINSTEPS® Rasch-Model computer program: Program Manual 3.68**, Beaverton, Oregon, 2009. Disponível em: <http://www.winsteps.com/a/winsteps.pdf>. Acesso em: 10 mar. 2010.

MASTERS, G. N. A. Rasch model for partial credit scoring. **Psychometrika**, Berlin: Springer, v.47, n. 2, p. 149-174, June 1982.

PANIZZON, D.; PEGG, J.; MCGEE, S. Incorporating Different Assessment tasks to gauge student understandings of planetary processes. In: AARE INTERNATIONAL EDUCATION RESEARCH CONFERENCE, 2004, Melbourne. **Proceedings...** Melbourne: Australian Association for Research in Education, 2004, p 1-18. (JEFFERY, P.L. (Ed.)), Disponível em: <<http://www.aare.edu.au/04pap/pan04654.pdf>>. Acesso em: 20 fev. 2010.

PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis: Vozes, 2003.

RASCH, G. **Probabilistic models for some intelligence and attainment tests**. Chicago: University of Chicago Press, 1980.

SOARES, M. **Letramento: um tema em três gêneros**. 2.ed. Belo Horizonte: Autêntica, 2005.

WALLMAN, K. K. Enhancing Statistical Literacy: Enriching our Society. **Journal of the American Statistical Association**, Alexandria, v. 88, n. 421, p. 1-8, Mar. 1993.

WATSON, J.M. Assessing statistical literacy using the media. In: GAL, I.; GARFIELD, J. B. (Eds.). **The assessment challenge in statistics education**. Berlim: International Statistical Institute, 1997. p. 107-121.

WATSON, J.M. Statistical Literacy at the School Level: What should students now and do? In: BULLETIN OF THE INTERNATIONAL STATISTICAL INSTITUTE (ISI), 54th, 2003, Berlim. **Proceedings...** Netherlands: International Statistical Institute, 2003. p. 1-4. Disponível em: <<http://www.stat.auckland.ac.nz/~iase/publications/3/3516.pdf>>. Acesso em: 15 fev 2010.

WATSON, J. M. Assessment in statistics education: Obstacle or opportunity? In: INTERNATIONAL CONGRESS ON MATHEMATICAL EDUCATION, 10th, 2004, Copenhagen, Denmark. **Proceedings...** Copenhagen, Denmark: International Commission on Mathematical Instruction, 2004. p 1-15. (NISS, M (Ed.)). Disponível em: <http://www.icme10.dk/proceedings/pages/regular_pdf/RL_Jane_Watson.pdf>. Acesso em: 20 fev 2010.

WATSON, J. M. **Statistical literacy at school: Growth and goals**. Mahwah, NJ: Lawrence Erlbaum, 2006.

WATSON, J.; CALLINGHAM, R.. Statistical Literacy: a complex hierarchical construct. **Statistical Education Research Journal**, Auckland, v. 2, n. 2, p. 3-46, Nov. 2003. Disponível em: <http://www.stat.auckland.ac.nz/~iase/serj/SERJ2%282%29_Watson_Callingham.pdf>. Acesso em: 15 fev. 2010.

WATSON, J. M.; MORITZ, J. B. Development of understanding of sampling for statistical literacy. **Journal of Mathematical Behavior**, Amsterdam: Elsevier, v.19, n.2, p. 109-13, 2000.

Submetido em Maio de 2010.
Aprovado em Novembro de 2010.

ANEXO

Tabela A. Variância residual padronizada da análise dos componentes principais

Variância	Autovalor	Empírico (%)	Modelado(%)
Variância bruta explicada pelas medidas	9,8	55,0	56,7
Variância bruta explicada pelas pessoas	4,8	26,8	27,7
Variância bruta explicada pelos itens	5,0	28,2	29,0
Variância bruta não explicada (total)	8,0	45,0	43,3
Variância não explicada - 1º contraste	2,1	11,7	26,0
Variância não explicada - 2º contraste	1,3	7,5	16,8
Variância não explicada - 3º contraste	1,2	6,8	15,0
Variância não explicada - 4º contraste	1,1	6,3	13,9
Variância não explicada - 5º contraste	1,0	5,5	12,2
Variância bruta total das observations	17,8	100%	100%

