

# Construção e validação de instrumentos: um desafio para a psicolinguística

Valéria Pinheiro Raymundo

PUCRS (Porto Alegre, Brasil)



**RESUMO** – O propósito deste artigo é apresentar uma pesquisa realizada por Raymundo, em 2006, que teve o objetivo de elaborar e validar um instrumento de avaliação do nível de consciência linguística do aprendiz adulto de inglês, falante nativo de português. Para efetivar o processo de validação completo, dez instrumentos foram elaborados e três tipos de validação foram realizados – conteúdo, critério e construto – seguindo um roteiro de procedimentos estipulado pela pesquisadora. O exame da validade desse tipo de instrumento mostrou que é possível construir um teste fidedigno, fundamentado em níveis de consciência linguística definidos como pré-consciente 1, pré-consciente 2, consciente e plenamente consciente a partir do julgamento, identificação, correção e/ou explicação de erros interlinguais e intralinguais.

**Palavras-chave:** validação de instrumento; nível de consciência linguística; inglês como L2.

**ABSTRACT** – The objective of this article is to present a research carried out by Raymundo, in 2006, with the purpose of elaborating and validating an instrument of linguistic consciousness level evaluation of adult learners of English, native speakers of Portuguese. Ten instruments were elaborated and a complete validity process was accomplished – content, construct and criterion-related – according to a stipulated script of procedures. The examination of the validity of this type of instrument showed that it is possible to construct a reliable test based on levels of linguistic consciousness defined as pre-consciousness 1, pre-consciousness 2, consciousness and full consciousness from the judgment, identification, correction and/or explanation of interlingual and intralingual errors.

**Keywords:** validation of instrument, level of linguistic awareness, English as L2.

## Introdução

Shohami (1994) comenta que testes podem contribuir para teorias e pesquisas da área de aquisição de segunda língua de diversas maneiras. Eles podem definir e identificar formas de medir habilidades, aprimorar a qualidade dos instrumentos de coleta de dados, além de identificar e testar hipóteses. No entanto, a autora lembra que essas contribuições são limitadas quando os testes não são validados empiricamente. Segundo ela, a inadequação dos testes em pesquisas e a falta de critérios de qualidade para instrumentos podem levar a más interpretações dos fenômenos ligados à aquisição de segunda língua e a consequentes conclusões erradas sobre programas e métodos de ensino de língua estrangeira.

Argumenta sobre isso que instrumentos usados como medida psicométrica em pesquisas de aquisição de segunda língua devem ser testados etnograficamente para se tornarem válidos. Um teste, para ser válido, deve também levar em consideração a primeira língua do respondente, já que a língua materna interfere no processo de aprendizagem de língua estrangeira. Não se pode tratar indivíduos de diferentes etnias da mesma maneira. Shohami (1994)

comenta ainda que é importante expandir o repertório de procedimentos para coletar dados de linguagem para além da abordagem tradicional de testes. Ela sugere que testes de julgamento poderiam trazer benefícios para o ensino/aprendizagem de língua estrangeira.

Em 2006, Raymundo realizou um estudo que teve como objetivo principal elaborar e validar um instrumento de avaliação do nível de consciência linguística do aprendiz adulto de inglês, falante nativo de português. O instrumento foi construído a partir da apresentação de frases certas e erradas, enfocando erros interlinguais e intralinguais<sup>1</sup> e formas afirmativas, negativas e interrogativas, de diferentes

<sup>1</sup> **Erros intralinguais:** erros que ocorrem dentro da própria língua estrangeira e que não são oriundos da interferência da língua materna. Exemplos desse tipo de erro são: *Where does she work?*, em vez de, *Where does she work?*, para expressar a seguinte ideia: “Onde ela trabalha?”; *I have studied yesterday*, em vez de, *I studied yesterday*, para expressar a seguinte ideia: “Eu estudei ontem.”

**Erros interlinguais:** erros oriundos da transferência de formas e significados da língua materna (português) para a língua estrangeira (inglês). Exemplos que ilustram erros ocasionados pela sobreposição dos dois sistemas linguísticos são: *She works in a hospital?*, em vez de, *Does she work in a hospital?*, para expressar a seguinte ideia: “Ela trabalha num hospital?”; *How long do you study here?*, em vez de, *How long have you studied here?*, para expressar a seguinte ideia: “Há quanto tempo você trabalha aqui?”.

tempos e aspectos verbais (*simple past, simple present, future, present continuous, past continuous, present perfect*).

## 1 Implicações da mensuração educacional

Segundo Vianna (1978), a experiência dos que se dedicam à mensuração educacional tornou possível estabelecer princípios que orientam a construção de instrumentos de medida do desempenho escolar. O primeiro princípio diz que a medida do desempenho escolar é fundamental para uma educação eficiente; o segundo afirma que os instrumentos de medida facilitam as observações que o professor faz do desempenho do aluno. O terceiro diz que todos os objetivos educacionais importantes podem ser mensurados.

Para Vianna (1978), o processo de medida não visa a obter apenas escores numéricos, mas a examinar as situações relativas à aprendizagem ocorridas através do ensino, às respostas dos indivíduos às situações apresentadas e à classificação dos indivíduos segundo suas respostas. As respostas dos examinandos constituem o elemento nuclear do processo de medida. É a partir delas que toda uma superestrutura estatística é constituída para possibilitar um julgamento do seu desempenho. A estatística, longe de ser apenas uma complicação matemática, tem se mostrado um instrumento extremamente útil na organização e na interpretação dos dados.

Não se pode, contudo, considerar as respostas como um valor em si, um produto final, mas como um elemento através do qual se pode inferir a qualidade do desempenho do estudante. Um teste educacional é, como cita Dryer (apud VIANNA, 1978), um processo que se inicia e termina por um julgamento humano. Por essa razão, exige esforço, tempo e muito pensamento criativo do construtor. Para minimizar a possibilidade de julgamentos subjetivos e transformar um teste num instrumento hábil, com objetivos de verificação e avaliação, é importante que possua dois requisitos básicos: a fidedignidade e a validade.

## 2 Validade de instrumentos

A validação é o processo de examinar a precisão de uma determinada predição ou inferência realizada a partir dos escores de um teste. Validar, mais do que a demonstração do valor de um instrumento de medida, é todo um processo de investigação. O processo de validação não se exaure, ao contrário, pressupõe continuidade e deve ser repetido inúmeras vezes para o mesmo instrumento. Valida-se não propriamente o teste, mas a interpretação dos dados decorrentes de um procedimento específico. A cada aplicação de um instrumento, pode corresponder, portanto, uma interpretação dos resultados. Um teste utilizado no concurso vestibular pode, por exemplo, ser usado para avaliar as deficiências do ensino médio.

A validade de um teste começa no momento em que se pensa em construí-lo e subsiste durante todo o processo de elaboração, aplicação, correção e interpretação dos resultados. A interpretação da validade de um teste também exige, por vezes, o cálculo de diferentes coeficientes. Existem três aspectos da validade que correspondem aos objetivos de um teste e se referem à validade de conteúdo, à validade de critério e à validade de construto.

A **validade de conteúdo** refere-se ao julgamento sobre o instrumento, ou seja, se ele realmente cobre os diferentes aspectos do seu objeto e não contém elementos que podem ser atribuídos a outros objetos. Ela não é determinada estatisticamente, ou seja, não é expressa por um coeficiente de correlação, mas resulta do julgamento de diferentes examinadores especialistas, que analisam a representatividade dos itens em relação às áreas de conteúdo e à relevância dos objetivos a medir.

O planejamento do teste tem grande influência na validade de conteúdo, pois é, nesse momento, que se organiza uma amostra representativa de conhecimentos, de processos cognitivos e de comportamentos. Nessa fase, os juízes devem relacionar os diversos itens do instrumento a fim de caracterizar o equilíbrio do teste, o universo dos conteúdos e objetivos do processo instrucional. O trabalho cooperativo para a construção dos instrumentos de medida é essencial para evitar julgamentos subjetivos.

É importante ressaltar que os itens ou as tarefas de um teste traduzem um determinado contexto educacional e curricular; por isso, a validade de conteúdo de um teste não é permanente. Ela pode alterar-se com o transcurso do tempo e em função de modificações na filosofia educacional e/ou na estrutura curricular que o instrumento procura refletir. Isso significa que um teste possui validade de conteúdo num determinado momento e em função de um contexto educacional específico.

Um teste geralmente informa apenas sobre aquilo que está medindo; no entanto, às vezes, é importante saber algo mais. A finalidade da **validade de critério** está em verificar se o instrumento é capaz de identificar os que são efetivamente melhores para uma determinada atividade. Vianna (1989) exemplifica esse tipo de validade, mencionando o caso específico da seleção de candidatos para a Universidade. A expectativa é a de que os de melhor desempenho no instrumento de seleção sejam igualmente os melhores na Universidade, se não houver a interveniência de fatores que alterem a natureza dessa associação.

A validade de critério é também chamada de preditiva ou concorrente e refere-se ao grau de correlação entre os escores de um teste e outras medidas do desempenho (critério) obtidas independentemente ou simultaneamente ao teste. Quando o instrumento e o critério são aplicados simultaneamente, fala-se de validade concorrente; quando o critério é avaliado no futuro, fala-se de validade preditiva.

A validade de critério é estimada estatisticamente e, se a correlação entre os escores do teste (X) e os escores da variável critério (Y) é alta, diz-se que o teste é válido para o fim a que se destina.

Não se pode deixar de comentar, no entanto, que a validade de critério (preditiva ou concorrente) de um teste está sempre relacionada a um determinado fator, que, entretanto, pode sofrer a influência de outros fatores que não estão associados à variável preditora (teste), podendo afetar a magnitude do coeficiente de validade. A ocorrência de baixas correlações nem sempre reflete, assim, falta de validade do preditor, mas indica a possibilidade de que o critério seja totalmente questionável. As notas dos alunos, por exemplo, nem sempre se revestem de características ideais para o processo de validação. Vianna (1989) explica que, em geral, as notas sofrem influência e refletem a diversidade de parâmetros de avaliação empregados habitualmente por diferentes professores em uma mesma área departamental. Isso, segundo ele, significa um total comprometimento da fidedignidade.

A validade de conteúdo e a validade de critério, atributos exigidos dos bons testes de escolaridade, não são suficientes para validar um instrumento, pois não se preocupam com a compreensão dos construtos<sup>2</sup> que os testes medem. Consequentemente, impõe-se uma nova abordagem para análise dos instrumentos de medida. A **validade de construto** possibilita determinar qual a característica educacional que explica a variância<sup>3</sup> do teste ou, então, qual o significado do teste.

A validade de construto refere-se à demonstração de que o instrumento realmente mede aquilo a que se propõe medir. As evidências necessárias para esse tipo de validação são obtidas fazendo-se uma série de estudos inter-relacionados, por meio de testes estatísticos, das construções teóricas sobre a relação entre as variáveis a serem medidas.

A constatação da validade de construto resulta do acúmulo, por diferentes meios, de várias provas, que precisam ser analisadas em todos os seus detalhes. Esse tipo de validação visa a detectar, entre outros aspectos, quais as variáveis com as quais os escores do teste se correlacionam, quais os tipos de itens que integram o teste, o grau de estabilidade dos escores sob condições variadas e o grau de homogeneidade<sup>4</sup> do teste, com vistas a ter elementos que possam esclarecer o significado do instrumento.

Vianna (1989) alerta, porém, que, ao analisar um teste, o interesse nem sempre se limita ao conhecimento do conteúdo dos itens, mas aprofunda-se e procura conhecer também o processo usado pelos examinados nas suas respostas aos itens. Assim sendo, qualquer processo de análise que identifique capacidades e habilidades pode, em princípio, esclarecer o significado do construto que o teste mede, ao indicar as variáveis que estão sendo medidas pelos itens do instrumento. Por essa razão, é preciso usar de cautela quando for empregado o processo de análise, pois, indiscutivelmente, diferentes pessoas poderão utilizar processos diversos, mas igualmente válidos, de resposta a um item, criando-se, desse modo, uma situação complexa que pode levar a falsas interpretações.

A validação de construto não se limita a validar um teste; o seu alcance é bem mais amplo, centrando-se o seu objetivo na validação da teoria em que se apoiou a construção do instrumento. Desse modo, o trabalho de validação de um construto é uma pesquisa científica empírica, porque, definidos os construtos que seriam responsáveis pelo desempenho no teste, o avaliador passa a formular hipóteses sobre a teoria de construtos e a testá-las empiricamente. Esse tipo de validade é estudado quando o investigador deseja entender melhor as questões cognitivas e psicológicas que estão sendo medidas pelo teste. No caso deste estudo, estamos buscando diagnosticar qual o nível de consciência linguística do aluno diante da correção de erros interlinguais e intralinguais e qual o grau de autoconfiança do aprendiz diante dos diferentes níveis de consciência linguística.

Os três tipos de validade são pertinentes a todos os tipos de testes e independentes apenas no nível conceitual, pois um estudo completo de um teste normalmente envolve informação de todos os tipos de validade. Além disso, o desenvolvimento de qualquer atividade científica depende da perfeição dos seus instrumentos de medida.

A determinação do grau de validade de um teste é um procedimento bastante complexo, pois depende da sua finalidade, da interpretação que se dá aos escores e do seu uso. Além disso, para que um teste seja válido, é importante que seja fidedigno, já que as medidas educacionais são geralmente usadas para a formulação de julgamentos que podem afetar a vida e o futuro do aprendiz.

### 3 Fidedignidade de instrumentos

A fidedignidade de um teste ou de qualquer outro instrumento de medida, que apresenta resultados consistentes daquilo que pretende medir, é condição necessária para a validade. É geralmente expressa por alguma forma de coeficiente. A fidedignidade de um teste, por exemplo, indica até que ponto as diferenças nos escores são decorrentes de variações na característica examinada e não de erros casuais. Ela também se refere à

<sup>2</sup> Construtos são traços, aptidões ou características supostamente existentes e abstraídos de uma variedade de comportamentos que tenham significado educacional (ou psicológico). Assim, fluência verbal, rendimento escolar, aptidão mecânica, inteligência, motivação, agressividade, entre outros, são construtos.

<sup>3</sup> Medida de dispersão dos escores em relação à média.

<sup>4</sup> A análise da homogeneidade do teste indica se o teste mede um único traço ou se, ao contrário, mede traços diversos. O grau de homogeneidade é obtido pela medida de consistência interna.

estabilidade dos resultados de um teste, ou seja, ao grau de consistência e precisão dos escores.

Operacionalmente, a fidedignidade pode ser definida como o coeficiente de correlação entre, pelo menos, duas medidas. Existem diferentes métodos para calculá-la. Entre eles, podemos citar o método do teste-reteste e o método das metades.

O primeiro método é também conhecido como coeficiente de estabilidade e refere-se à estabilidade do examinando. O índice é obtido através da correlação dos escores de um teste com os escores de uma segunda aplicação deste aos mesmos sujeitos. Nesse caso, a variância do erro corresponde às flutuações aleatórias do desempenho do examinando. O método das metades, ou coeficiente de consistência, é utilizado quando uma única forma do teste é aplicada numa única sessão e se deseja conhecer apenas a influência da amostragem, mas não a da variação das respostas. Nesse caso, dividem-se os itens do teste em duas metades equivalentes.

Um teste não possui uma única fidedignidade. O seu cálculo é feito em função de um determinado grupo, que sofre a influência de diversas variáveis. O mesmo teste, quando aplicado a outros grupos, terá, muito possivelmente, coeficientes de fidedignidade diferentes, dada a influência de vários fatores. Assim, esses fatores precisam ser levados em consideração quando determinada fidedignidade é interpretada. É preciso estabelecer um controle sobre os mesmos para evitar imprecisão dos resultados e falta de validade dos julgamentos.

A fidedignidade de um teste é afetada em diferentes graus por fatores relativos ao instrumento e ao examinando (VIANNA, 1989). Entre os fatores relativos ao instrumento que podem afetar a fidedignidade do teste estão: número de itens (quanto maior o número de itens, maior a fidedignidade); grau de dificuldade dos itens (itens com grau de dificuldade média são os que mais contribuem para a fidedignidade); homogeneidade do teste (quanto mais homogêneo o teste em sua composição, maior a fidedignidade).

Os fatores relativos ao examinando referem-se principalmente à motivação (testes realizados por examinandos motivados possuem fidedignidade alta), à compreensão das instruções (se as instruções não são claras e o examinando não compreende o que se pede, o grau de precisão das respostas é baixo e a fidedignidade do instrumento também) e às características do respondente (conhecimento, aptidões, reações emocionais, esforço e sorte na seleção de respostas através da “adivinhação” podem alterar a fidedignidade do teste).

#### **4 Metodologia da elaboração de instrumentos**

A construção de um teste que utiliza medidas psicométricas só é possível devido à acessibilidade a

computadores e a pacotes estatísticos, uma vez que envolve cálculos muito complexos para serem solucionados manualmente, especialmente quando abrangem um número avantajado de dados. A construção de um teste desse tipo obedece, de acordo com Kline (1995), à seguinte sequência:

1. fundamentação teórica do teste: momento em que se realiza uma revisão bibliográfica cuidadosa sobre o objeto que se deseja avaliar e também estudos exploratórios, especialmente quando a literatura técnica não possui uma produção suficiente para fundamentar a formulação de itens;
2. formulação de itens do teste: aqui se elabora um número de itens superior ao que se espera obter nas escalas que comporão o teste;
3. análise preliminar de dificuldade de itens: emprego da técnica de juízes;
4. análise da fidedignidade: busca-se verificar a consistência interna do teste;
5. validação do conjunto final de itens do teste para demonstrar que o fator não é apenas uma abstração aleatória, mas um construto que permite uma compreensão do objeto em análise;
6. padronização: momento em que se descreve o processo de aplicação, avaliação e interpretação do teste. Kline alerta para o fato de que o número de itens em um teste não deve ser tão grande que venha a aborrecer os sujeitos e propõe que a aplicação não ultrapasse uma hora para sujeitos adultos.

Schwab (1976) e Freedman e Stumpf (1976) elaboraram e validaram instrumentos em pesquisas na área comportamental. Esses pesquisadores desenvolveram instrumentos para medir percepções dos alunos relativas a professores e disciplinas que pudessem conduzir a experiências de alta e baixa aprendizagem. Alguns procedimentos utilizados por esses pesquisadores inspiraram a metodologia de Raymundo (2006).

Os procedimentos adotados para validar um instrumento de avaliação do nível de consciência linguística englobam nove etapas que compõem três versões do instrumento a ser validado. O Quadro 1, a seguir, foi elaborado para melhor visualizar em que consiste cada etapa e compreender o processo de validação proposto por Raymundo (2006).

O processo de validação implicou a elaboração de dez instrumentos, que serviram como Instrumentos de Coleta de Erros; como Instrumentos Iniciais de consciência linguística; como Instrumento Final de consciência linguística e como Instrumento de Desempenho. A aplicação desses testes foi realizada em quatro momentos distintos. No primeiro semestre de 2004, foram aplicados quatro instrumentos de coleta em uma amostra composta

QUADRO 1 – Procedimentos utilizados para a validação do instrumento.

Versão do instrumento	Etapa	Em que consiste
Primeira Versão (Coleta de Erros)	1) Geração de itens	Coleta de erros para a montagem do instrumento.
	2) Análise de redundância agregada à composição	Agrupamento dos erros segundo a semelhança dos itens e a composição do instrumento.
	3) Validação de conteúdo	Análise da representatividade dos itens por examinadores especialistas
Segunda Versão (Instrumentos Iniciais)	4) Instrumentos Iniciais	Formatação e aplicação.
	5) Validação de construto: fidedignidade do instrumento e do item	Procedimentos estatísticos que visam a calcular coeficientes de fidedignidade para o instrumento e mensurar a consistência interna de cada item e de cada parte do instrumento.
Terceira Versão (Instrumento Final)	6) Validação de construto: retenção de um item no instrumento final	Verificação do grau de contribuição de cada item para a confecção da terceira versão do instrumento: itens com grau de dificuldade média e bom grau de discriminação poderão ser mantidos
	7) Instrumento final	Composição do Instrumento Final e sua aplicação em amostra com o mesmo padrão da amostra usada para a confecção da segunda versão a fim de fazer ajustes finais.
	8) Validação de construto: fidedignidade do instrumento e do item	Procedimentos estatísticos que visam a calcular coeficientes de fidedignidade para o instrumento e mensurar a consistência interna de cada item e de cada parte do instrumento.
	9) Validação de critério	Verificação de uma possível correlação entre os escores do teste de consciência linguística e os escores de um teste de desempenho.

de 114 alunos; no segundo semestre do mesmo ano, foram aplicados quatro Instrumentos Iniciais em uma amostra composta de 102 alunos. O terceiro momento envolveu a aplicação de um Instrumento Final de consciência linguística, em uma amostra composta de 104 alunos, no primeiro semestre de 2005; a quarta aplicação, no mesmo semestre, consistiu na realização de um Instrumento de Desempenho por 74 alunos.

## 5 Construção do instrumento

Com o objetivo de trabalhar com dados empíricos, foram criados exercícios que serviram como **Instrumentos de Coleta de Erros**. Com o título: *How would you say that in English?*, esses exercícios continham frases em português a serem vertidas para o inglês. Devido ao grande número de frases propostas, eles foram divididos em quatro instrumentos. Cada um deles foi constituído de 20 frases simples em português cujas versões envolvem o uso dos seguintes aspectos gramaticais do inglês: *simple present, simple past, future with going to, present continuous, past continuous, present perfect*. Versões afirmativas, negativas, interrogativas do tipo *yes/no questions* e interrogativas do tipo *wh questions* das mesmas frases foram distribuídas entre os quatro exercícios.

Ex.: *How would you say that in English?* (Instrumento 1)

- 1) Ele está com fome. \_\_\_\_\_
- 2) Ela não está com dor de cabeça. \_\_\_\_\_
- 3) Você está aqui há bastante tempo? \_\_\_\_\_
- 4) Onde fica o correio? \_\_\_\_\_

Os instrumentos iniciais de avaliação da consciência linguística, também em número de quatro, foram elaborados com base nos dados obtidos através da coleta de erros. Os Instrumentos Iniciais foram organizados

da seguinte forma: cada instrumento foi dividido em duas partes. A primeira foi constituída de quatro dados de identificação e a segunda, de doze questões a serem respondidas. Os dados de identificação incluíram o nome completo do informante, sua idade, seu nível de inglês na instituição (PUCRS) e o seu curso universitário. As doze questões a serem respondidas foram expressas em frases sublinhadas que faziam parte de pequenos diálogos em inglês. Quatro frases estavam certas e oito, erradas. Cada diálogo foi acompanhado de um quadro a ser completado pelo respondente. Foi solicitado também que o informante marcasse seu grau de segurança em relação às suas respostas (não estou certo – *not sure*, estou quase certo – *almost sure*, estou totalmente certo – *totally sure*).

Ex.: **INSTRUMENT 2**

### Part 1 – IDENTIFICATION DATA

( ) Male ( ) Female

Age: \_\_\_\_\_

Level in the English course: \_\_\_\_\_

Academic course: \_\_\_\_\_

### Part 2 – RIGHT OR WRONG?

A: My two best friends are veterinarians.

B: Really! *Where they work?*

A: They are out of work.

Observe the underlined sentence and answer the following questions	not sure	almost sure	totally sure
Is the sentence right or wrong?			
Where is the error?			
Which is the correct form?			
Why is this the correct form?			

Para compor o **Instrumento final**, procedeu-se a quatro tipos de análise. A primeira análise consistiu em fazer uma avaliação comparativa dos quatro Instrumentos Iniciais, considerando os coeficientes de fidedignidade.

A segunda consistiu em fazer uma avaliação de todos os itens de cada instrumento inicial, considerando o nível de dificuldade e o nível de discriminação. A terceira consistiu na seleção dos itens considerados bons e na eliminação dos que deveriam ser rejeitados por não estarem contribuindo para a precisão do instrumento. O último tipo de análise consistiu na busca pelos melhores itens de todos os instrumentos e na composição de um único instrumento – o Instrumento Final.

O Instrumento Final foi formatado com um número maior de itens, porque, após sua aplicação, sua fidedignidade seria novamente calculada e, dependendo do coeficiente obtido, talvez fosse necessário eliminar algum item para aumentar a precisão do instrumento. A seleção, estrutura e composição do Instrumento Final foram organizadas no Quadro 2, abaixo, para melhor visualização.

O instrumento de desempenho criado é um teste objetivo de escolha simples, intitulado *Choose the best answer* e acompanhado por uma Grade de Respostas *Answer sheet*. Testes desse tipo são muito utilizados para avaliação de conhecimento e classificação de alunos de língua estrangeira; por isso, aceitos como válidos para fins acadêmicos. Para o estudo, o teste foi composto de sete diálogos em inglês, com lacunas numeradas a serem preenchidas pelo informante através da escolha de apenas uma das três alternativas sugeridas abaixo de cada espaço em branco. O instrumento foi composto de trinta lacunas no total. O conteúdo do teste envolveu as mesmas estruturas verbais utilizadas para avaliar o nível de consciência do aluno de nível intermediário. Essas

estruturas são: *simple present, simple past, future, present continuous, past continuous, present perfect*.

Ex.: INSTRUMENT B

CHOOSE THE BEST ANSWER

Fill in *a, b* or *c* on your Answer Sheet. An example has been done for you.

**Mr. Brown** Do you \_\_\_\_\_  $\Sigma$  \_\_\_\_\_ any children?  
 a) have  
 b) has  
 c) had

**Harry** Yes, I do. A boy and a girl.  
 $\Sigma$  The correct answer is (a) have.

Por razões de compatibilidade e acessibilidade, a amostra da pesquisa foi composta de alunos das disciplinas de inglês dos cursos de Inglês Geral (níveis VI, VII e VIII), Letras (níveis I e II) e Ciências Aeronáuticas (níveis I, II e III) da Pontifícia Universidade Católica do Rio Grande do Sul, ministradas por professores da Faculdade de Letras.

As questões que nortearam o trabalho se referiam ao exame do processo de validação do instrumento elaborado para avaliar o nível de consciência linguística, ao grau de precisão do instrumento através do teste das metades e da consistência interna de cada metade e à manutenção e eliminação de itens responsáveis por alta precisão e baixa precisão. Essas questões se fundamentaram em níveis de consciência linguística propostos por Raymundo (2001), os quais foram operacionalmente definidos como pré-consciente 1, pré-consciente 2, consciente e plenamente consciente, a partir do julgamento, identificação, correção e/ou explicação do erro.<sup>5</sup>

QUADRO 2 – Estrutura, conteúdo e organização do Instrumento Final

Item selecionado	Frases a ser analisada	Tempo e aspecto verbal	Estrutura da frase	Grupo de questões
1 (Instr. 2)	1. <i>Where they work?</i>	<i>Simple present</i>	<i>Wh question</i>	erro interlingual
10 (Instr. 1)	2. <i>Do they always make a trip on their holiday?</i>	<i>Simple present</i>	<i>Y/N question</i>	certa
3 (Instr. 2)	3. <i>She found her book in the library?</i>	<i>Simple past (irregular verb)</i>	<i>Y/N question</i>	erro interlingual
4 (Instr. 2)	4. <i>He is studying at the moment?</i>	<i>Present continuous</i>	<i>Y/N question</i>	erro interlingual
5 (instr. 1)	5. <i>What are you going to do tomorrow?</i>	<i>Future</i>	<i>Wh question</i>	certa
6 (Instr. 1)	6. <i>How long are you here?</i>	<i>Present perfect</i>	<i>Wh question</i>	erro interlingual
3 (Instr. 3)	7. <i>When she studied English?</i>	<i>Simple past (regular verb)</i>	<i>Wh question</i>	erro interlingual
4 (Instr. 3)	8. <i>They were studying yesterday afternoon?</i>	<i>Past continuous</i>	<i>Y/N question</i>	erro interlingual
5 (Instr. 3)	9. <i>She taught English from 1987 to 1999.</i>	<i>Simple past</i>	<i>Affirmative</i>	certa
8 (Instr. 2)	10. <i>She don't have a headache now.</i>	<i>Simple present</i>	<i>Negative</i>	erro intralingual
9 (Instr. 1)	11. <i>She didn't found her book in the library.</i>	<i>Simple past (irregular verb)</i>	<i>Negative</i>	erro intralingual
11 (instr. 2)	12. <i>What does he doing at the moment?</i>	<i>Present continuous</i>	<i>Wh question</i>	erro intralingual
10 (instr. 3)	13. <i>Who are you married to?</i>	<i>Simple present</i>	<i>Wh question</i>	certa
12 (Instr. 2)	14. <i>I have worked all day yesterday.</i>	<i>Present perfect</i>	<i>Affirmative</i>	erro intralingual
9 (Instr. 4)	15. <i>Did you worked yesterday?</i>	<i>Simple past (regular verb)</i>	<i>Y/N question</i>	erro intralingual
11 (Instr. 4)	16. <i>Are you studying yesterday afternoon?</i>	<i>Past continuous</i>	<i>Y/N question</i>	erro intralingual

<sup>5</sup> **Níveis de consciência linguística:** inconsciente, pré-consciente 1, pré-consciente 2, consciente e plenamente consciente.

**Nível inconsciente:** quando o aluno erra o julgamento da frase e não consegue perceber quando a frase está errada ou não. Este nível não faz parte da experiência consciente propriamente dita.

**Nível pré-consciente 1:** primeiro nível da experiência consciente. Este estágio envolve a percepção do erro, quando o aluno sabe julgar a frase.

**Nível pré-consciente 2:** segundo nível da experiência consciente. Este estágio envolve a identificação do erro, quando o aluno sabe onde está o erro.

**Nível consciente:** terceiro nível da experiência consciente. Este nível envolve dois estágios de observação que oscilam: a observação da forma, quando o aluno sabe corrigir o erro, ou a observação do uso, quando o aluno sabe explicar a sua correção.

**Nível plenamente consciente:** último nível da experiência consciente. Este nível envolve a compreensão do erro e implica o correto julgamento da frase, a identificação, a correção e a explicação do erro.



O exame da validade desse tipo de instrumento exigiu diferentes procedimentos de análise qualitativa e quantitativa, envolvendo três tipos de validação (conteúdo, construto e critério). A conclusão do processo de validação permitiu constatar que é possível construir um instrumento válido de avaliação do nível de consciência linguística, seguindo as etapas propostas e efetivadas. Contudo, o modelo de instrumento, na condição presumível de pioneiro, deverá ser enriquecido com mais dados, corrigido e melhorado em alguns aspectos, à medida que seja usado com o passar dos tempos. Conforme afirma Vianna (1989), o processo de validação não se exaure, ao contrário, pressupõe continuidade e deve ser repetido inúmeras vezes para o mesmo instrumento.

Além das questões norteadoras, através do instrumento foi possível avaliar outros tópicos, com o intuito de complementar a investigação, enriquecer a interpretação dos dados e estimular a elaboração de técnicas de ensino. Essas questões complementares consistiram em verificar a possibilidade de identificação do nível de consciência linguística dos alunos em relação a erros do tipo interlinguais e intralinguais, a possibilidade de identificação do grau de segurança dos alunos em relação aos diferentes níveis de consciência linguística e em investigar a intensidade da relação entre o nível de consciência do aluno e sua idade.

Os resultados também possibilitaram comparar o comportamento da relação entre o nível de consciência dos alunos e o curso ao qual estavam vinculados e o comportamento da relação entre o nível de consciência linguística dos alunos do sexo feminino e o dos alunos do sexo masculino. Por fim, foi possível verificar a existência de uma possível relação entre o nível de consciência linguística do aprendiz dos diferentes cursos e seu desempenho; e propor uma classificação dos sujeitos de acordo com o uso do processo consciente.

A investigação dessas questões levou às seguintes conclusões: não foi possível identificar, estatisticamente, o nível de consciência linguística do aluno diante da correção de erros interlinguais e intralinguais, ou o grau de segurança dos alunos em relação aos diferentes níveis de consciência linguística.

Constatou-se que não houve relação entre o nível de consciência do aluno e sua idade em termos estatísticos; nem diferença significativa entre o nível de consciência linguística dos alunos dos três cursos. Verificou-se, porém, que houve diferença significativa entre o nível de consciência dos alunos do grupo feminino e do grupo masculino, sendo que o primeiro grupo apresentou nível de consciência superior ao do segundo.

Para a relação entre nível de consciência e desempenho dos alunos dos diferentes cursos, observou-se que, enquanto houve correlação entre as variáveis para os alunos do Curso de Inglês Geral e de Letras, não

houve relação significativa para os alunos de Ciências Aeronáuticas.

Finalmente, diante da proposta de classificação dos alunos conforme o uso do processo consciente, retomou-se a teoria de Krashen (1982) para redefinir sua concepção de superusuários, subusuários e ótimos usuários do monitor,<sup>6</sup> avaliando o nível de consciência e a predominância de grau de convicção de cada sujeito. Com isso, chegou-se a uma nova configuração da hipótese do monitor e a uma nova organização entre os sujeitos.

Através dessa organização, observou-se que a maioria dos alunos do Curso de Letras foram classificados como ótimos usuários do monitor; os alunos de Inglês Geral, como subusuários do monitor e os alunos de Ciências Aeronáuticas oscilaram entre duas classificações, subusuários e ótimos usuários do monitor. Essa oscilação entre os alunos de Ciências Aeronáuticas reforça a inexistência de relação entre nível de consciência e desempenho dos alunos desse curso.

Delineado esse novo quadro, recorreu-se às observações de Schultz (2005), que atribui a características individuais um efeito sobre o monitoramento e o desempenho do aluno de língua estrangeira. Segundo ele, alunos introvertidos, pouco autoconfiantes (superusuários) e extrovertidos, muito autoconfiantes (subusuários) serão pouco beneficiados pelo conhecimento consciente. Ele acredita que os únicos que se beneficiarão desse tipo de conhecimento são os que se situam em um ponto intermediário entre introversão e extroversão (ótimos usuários).

Com isso, constatou-se que os alunos de Letras, em sua maioria, e parte dos alunos de Ciências Aeronáuticas são os que, possivelmente, mais se beneficiam do aprendizado formal das estruturas propostas neste estudo. Essas observações levam a questionar a eficiência do ensino formal de estruturas da língua estrangeira e a refletir sobre a necessidade de identificar atributos de personalidade dos aprendizes para otimizar técnicas de ensino, sugestões que damos para próximas pesquisas.

Para concluir, é importante comentar que o aprendizado de uma língua estrangeira é resultado do

<sup>6</sup> A variação individual observada na produção do adulto resulta em diferentes usos do monitor. Esses usos sugerem que existem basicamente três tipos de aprendiz. O primeiro tipo usa excessivamente o monitor e é classificado como *over-user*. Este recorre ao monitor o tempo todo, constantemente conferindo seu desempenho através do conhecimento consciente. Ele é dependente do aprendizado e não se sente seguro em usar apenas o sistema adquirido. O segundo tipo faz pouco uso do monitor e é classificado como *under-user*. Este prefere não utilizar seu conhecimento consciente, mesmo estando sob condições favoráveis para o uso do monitor. Este tipo de aprendiz não é sensível à correção de erros e usa apenas o seu sistema adquirido para fazer autocorreções. O terceiro tipo de aprendiz é chamado de *optimal monitor user*. Este aprendiz usa o monitor quando é favorável, principalmente na escrita, valendo-se de seu conhecimento aprendido como um suplemento ao seu conhecimento adquirido e tornando o seu discurso mais preciso. O aprendiz, nesse caso, sabe quando usar o monitor com o objetivo de melhorar a sua acuidade gramatical sem prejuízos à comunicação.

acúmulo de informações conscientes, do desenvolvimento de habilidades inconscientes, dos interesses e necessidades de cada grupo e da personalidade de cada aluno. Por essa razão, o ensino de uma língua estrangeira é uma tarefa árdua, que requer muita habilidade, talento e interesse por parte de quem ensina. A construção de testes fidedignos pode auxiliar o professor a diagnosticar falhas de ensino/aprendizado e a desenvolver técnicas adequadas para lidar com essa complexa realidade.

## Referências

- FREEDMAN, R. D.; STUMPF, S. A. *Course faculty instrument: development and application*. New York State University, Graduate School of Business Administration: Oct., 1976.
- KLINE, P. *The handbook of psychological testing*. London: Routledge, 1995.
- KRASHEN, S. D. *Principles and practice in second language acquisition*. New York: Pergamon Institute of English, 1982.
- RAYMUNDO, V. P. *O papel do tratamento de erros na ativação da experiência consciente e na promoção do desempenho de alunos adultos de língua estrangeira: um estudo em língua inglesa*. Dissertação (Mestrado em Linguística Aplicada) – Pontifícia Universidade Católica do Rio Grande do Sul, 2001.
- RAYMUNDO, V. P. *Elaboração e validação de um instrumento de avaliação de consciência linguística*. Tese (Doutorado em Linguística Aplicada) – Pontifícia Universidade Católica do Rio Grande do Sul, 2006.
- SCHWAB, D. P. *Manual for the course evaluation instrument*. Madison, Graduate School of Business and Industrial Relations Research Institute, University of Wisconsin, Nov. 1976.
- SHOHAMY, E. The role of language tests in the construction and validation of second-language acquisition theories. In: TARONE, E. E.; GASS, S. M.; COHEN, A. D. (Org.). *Research methodology in second language acquisition*. New Jersey: Lawrence Erlbaum Associates publishers, 1994.
- SHULTZ, R. “*Interlíngua e fossilização*” *English made in Brazil*. [Online] Disponível em: <<http://www.sk.com.br/sk-interfoss.html>>. Acesso em: 15 maio 2005.
- VIANNA, H. M. *Testes em educação*. São Paulo: IBRASA, 1978.
- VIANNA, H. M. *Introdução à avaliação educacional*. São Paulo: IBRASA, 1989.

Recebido: 24/10/2008  
Aprovado: 28/04/2009  
Contato: <val@puers.br>