

INTRODUÇÃO

Avaliar é uma atividade que está sempre presente em nossa prática diária. Estamos todo momento avaliando para tomarmos decisões. Ao fazermos uma simples compra em um supermercado, avaliamos a qualidade e o preço dos produtos que queremos, e constantemente somos avaliados pelos nossos familiares, colegas de trabalho e da escola. Na verdade, esquecemos todos esses momentos informais em que avaliamos e somos avaliados e só pensamos na avaliação como uma testagem formal, relacionada ao ensino.

A palavra avaliação causa arrepios em muitas pessoas e, de fato, pode causar um grande impacto no nosso dia a dia. Entretanto, a avaliação é necessária para podermos analisar, compreender, melhorar e até mudar uma série de aspectos da vida em sociedade, em todos os campos e áreas.

Em Educação, a avaliação na escola é um tema sempre bastante discutido e autores como Luckesi (1999), Hoffman (1993; 1998) e Vasconcelos (2000) questionam a problemática que esse assunto envolve, e as várias opiniões divergentes que acabam por surgir principalmente pela complexidade do processo avaliativo e seus resultados, o que traz incertezas e questionamentos por parte daqueles envolvidos nesse processo: os protagonistas, também chamados de *stakeholders*¹.

Em um momento no qual a globalização se faz presente, verifica-se que em diferentes países o tema avaliação é considerado prioritário. Segundo a Organização para Cooperação e Desenvolvimento Econômico - OCDE (Afonso, 2000) desde a década de 1990 há uma “renovação do interesse pela avaliação, sendo que esse interesse se deve a três razões:

- a) à necessidade que os países têm de dispor de uma mão de obra qualificada;

¹ O termo *stakeholder*, na literatura de avaliação de língua, refere-se a todas as pessoas e/ou instituições que fazem parte do processo de avaliação em algum nível: os responsáveis pela formulação e correção, os examinandos, e as instituições ou pessoas que usam os resultados das avaliações para algum fim seletivo, como uma empresa num exame de seleção para um emprego, uma instituição educacional para o ingresso num programa de estudos, uma agência financiadora para distribuição de bolsas de pesquisa, entre outros usos.

- b) à necessidade de melhorar a qualidade da educação e da formação para uma melhor utilização dos recursos;
- c) à nova partilha de responsabilidades entre as autoridades centrais e locais na gestão das escolas”.

Em termos de políticas de avaliação, estamos assistindo a uma crescente demanda, com origens em agências nacionais e internacionais, que valorizam a utilização de provas e exames padronizados com fins comparativos, a partir dos quais se pretende induzir mudanças nas escolas e nos professores. Para se adaptarem às lógicas próprias do mercado e da competitividade econômica, a avaliação aparece como um pré-requisito para que seja possível a implementação de mecanismos de mudança no ensino que, por sua vez, impõem a prévia definição de objetivos, pois sem objetivos claros e previamente definidos, não é possível criar indicadores que avaliem o desempenho dos sistemas educativos.

Assim, segundo Carson (2003):²

Os testes sempre foram um aspecto importante na Educação. Os educadores usam testes para determinar o conhecimento do aluno e suas habilidades no que diz respeito ao currículo estar sendo cumprido. Muitas decisões importantes são tomadas com base nos resultados de determinados testes, incluindo testes de classificação em programas educacionais específicos, aulas extras, a atribuição de diplomas e a elegibilidade para a educação em nível superior e assistência financeira. (CARSTON, 2003:4-5).

Se avaliar conteúdos em língua materna implica em tarefa árdua, o processo se torna bem mais difícil em língua estrangeira, pois vários aspectos podem ser contemplados desde que criteriosamente relacionados ao processo de aprendizagem como: estrutura gramatical, fala, pronúncia, compreensão oral, leitura e interpretação, prática escrita por meio de elaborações de frases e textos, e uso de todo esse conhecimento adquirido em um contexto social adequado.

Entretanto a avaliação é necessária não somente para se verificar o quanto um aluno conseguiu aprender daquilo que lhe foi ensinado, mas também para uma reflexão

² No original: “Testing has always been an important aspect of education. Educators use test to determine a student's knowledge and skill with respect to the curriculum being taught. Many important decisions are made based on the results of particular tests including placement in specific educational programmes, provision of remedial instruction, the awarding of diplomas, and eligibility for higher education and related financial assistance. Carston, 2003:4-5).

do professor diante daquilo que foi ensinado, pela escola para verificar se o que foi ensinado pelo professor foi aprendido e compreendido pelo aluno e, por órgãos reguladores e fiscalizadores, para constatar se um determinado programa de ensino está atingindo o objetivo esperado.

Contextualização do problema

Desde final do século XX e início do século XXI mudanças na ordem econômica mundial têm tido repercussão no mundo do trabalho e, como consequência, para acompanhar essas mudanças, alterações e implementações nos currículos dos cursos de formação profissional se fizeram necessárias.

No Estado de São Paulo, a oferta de cursos de educação profissional em nível médio e superior nas redes estaduais foi e está sendo ampliada ano a ano.

O progresso tecnológico tem contribuído para o crescimento do poder dos Mercados, conferindo-lhes agilidade e proporções inéditas na História. Para responder às exigências do Mercado Internacional, não são poupados esforços para incorporar avanços tecnológicos e mudanças nos modos de formação e de gestão das forças produtivas que permitam às diversas instituições alcançar maior competitividade. Como decorrência, aparece a exigência de uma maior escolaridade e uma qualificação profissional mais especializada, mesmo para trabalhadores que atuam em ofícios considerados mais simples. (FERNANDES, 2011:1).

O mercado de trabalho estipula qual é o perfil do trabalhador desejado: capacidade de observação, raciocínio lógico, hábito de leitura diária, rigor nas análises, precisão, clareza, uso correto da linguagem oral e escrita e domínio da língua inglesa. Com a economia internacionalizada, o conhecimento do idioma é um pré-requisito cada vez mais importante na disputa por melhores empregos. O conhecimento de outros idiomas além do oficial dominante dentro do país deve compor a qualificação de profissionais, principalmente quando a profissionalização exige contato com o mercado externo. Um dos fatores considerados por empresas estrangeiras para terceirizar processos de negócios para outros países é o nível de educação da população e proficiência em um idioma, principalmente em inglês.

Os países em desenvolvimento reconhecem que uma das formas para ter uma economia de serviço forte é ter graduados capazes de se comunicar em inglês, de forma que muitas escolas estão integrando o inglês desde a Educação Infantil até o Ensino Médio, usando a língua como meio de instrução e não apenas ensinando o idioma como uma disciplina separada, que faz parte do currículo escolar. Cada vez mais, o inglês está sendo incluído em testes nacionais padronizados, como a recente inserção da língua inglesa na prova do Exame Nacional do Ensino Médio (ENEM) e testes internacionais como Testing of English as a Foreign Language (TOEFL), Test of English for International Communication (TOEIC), The Business Language Testing System (Bulats), International English Language Testing System (IELTS), entre outros.

A qualificação em idiomas é um dos maiores desafios do país em seu processo de internacionalização. Em seu trabalho sobre o futuro da língua inglesa, Graddol (2006) relata que haverá um aumento no uso da língua inglesa por grupos de estudantes, trabalhadores imigrantes, e das populações nas nações que estão emergindo como os poderes econômicos do futuro. O que todos esses grupos compartilham é o contexto no qual eles desejam utilizar a língua inglesa: que seja direcionado para sua área de trabalho.

Pela grande quantidade de institutos de idiomas que oferecem cursos de inglês no Brasil percebemos a valorização do idioma por parte da sociedade, pois é fato notório que o mundo inteiro e, naturalmente o Brasil, reconhecem que aprender inglês é de vital importância a fim de se entrar em contato com as últimas inovações tecnológicas e científicas e que aqueles que têm um bom domínio do idioma têm muito mais vantagens sobre aqueles que não o têm.

Apesar de a língua inglesa ser um requisito básico para determinadas profissões como diplomacia e tradução, temos que o conhecimento da língua inglesa está implícito em quase todos os cargos de setores da economia, administração, marketing, engenharias e na área da saúde. As empresas que oferecem vagas para “*trainees*” eliminam candidatos que não falam inglês, sendo que o mesmo critério vale para a contratação de executivos mais experientes e aqueles com habilidades em inglês acima da média local recebem, em média, salários 30-50% mais altos que candidatos com as mesmas qualificações, mas sem o domínio do idioma. Assim, temos um círculo virtuoso

no qual a melhor proficiência leva a melhores salários, que pode resultar em um melhor desempenho por parte dos funcionários dentro das empresas.

Uma das preocupações dos cursos superiores em geral, e em especial das faculdades de tecnologia é com a inserção de seus formandos no mercado de trabalho. Para tal, as faculdades de tecnologia têm buscado para seus alunos uma formação profissional privilegiada, estando aí inserido o domínio de uma língua estrangeira, no caso o inglês. Em algumas delas, principalmente as estaduais paulistas, reestruturações foram feitas no currículo dos cursos tecnológicos de três anos inserindo como disciplina obrigatória 6 (seis) semestres de Língua Inglesa.

A faculdade que faz parte do corpus de pesquisa deste estudo institui 6 (seis) semestres de língua inglesa em seu currículo. Para que as aulas pudessem ser melhor aproveitadas, com uma menor quantidade de alunos em sala de aula, foi instituído um teste de classificação em Língua Inglesa (*placement test*) mediado pelo computador, para todos os alunos ingressantes, cujo conteúdo está atrelado à ementa do curso, que está direcionada à área de Inglês para Negócios. O objetivo do teste é avaliar diferentes níveis de desempenho desses alunos e posicioná-los em classes cujos alunos tenham níveis linguísticos semelhantes.

Objetivos do estudo

A utilização de testes tem sido uma prática comum nos sistemas educativos. Vários tipos de testes são usados para objetivos diversos como classificação, seleção, desempenho e certificação.

Para a elaboração de um teste que seja coerente com os objetivos delineados pela instituição é importante que tenhamos um bom conhecimento quanto às funções e tipos de avaliação. Assim, os exames normalmente são divididos em categorias como: entrada/ classificação, rendimento, progresso/domínio, diagnóstico, proficiência.

No caso deste trabalho de pesquisa, analisaremos um teste de classificação que tem como característica avaliar o nível de habilidade linguística do aluno de forma que possa ser colocado em um estágio de um programa de ensino apropriado as suas habilidades.

Um teste de classificação é um elemento importante na maioria dos programas, devido à necessidade de classificar os alunos em grupos de capacidade linguística relativamente homogêneas, às vezes dentro de áreas de habilidades específicas e diversas.

O propósito desta pesquisa é coletar e analisar evidências relacionadas a um teste de classificação que foi aplicado em cursos superiores de tecnologia durante cinco semestres consecutivos – do 2º semestre de 2010 ao 2º semestre de 2012. Para tanto utilizaremos os dados estatísticos de uma faculdade de tecnologia pública localizada a 60 km da capital de São Paulo.

O objetivo é avaliar a qualidade dos itens que compõem cada nível desse teste de classificação, fazendo-se uma comparação com o conteúdo programático da disciplina Inglês, que consta na ementa dessa faculdade de tecnologia, além da validação de seu construto, que se destina a conhecimentos sobre a língua inglesa e formas de como usar essa língua em situações diversas, principalmente, na área de negócios. Além disso, serão também observados os fatores ligados à validade de construto, a validade de conteúdo e a validade de face desse teste.

A avaliação desenvolvida por meio de testes de classificação é, provavelmente, uma das mais utilizadas nas instituições de ensino, e segundo autores como Hughes (1989); Bachman & Palmer (1996) e Brown (1996), esses testes dispõem de duas funções principais:

- investigar o domínio das habilidades linguístico-comunicativas do aluno;
- ajudar na criação de grupos de alunos com habilidades linguístico-comunicativas homogêneas.

Normalmente, a primeira função é a que predomina, entretanto, a segunda é utilizada como resultado de considerações práticas tais como: o número e a diversidade dos alunos, flexibilidade administrativa, a eficácia nos custos, ou a necessidade educacional de se posicionar grupos com as mesmas habilidades linguístico-comunicativas em um mesmo ambiente de sala de aula. (Bachman (1990), Brown, 1996).

Não há muitas pesquisas e literatura relacionadas sobre análise de testes de classificação e em relação à validação dos mesmos, incluindo conceitos como: confiabilidade, praticidade, autenticidade e validade de tal procedimento. Podemos mencionar alguns autores que pesquisaram esse tipo de teste, entre eles Brown (1989); Bradshaw (1990); Blais & Laurier (1995), Fulcher (1997a, 1999b) que têm discutido o potencial dos resultados dos testes de classificação para informar aspectos do ensino e contribuir para a avaliação de programas educacionais, Wall et al. (1994) que descrevem o uso de testes de classificação em um cenário de Inglês para Fins Acadêmicos (EAP – English for Academic Purposes) para diagnosticar determinadas áreas de fragilidade que podem ser solucionadas por estudos de Inglês como língua estrangeira, e Santos (2007) que analisa um teste para nivelamento de alunos de Português como Língua Estrangeira.

Justificativa da pesquisa

Em 1970, por meio do Parecer CEE/SP nº 50/70, p.12, o Conselho Estadual de Educação de São Paulo autorizou a instalação e o funcionamento dos primeiros cursos de tecnologia do Centro Estadual de Educação Tecnológica de São Paulo, ressaltando que “(...) o tecnólogo virá preencher a lacuna geralmente existente entre o engenheiro e a mão de obra especializada, (...) deverá saber resolver problemas específicos e de aplicação imediata ligados à vida industrial...” e que “vem a ser uma espécie de ligação do engenheiro e do cientista com o trabalhador especializado (...) e está muito mais interessado na aplicação prática da teoria e princípios, do que no desenvolvimento dos mesmos (...)”.

De acordo com a Resolução nº 3 de 18 de dezembro de 2002, os cursos de Educação Profissional de Nível Tecnológico serão designados como Cursos Superiores de Tecnologia e deverão:

- I. incentivar o desenvolvimento da capacidade empreendedora e da compreensão do processo tecnológico, em suas causas e efeitos;
- II. incentivar a produção e a inovação científico-tecnológica, e suas respectivas aplicações no mundo do trabalho;

III. desenvolver competências profissionais tecnológicas, gerais e específicas, para a gestão de processos e a produção de bens e serviços;

IV. propiciar a compreensão e a avaliação dos impactos sociais, econômicos e ambientais resultantes da produção, gestão e incorporação de novas tecnologias;

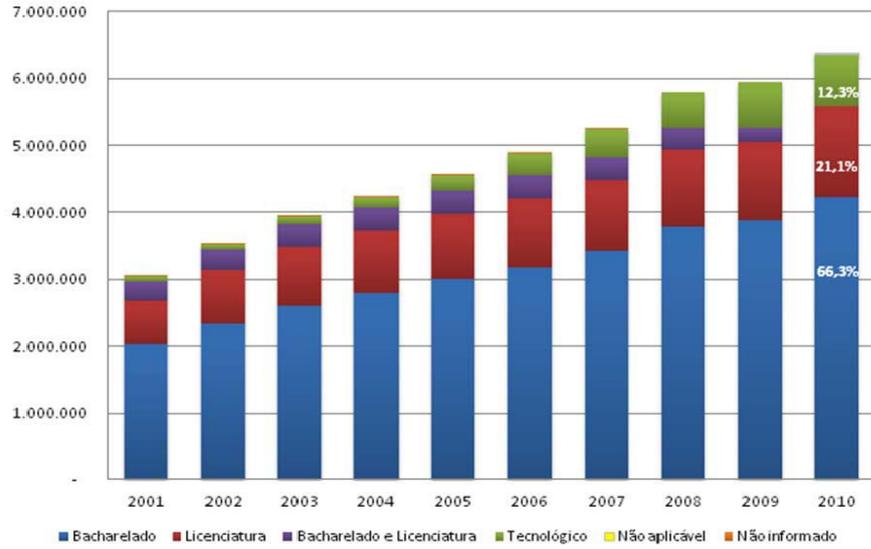
V. promover a capacidade de continuar aprendendo e de acompanhar as mudanças nas condições de trabalho, bem como propiciar o prosseguimento de estudos em cursos de pós-graduação;

VI. adotar a flexibilidade, a interdisciplinaridade, a contextualização e a atualização permanente dos cursos e seus currículos;

VII. garantir a identidade do perfil profissional de conclusão de curso e da respectiva organização curricular.

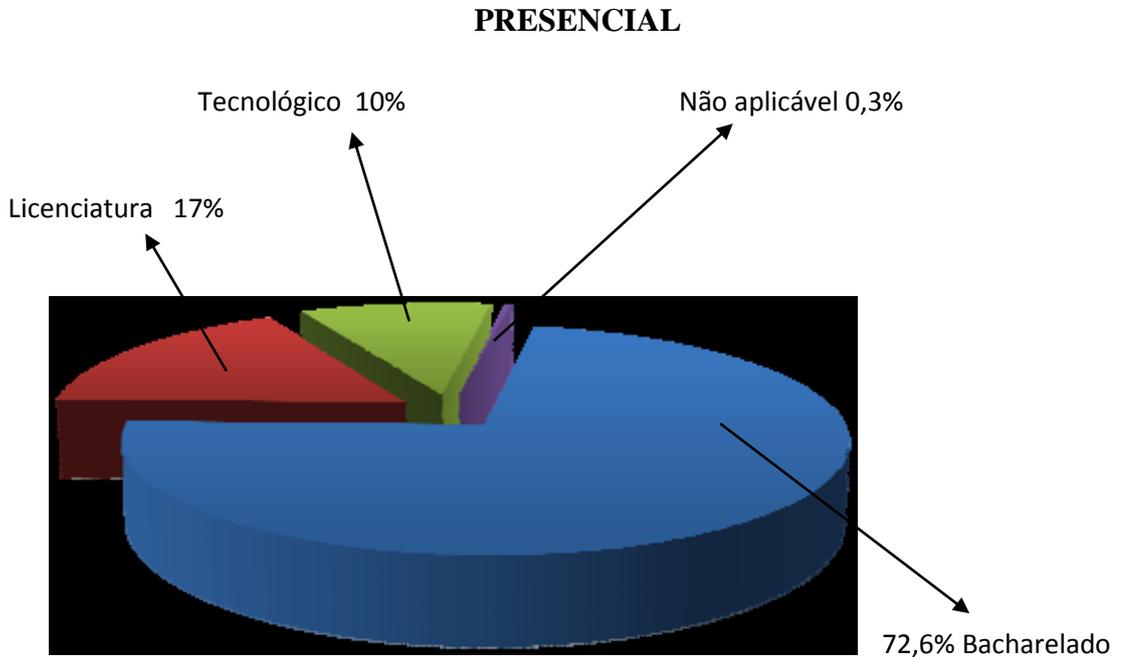
Desse modo, com maior propriedade, ficou evidenciado que o que caracteriza os cursos superiores de tecnologia não é a sua duração e, sim, o seu perfil profissional de conclusão.

O Censo da Educação Superior 2010 realizado pelo INEP/MEC aponta o aumento das matrículas nos cursos tecnológicos. Os cursos presenciais atingiram os totais de 3.958.544 matrículas de bacharelado, 928.748 de licenciatura e 545.844 matrículas de grau tecnológico. Em 2001, esse número era de 69.797 e atingiu em 2010 um total de 781.609 matrículas. Assim, houve uma elevação de 2,3% em 2001 para 12,3% em 2010, como podemos observar nos gráficos a seguir, demonstrando o interesse dos alunos pelo ensino tecnológico.



Não informado: corresponde aos cursos que não informaram grau acadêmico nos Censo da Educação Superior até o ano de 2008.
 Não aplicável: corresponde à Área Básica de Curso.
 Fonte: MEC/Inep, 2012:12

Gráfico 1: Evolução do Número Total de Matrículas (presencial e a distância) por Grau Acadêmico – Brasil – 2001-2010



* A categoria “Não aplicável” corresponde à Área Básica de Curso

Fonte: MEC/Inep, 2012:10

Gráfico 2: Número de Matrículas por Modalidade de Ensino e Grau Acadêmico – Brasil – 2010

A partir dos gráficos podemos identificar que o crescimento pelo interesse nos cursos superiores em tecnologia está em ascensão, devido a investimentos na educação profissional de nível superior, principalmente pela iniciativa privada.

Ainda, segundo o Censo da Educação Superior 2010, as matrículas nos cursos tecnológicos foram, em grande parte, nas áreas de Gerenciamento e Administração, centralizando 44% das matrículas, constando em 2º lugar a área de Processamento da Informação, com 8,5% das matrículas, seguindo-se as áreas de Ciência da Computação, com 6,6%, Marketing e Publicidade, com 6,1% e Proteção Ambiental (cursos gerais), com 5,1%.

Foram realizadas pesquisas pelos órgãos diretivos dessas faculdades de tecnologia e por uma equipe de professores de Inglês que identificaram a situação dos recém-formados e desses cursos de tecnologia e o mercado de trabalho nas áreas dos diversos cursos oferecidos por essas instituições de ensino. Surpreendentemente, constatou-se que o mercado vem contratando profissionais de outras áreas para funções para as quais aqueles ex-alunos apresentam toda a especialização necessária, e que essa contratação ocorre em decorrência da falta de proficiência na língua inglesa. Assim, profissionais formados em diferentes áreas – Letras, Tradução, Turismo, Publicidade, por exemplo, vêm sendo contratados como profissionais nas áreas de Logística, Secretariado, Análise de Sistemas e Tecnologia da Informação, para citar aqui apenas algumas, pelo fato de serem capacitados para a comunicação em língua inglesa.

A equipe de professores de Inglês que pesquisou a situação dos egressos desenvolveu um teste de classificação em Língua Inglesa por meio eletrônico, que busca avaliar a capacidade dos ingressantes em utilizar a língua inglesa em situações comuns do dia a dia, bem como em situações com foco em Inglês para Negócios.

O teste, que é feito *online*, possui duas etapas iniciais. A primeira parte consiste de um questionário sociocultural, que tem por objetivo saber se o examinando tem algum conhecimento da língua-alvo, estuda ou estudou a língua-alvo somente no ensino regular ou frequenta também institutos de idiomas, se já viajou para o exterior, ou se utiliza a língua-alvo no seu trabalho ou em outras atividades.

A segunda parte é o teste em si, dividido em 6 (seis) níveis, sendo que cada nível é composto de 10 (dez) questões de múltipla escolha, cujo enfoque pode ser gramatical, situacional, leitura e compreensão de textos e uma parte de recepção oral. Nesta última parte o examinando faz uso do fone de ouvido para ouvir um diálogo e escolher uma opção de resposta, entre 4 (quatro) alternativas, no computador. Em cada fase o examinando precisa acertar 60% das questões, ou 6 (seis) itens para passar para a próxima fase. A razão da nota 6,0 (seis) é que a média para aprovação em qualquer disciplina que consta no currículo dessas faculdades é 6,0 (seis). Se o examinando atingir o nível 3(três) ou ultrapassar esse nível, ele será convidado a fazer uma entrevista com um examinador, que normalmente é o professor da disciplina Inglês, para verificar a oralidade do examinando, sendo que a entrevista tem a função de confirmar o nível adequado que esse examinado será posicionado.

A opção pela realização dessa etapa da avaliação através do computador está alicerçada na necessidade de se classificar os alunos ingressantes nos devidos níveis de proficiência no menor tempo possível, além disso, por esses alunos terem ingressado em cursos de tecnologia, o objetivo é também colocá-los diretamente em contato com a tecnologia disponível no mercado e que estará acessível a eles durante os cursos.

Perguntas de pesquisa

A avaliação representa um importante desafio no trabalho realizado por professores e alunos, e vem despertando, nos últimos anos, crescente atenção daqueles que trabalham com educação, tanto no ensino superior como em outros níveis escolares.

Embora a influência e a importância dos exames de línguas sejam consideradas como imprescindíveis, normalmente o processo de avaliar sempre foi feito com certa relutância. Porém, há hoje um reconhecimento de que esse estado em relação à avaliação não se aplica mais, pois ao invés de se ver os exames por um lado negativo, como uma força conservadora, uma ameaça, reconhece-se pontos positivos na sua aplicação, uma vez que o exame exerce uma força poderosa nos aprendizes e no ensino.

Normalmente um teste de classificação pode não ser considerado de alto impacto (*high stakes test*)³. Wall, Clapham, e Alderson, (1994) oferecem um dos poucos estudos empíricos de um teste de classificação que se destina a mapear as deficiências em língua inglesa de alunos que entram em uma universidade britânica, de forma que algum problema na competência linguístico-comunicativa possa ser detectada e trabalhada para não impedir seu progresso nos estudos de graduação e pós-graduação.

Tendo como objeto de estudo o teste de classificação já mencionado no item anterior, os questionamentos que direcionam este trabalho são os seguintes:

- 1) O teste de classificação proposto é válido para o posicionamento dos alunos dos cursos superiores em tecnologia analisados neste trabalho? (Cursos de Eventos, Logística e Gestão da Tecnologia da Informação).
- 2) O conteúdo do teste é condizente com a ementa e o conteúdo programático do curso?
- 3) Os examinandos identificam que o resultado do teste os está classificando adequadamente?
- 4) Em que medida tem sido um norteador para o planejamento do ensino e atividades nos cursos superiores em tecnologia?

O contexto da pesquisa

Com as novas políticas públicas implantadas e implementadas pelos governos federal, estadual e municipal, os Cursos Superiores de Tecnologia têm tido nos últimos anos um crescimento considerável, conforme já apontamos. De acordo com o Censo do Ministério da Educação um maior número de alunos das classes menos privilegiadas, que tinham pouca possibilidade de chegar ao ensino médio e terminá-lo, está tendo essa oportunidade de concretizar o sonho de cursar uma faculdade.

Os cursos de tecnologia gratuitos oferecidos por essas instituições em várias áreas de conhecimento como análise e desenvolvimento de sistemas, gestão da tecnologia da informação, logística, manutenção industrial, automação, somente para

³ Testes de alta relevância - O termo "testes de alta relevância" foi incorporado do inglês (*high-stakes test*), expressão concebida na década de 80 no meio acadêmico norte-americano para designar avaliações externas de prestígio, que têm consequências importantes para o candidato (em vários aspectos), ou é tomado como base para grandes decisões, como admissão em cursos ou em empregos. (McNamara, 2000: 133).

citar alguns, são disputados principalmente por candidatos oriundos de escolas públicas, que veem neles perspectivas promissoras em termos de mercado de trabalho.

No caso do contexto desta pesquisa e no período pesquisado, que vai do 2º semestre de 2010 ao 2º semestre de 2012, 61% dos alunos são provenientes de instituições públicas. Sabe-se também que, em geral, o ensino de língua estrangeira nas escolas fundamental e média pública ainda é deficiente e as escolas não conseguem habilitar seus alunos a terem uma competência da língua estrangeira adequada para atuar no mercado de trabalho. No capítulo IV iremos detalhar o cenário da pesquisa e seus participantes.

Além do documento de pesquisa que é a análise do teste em si, questionários foram aplicados aos alunos ingressantes e professores da unidade de ensino. Os resultados do Questionário II denominado “Questionário de percepção da prova” serão apresentados e discutidos no Capítulo V, com o objetivo de ratificar as conclusões referentes aos dados da validade de conteúdo, bem como o Questionário III, que apresenta a visão de duas professoras que administram a disciplina Inglês na unidade de ensino.

Organização da tese

O presente estudo está organizado de maneira que após a introdução, onde contextualizamos este estudo e apresentamos o objetivo e justificativa para a pesquisa, apresentamos a metodologia adotada para a condução deste trabalho, que tem como objetivo situar o leitor sobre o percurso da pesquisa, coleta de dados, participantes e cenário dos envolvidos, o objeto de estudo e por entendermos que a metodologia também embasa a escolha do referencial teórico que apoia todo este estudo.

No primeiro capítulo discutiremos o conceito de competência linguístico-comunicativa e a visão de linguagem que norteia o ensino de línguas nessa faculdade de tecnologia.

No segundo capítulo um breve histórico da avaliação de línguas será apresentado, além dos conceitos relacionados às práticas avaliativas, que são: praticidade, confiabilidade, autenticidade, efeito retroativo e impacto, uma discussão

sobre o que é validade, o conceito tradicional e contemporâneo do termo, bem como os tipos de validade e formas de validação de testes.

No terceiro capítulo trataremos exclusivamente sobre a avaliação na área de línguas estrangeiras, abordando os tipos de exames de línguas, e formas como esses exames podem ser elaborados, principalmente a prova objetiva e formato de questões de múltipla escolha, além da forma como os testes de recepção/produção oral e escrita são elaborados, além do impacto da tecnologia na elaboração de testes.

Em seguida, no capítulo 4, apresentaremos a metodologia de pesquisa escolhida para conduzir esse estudo: o estudo de caso, e sua fundamentação teórica, ancorado por dados quantitativos e qualitativos, que servirão de base para a análise dos dados.

A análise de todos os itens que compõem o teste de classificação a partir de uma comparação com as ementas e o conteúdo programáticos dos 6 (seis) níveis de língua inglesa será apresentada no quinto capítulo.

No capítulo 6 discutiremos os resultados obtidos por meio da análise do teste e questionários, além de apresentarmos o processo de validação e a caracterização das validades de construto, de conteúdo e de face, e retomamos as perguntas de pesquisa.

Nas considerações finais, sintetizamos as contribuições e os principais pontos do estudo, além de levantar limitações e desafios, e de apontar encaminhamentos. Por fim, reunimos as referências bibliográficas que fundamentam a pesquisa, seguidas dos anexos e apêndices.

CAPÍTULO 1 – A LINGUAGEM NA AVALIAÇÃO DE LÍNGUAS

O objetivo deste capítulo é trazer a fundamentação teórica referente à literatura de pesquisa da área de Linguística Aplicada ao longo das últimas três décadas, a fim de usá-la como arcabouço teórico para a investigação que está sendo desenvolvida neste trabalho. Para tanto procuraremos apresentar diferenças terminológicas referentes às palavras “assessment”, “evaluation”, “test”, as visões de linguagem pelas quais a avaliação em língua estrangeira passou e sua importância no ensino; o conceito de competência comunicativa e proficiência; e a linguagem utilizada em testes para fins específicos, com foco em inglês para negócios.

1.1. A natureza da avaliação de línguas

Testes de línguas ocorrem em diversos contextos, entretanto, para Shohamy (1992:513) os dois mais importantes são: 1) o contexto escolar, em que os testes e outros procedimentos de avaliação são usados como parte do processo de instrução e para melhorar o ensino e aprendizagem na escola; 2) no contexto externo, em que os testes são usados para tomada de decisões sobre o futuro dos indivíduos, como quando aceitamos candidatos para um determinado programa, colocando-os em determinados níveis.

O uso principal de qualquer avaliação na área de línguas é coletar informações para uma tomada de decisões. Além disso, o uso de uma avaliação e a tomada de decisões terão consequências para as partes interessadas, os alunos e os programas de ensino no cenário educacional e social nos quais as avaliações ocorrem. As avaliações de línguas são então usadas para nos ajudar a tomar decisões que idealmente deveriam levar a consequências benéficas para as partes interessadas: alunos, professores, instituições de ensino.

Entretanto, na área de avaliação, os termos “assessment”, “measurement”, “test” e “evaluation” são normalmente utilizados para se referir a práticas muito semelhantes: coletar informações. No entanto, parece não haver um acordo referente à

natureza precisa das diferenças entre esses termos. As palavras “*test*”, “*assessment*” e “*evaluation*” quando traduzidas para o português e utilizadas no âmbito escolar podem trazer uma ideia única de significado, e podemos até pensar que os termos são sinônimos. Como este trabalho trata especificamente de avaliação, é, no entanto, importante buscar esclarecer a definição aplicada a cada um desses termos.

As palavras “*testing*” e “*assessment*” são dois termos que muitas vezes são usados indistintamente, passando a ideia de que ambos têm a mesma definição. No entanto, são diferentes. O teste é realmente uma das formas de avaliar alguma coisa. Os testes, de fato, podem servir como um dos métodos de avaliação para coleta de informações sobre a aprendizagem dos alunos. O objetivo da maioria dos testes é a de atribuir notas aos alunos. Eles oferecem informações de diagnóstico limitado para identificar áreas de melhoria.

Para atingir o objetivo final de avaliação - para melhorar a aprendizagem do aluno - temos de responder não só questões sobre o quanto os alunos aprenderam, mas também perguntas sobre como eles aprenderam e por que certos resultados ocorreram. Essas informações são obtidas através de diferentes testes, incluindo observações, pesquisas, entrevistas, tarefas de desempenho, seminários, portfólios, etc. Assim, “*assessment*” é um conceito abrangente, centrando seus esforços na aprendizagem dos alunos, e servindo ao propósito de melhoria do aluno e desenvolvimento através de uma variedade de maneiras.

Bachman (1990, 2004b, 2010) aponta algumas distinções entre os termos. “*Assessment*” é o processo de coleta de informações sobre algo que nos interessa, de acordo com procedimentos que são sistemáticos e que têm uma base substancial. O resultado desse processo, como uma pontuação ou uma descrição verbal, recebe o nome de “*assessment*”. Na avaliação de línguas, é imprescindível fazer uma interpretação a respeito das habilidades linguísticas do examinando. Nesse caso, o termo “*evaluation*” envolve julgamento de valor e tomada de decisões sobre as informações coletadas.

Frequentemente se usa “*assessment*” para avaliação em programas educacionais, nos quais podemos identificar os pontos fortes e as dificuldades dos alunos para poder ajudá-los na sua aprendizagem, para selecionar e classificar alunos em programas instrucionais, ou decidir quais alunos serão promovidos de nível.

Os termos “assessment” e “evaluation” constantes do Quadro Europeu Comum de Referência de Línguas na versão inglesa aparecem na versão para o português traduzidos por uma só, a palavra “avaliação”, como podemos observar no trecho abaixo.

Neste capítulo, o termo ‘avaliação’ é usado no sentido de avaliação da proficiência do usuário da língua. Todos os testes de língua são uma forma de avaliação, mas há também muitas outras formas de avaliação (p. ex.: as listas de verificação usadas na avaliação contínua, a observação informal do professor), que não são descritas como testes. A avaliação é um termo mais vasto que a testagem. Qualquer testagem é uma forma de avaliação, mas num programa de língua existem muitos outros aspectos, para além da proficiência do aprendiz, que também são avaliados – p. ex.: a rentabilidade de determinados métodos ou materiais, o tipo, a qualidade do discurso efetivamente produzido, a satisfação do professor e do aprendiz, a eficácia do ensino, etc. (CEFR, 2001:177).

O termo ‘avaliação’ também é empregado como sinônimo de ‘teste’ ou ‘prova’, termos estes que nesta pesquisa serão empregados com o sentido de instrumentos de avaliação pelos quais se faz a aferição do processo de ensino/aprendizagem.

As relações de uso entre os termos: *assessment*, *measurement* e *tests* estão ilustrados na figura 1.

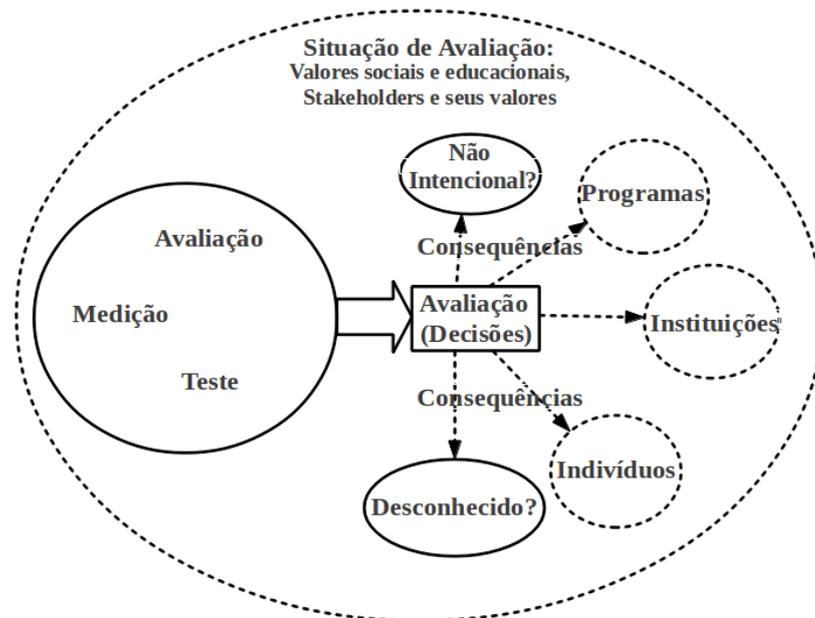


Figura 1: Relações entre *assessment/measurement/test*, seus usos para a avaliação e as consequências de seu uso.

Fonte: Bachman, 2010:22

Como indicado na figura 1, as avaliações que usamos e as decisões que tomamos estão situadas nos cenários social e educacional, e terão consequências para os indivíduos, programas e instituições que estão inseridos nesses cenários. Inerente a tudo isso estão os valores do indivíduo, das instituições e das comunidades e os vários protagonistas envolvidos com as comunidades sociais e educacionais, bem como as leis, regulamentações e as políticas públicas que delimitam as formas como essas avaliações devem ser usadas. Assim, como este estudo versa sobre o tema avaliação, temos de levar em consideração as consequências potenciais do uso da avaliação e as decisões que são tomadas para os diferentes protagonistas numa situação de avaliação.

Segundo Bachman (2010: 22-25) uma das formas de se pensar no uso da avaliação de línguas é como uma série de inferências a partir do desempenho do candidato até suas consequências, que estão ilustradas nas figuras 2 e 3.

As decisões que tomamos são geralmente sobre os examinandos e os programas. Bachman (2010:24-25) resume essas decisões em:

1. Decisões sobre os indivíduos (micro-avaliação):
 - seleção para admissão em um programa ou para trabalho.
 - classificação em um curso ou estudo.
 - fazer alterações em instruções ou nas atividades de ensino/aprendizagem.- aprovar ou reprovar um aluno em um curso com base na sua avaliação de rendimento.
 - predição de um desempenho futuro.
2. Decisões sobre o programa (macro-avaliação):
 - Formativa, relacionada a mudanças para o melhoramento de um programa pré-existente.
 - Somativa, relacionada à continuidade de um programa pré-existente ou implementação de um novo programa.
3. Decisões sobre a pesquisa:
 - Decisão sobre novas perguntas de pesquisa ou metodologia para pesquisas futuras.

- Mudança de ponto de vista sobre um fenômeno linguístico particular.
- Modificação de entendimento ou explicação de um fenômeno linguístico particular. (Bachman, 2004b:10)

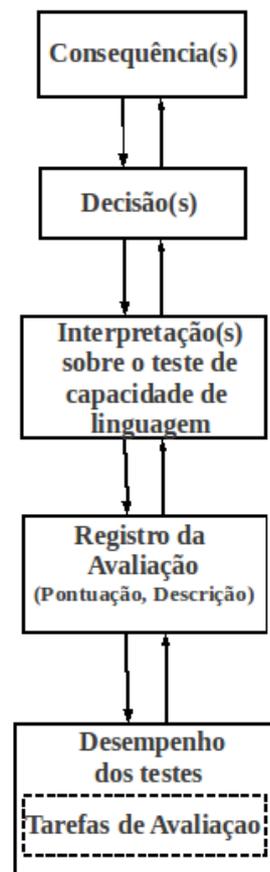


Figura 2: Ligações entre o desempenho do examinando aos usos pretendidos (decisões e consequências)

Fonte: Bachman, 2010:23

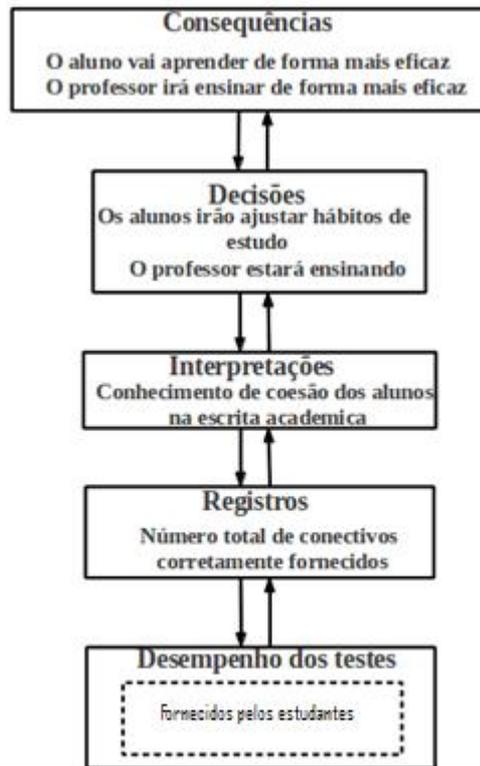


Figura 3: Relações entre o desempenho do examinando e as consequências de um teste

Fonte: Bachman, 2010:24

Além disso, como está indicado nas figura 2 e 3 , poderão haver consequências indesejadas ou desconhecidas. É responsabilidade das pessoas que estão envolvidas na preparação de um teste identificar as possíveis consequências indesejáveis e tentar estabelecer procedimentos para que as mesmas sejam minimizadas. Entretanto, apesar de muito planejamento, ainda podem surgir imprevistos que não foram detectados ou antecipados.

Ao falarmos de avaliação, principalmente em testes, um ponto muito importante é assegurar que um teste seja bem construído, coerente e claro, e que seja um instrumento que forneça informações precisas das habilidades do candidato em um determinado domínio. Para tanto, existem certos critérios que devemos observar na elaboração de testes que garantam que os mesmos avaliem o domínio pretendido, sem incluir fatores desnecessários, que possam interferir nos resultados e/ou na interpretação dos mesmos.

Um desses fatores é a visão de linguagem que o professor adota para sua prática docente, que tem consequências no tipo de avaliação praticada. É a partir da visão de linguagem postulada pelo professor que sua prática se fundamenta. Schlatter et alii (2004:366) afirmam que “*conceitos de avaliação e de uso de linguagem tornam-se (...) indissociáveis, influenciando-se mutuamente*”.

1.2. Visões de Linguagem

No final da década de 80 e começo da de 90 a área de avaliação começa a ter destaque dentro da Linguística Aplicada, *status* este corroborado por Bachman (1991) que afirma que a partir desse período a avaliação é reconhecida como uma área de pesquisa e influência para o ensino de língua estrangeira e Alderson (1991) também comenta a criação da Revista *Language Testing*, que auxiliou na disseminação de assuntos relacionados à área de Avaliação e testes em língua estrangeira, sendo que uma das preocupações dos pesquisadores é a elaboração de testes que apresentem questões de forma contextualizada, simulando situações de linguagem na qual o aluno passaria na vida real. Bachman e Palmer (1996) apontam que não há fórmulas prontas para a elaboração de testes e que não existe o melhor teste. O que deve ser levado em consideração é o objetivo para determinado teste, quem são os personagens envolvidos, para que se tenham tarefas confiáveis e válidas.

Ao se falar em avaliação em contextos escolares, e no caso específico deste trabalho, da avaliação em língua estrangeira, temos de pensar que ela é um elemento fundamental em todo o processo de ensino e aprendizagem, de forma que deve ter objetivos coerentes com a visão de linguagem que orienta o ensino.

Como grande parte das dificuldades do professor no exercício de sua prática avaliativa é gerada pela ausência de uma reflexão acerca do construto que irá fundamentar sua avaliação, seria importante que o professor avaliador de língua estrangeira estivesse capacitado para responder às seguintes perguntas: Qual é a minha visão de linguagem? E de língua estrangeira? Qual minha visão de ensino/aprendizagem de língua estrangeira? Minha prática de ensino é consistente com minha visão de língua estrangeira e de ensino/aprendizagem de língua estrangeira? Minha prática avaliativa é consistente com essa visão de ensino/aprendizagem? A seleção dos materiais didáticos, métodos e técnicas por mim usados são consistentes com essa visão? (SCARAMUCCCI, 2006: 58-59).

Neste momento, uma das visões que direciona o ensino de língua estrangeira é a perspectiva sócio-interacionista, baseada em Vigotsky (1993), que afirma que o desenvolvimento cognitivo se dá através das interações sociais, principalmente através da linguagem.

Na perspectiva bakhtiniana (Bakhtin,1992) a linguagem é entendida como um fenômeno social e histórico, em que a dialogia está em destaque pois para o autor, a interação entre falantes influencia as escolhas feitas em termos da linguagem, pois é através do discurso que trazemos nossos valores, as nossas concepções culturais e sociais.

Nessa visão, aprender é uma forma de estar no mundo social com alguém, em um contexto histórico, cultural e institucional, mediado pela interação, visando à construção de um conhecimento conjunto a ser realizado por aluno e professor.

A partir da perspectiva sócio-interacionista várias implicações surgem na área de ensino e aprendizagem de línguas e conseqüentemente na área de avaliação. É importante para o ensino que os alunos interajam com seus colegas por meio do uso de tarefas que simulem situações reais, visando sempre o aspecto social da linguagem, a comunicação entre as pessoas.

Se o conhecimento é construído a partir das interações sociais, podemos concluir que a noção de proficiência está relacionada à capacidade do usuário/examinando de interagir com alguém no ambiente social para construir sua comunicação. Dessa forma, a avaliação deve seguir a mesma linha de ação, proporcionando situações em que o examinando possa interagir com alguém, pensando nas condições de produção de seu texto: oral ou escrito.

Segundo Clark (2000:49) “o uso da linguagem é realmente uma forma de ação conjunta que é aquela ação levada a cabo por um grupo de pessoas agindo em coordenação uma com a outra”. Essa “ação em língua estrangeira pressupõe muito mais do que simplesmente dominar regras gramaticais ou funções comunicativas; significa, sobretudo, interagir dentro e fora da comunidade escolar”. O que é necessário para um falante ser proficiente é mais do que conhecer o sistema gramatical de uma língua. É saber usá-la adequadamente em determinadas situações e condições, como pedir informações, desculpar-se, negociar preços, questionar, responder. Podemos então dizer

que as práticas linguísticas precisam representar situações reais, que um teste precisa trazer práticas contextualizadas.

Em se tratando de avaliação em língua estrangeira, temos de pensar em nossos objetivos ao ensinar uma língua. Se pensarmos que todo ensino tem um objetivo, precisamos utilizar algum instrumento avaliativo para aferir se as metas foram atingidas, ou não.

No caso do exame de classificação que estamos analisando, a primeira interação do aluno é com uma máquina, que simula situações de uso da língua. O segundo passo é uma interação face a face com um interlocutor, que é um professor, e, desta forma, a construção de um texto, de forma que esse interlocutor possa entendê-lo, sendo que nessa perspectiva de ensino, a linguagem é usada como ação social.

Um conceito também muito importante para uma abordagem de ensino e conseqüentemente da avaliação, porém muito controverso é o conceito de competência linguístico-comunicativa, e, por conseqüência, proficiência, sendo que a partir deles outros termos foram cunhados.

À primeira vista, competência, competência linguística, competência comunicativa, competência linguístico-comunicativa e proficiência, parecem ser sinônimas e a única diferença parece ser uma questão terminológica, porém, a escolha ou rejeição de um termo revela divergências teóricas a respeito do conceito e a existência de uma variedade de modelos, dependendo sempre da visão de linguagem predominante na época em que o modelo foi proposto. Numa revisão de estudos que tratam especificamente do assunto Scaramucci (2000) mostra uma grande quantidade de sinônimos e definições para o termo, como: competência comunicativa, capacidade linguístico comunicativa, proficiência comunicativa, proficiência linguístico-comunicativa, competência transicional, são alguns deles.

1.3. O conceito de competência comunicativa

A palavra competência tem sido usada por vários autores e com conceitos diferentes. Para Perrenoud (1997:7) “competência é a capacidade de o sujeito mobilizar recursos visando abordar uma situação complexa.”

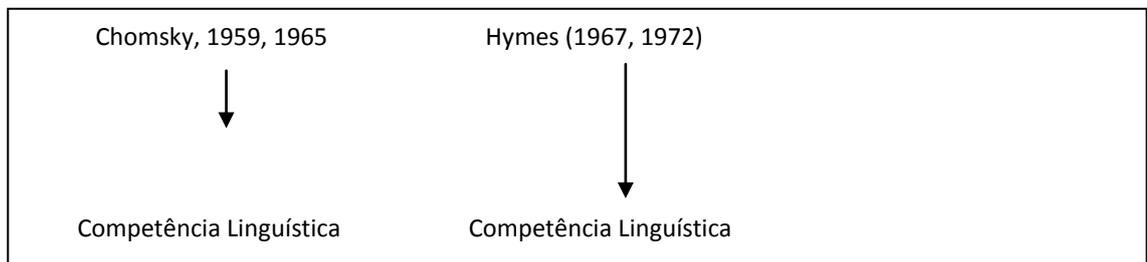
O conceito relaciona três aspectos importantes. O primeiro é relacionar competência como uma capacidade do sujeito “ser capaz de”. O segundo é ligado ao verbo *mobilizar* que significa movimentar com força interior, o que é diferente de deslocar, transferir de um lado para o outro. O terceiro está ligado à palavra *recursos*, porque a competência exige além dos recursos da cognição, isto é, do conhecimento intelectual, recursos do domínio emocional.

O conceito de competência (o conhecimento gramatical que o nativo possui) e *performance* (sua manifestação externa) desenvolvido por Chomsky (1965) foi contestado por Hymes (1972) que considerou a conceituação de Chomsky muito abstrata. Para Chomsky competência refere-se ao conhecimento tácito que o falante-ouvinte possui da estrutura da sua língua, e sugere que para que um indivíduo seja competente linguisticamente ele deve possuir conhecimento internalizado de estruturas gramaticais de seu sistema linguístico. Chomsky focalizou o falante nativo ideal, aquele que sabe a língua perfeitamente, e não é afetado por limitações da memória, lapsos ocasionais, inadequações lexicais e gramaticais.

Para Hymes (1972), competência é a capacidade de usar a língua, sendo que o mesmo implantou o termo *competência comunicativa*, que marca uma dissidência com relação ao conceito de Chomsky. Em seu modelo, Hymes focaliza o falante e ouvinte “real”, em interação, além de incluir conhecimento de outros aspectos da linguagem, que vão além da gramática, também inclui a capacidade de uso quanto a códigos de conduta social. Assim, seu conceito de competência comunicativa se refere a conhecimentos de regras abstratas de uma língua, regras essas exigidas para produzir correspondências de som/significado e a habilidade para usar tais correspondências entre som/significado e forma de modo social e culturalmente apropriado (vide quadro 1).

Taylor (1988) afirma que quando Hymes introduziu a noção de capacidade de uso no termo competência, fez com que ela perdesse seu significado, confundindo estado e processo. O autor propõe que o termo competência comunicativa (para ele, um conceito estático) seja abolido da literatura da área (cujo foco principal é o processo de ensino/aprendizagem) e em seu lugar seja utilizado o termo proficiência comunicativa.

Quadro 1 –Competência Comunicativa



Fonte: Celce-Murcia, 1995:700

Sandra Savignon, que desenvolveu uma das primeiras pesquisas sobre competência comunicativa na Universidade de Illinois em 1972, usou o termo *competência comunicativa* para caracterizar a habilidade de aprendizes de língua de interagirem com outros falantes a fim de produzirem significado. Em seu livro “*Communicative Competence: Theory and Classroom Practice*” (1983) a autora define comunicação como um processo contínuo de expressão, interpretação e negociação. As oportunidades para a comunicação são infinitas e incluem sistemas de sinais e símbolos dos quais a linguagem faz parte.

Para Savignon (1983:8), competência comunicativa é um intercâmbio dinâmico em que a competência linguística tem de ser adaptada a um insumo totalmente informacional, tanto linguístico como paralinguístico, de um ou mais interlocutores e apresenta as seguintes características:

Competência comunicativa é dinâmica e não estática. Ela depende da negociação de significado entre duas ou mais pessoas que compartilhem do mesmo sistema simbólico.

Competência comunicativa aplica-se tanto para a escrita como para a fala, assim como qualquer sistema simbólico.

Competência comunicativa está embasada num contexto específico. Comunicação acontece numa infinita variedade de situações, e o sucesso de determinados papéis depende do indivíduo compreender o contexto e de

experiência a priori similar. Requer fazer escolhas apropriadas de registros e estilos em termos de situação e dos outros participantes.

Existe uma diferença teórica entre competência e desempenho. Competência é definida como habilidade subjacente presumida, e desempenho como clara manifestação daquela habilidade. Competência é o que alguém *sabe*; desempenho é o que alguém *faz*. Somente o desempenho é observável, entretanto, e é somente através do desempenho é que competência pode ser desenvolvida, mantida e avaliada.

Competência comunicativa é relativa, não absoluta, e dependente da cooperação de todos os participantes que estão envolvidos. Faz sentido falarmos sobre graus de competência comunicativa (SAVIGNON, 1983:8)

Na década de 80 o conceito de competência comunicativa ganha grande destaque com o trabalho de Michael Canale e Merrill Swain.

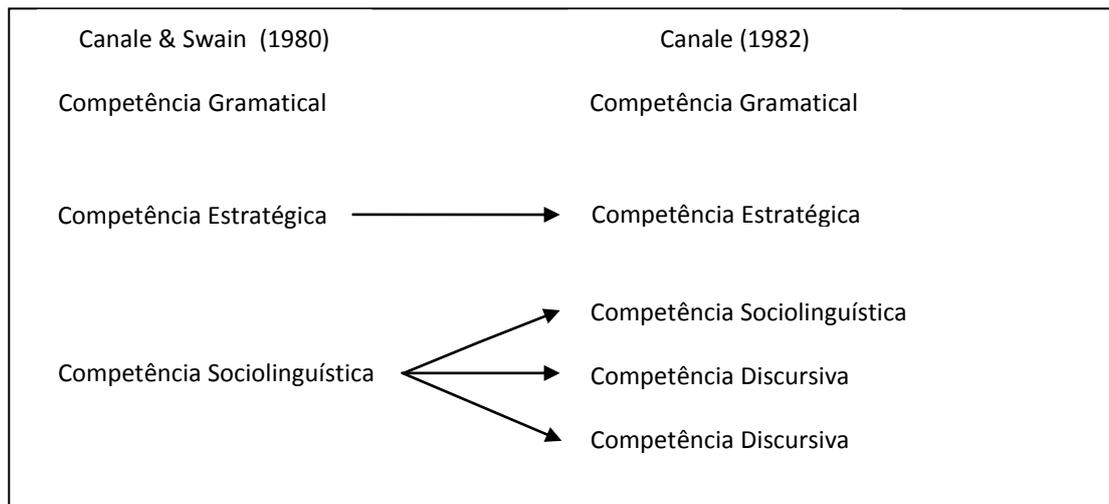
Para Canale a natureza da comunicação seria:

1. uma forma de interação social, sendo portanto adquirida e usada em interações sociais;
2. envolve um alto grau de imprevisibilidade e criatividade na forma e na mensagem;
3. acontece em contextos socioculturais e discursivos que fazem com que se utilize a linguagem adequada e também pistas da interpretação correta das mensagens;
4. é desenvolvida sob condições psicológicas limitadas e outras condições como memória, fadiga e distrações;
5. sempre tem um objetivo (por exemplo, estabelecer relações sociais, persuadir ou fazer promessas);
6. envolve autenticidade, sendo oposta a linguagem artificial de livros didáticos; é julgada como satisfatória ou não baseado em resultados concretos (CANALE, 1983:3-4).

Levando-se em conta esse conceito de comunicação, Canale & Swain (1980) no Canadá ofereceram uma especificação de três componentes, mais tarde modificado por Canale (1983), que propôs um modelo final que integrasse as habilidades – ouvir, falar, ler e escrever – em 4 componentes – competência gramatical, competência sociolinguística, competência discursiva e competência estratégica. A **competência gramatical ou linguística** se atém ao código linguístico (vocabulário, formação de palavras, formação de frases, significado literal de enunciados), **competência sociolinguística** (considera o papel do falante no contexto e a sua escolha de registro e estilo), **competência discursiva** (considera a questão da coesão e da coerência relevantes no determinado contexto) e a **competência estratégica** (considera que não há falantes e ouvintes ideais, sendo necessário, portanto, que se faça uso de estratégias de comunicação verbais ou não verbais para se compensar as quebras na comunicação), como podemos visualizar no quadro 2.

Para Canale & Swain (1980) as teorias sobre competência comunicativa de modo geral não dão a devida atenção às estratégias de comunicação que os falantes utilizam em situações de comunicação autêntica como hesitações, como se dirigir a estranhos, como evitar formas gramaticais que não dominam. Por essa razão, Canale (1983) acrescenta em sua teoria a quarta competência, a estratégica.

Quadro 2 – ELABORAÇÕES DA COMPETÊNCIA COMUNICATIVA



Fonte: Celce-Murcia, 1995:701

Observando-se a definição de Canale sobre competência comunicativa como “sistema de conhecimentos e habilidades exigidos para a comunicação”, podemos verificar que há a junção das palavras competência/desempenho. Em seu quadro teórico Canale descreve a competência comunicativa como parte de uma **comunicação real**: a competência comunicativa seria prejudicada por fatores ambientais externos e estados psicológicos como memória, fadiga, nervoso (Canale 1983:5) . É importante observar que tanto em Canale & Swain (1980) e como em Canale (1983) “competência comunicativa” é usada para se referir tanto a conhecimento quanto a habilidade em usar o conhecimento. A habilidade em se comunicar não é para esses pesquisadores uma parte da teoria sobre desempenho.

Infelizmente os autores não fizeram uma descrição de como esses componentes interagiriam entre si, deixando o assunto para uma pesquisa posterior, entretanto, Savignon (1983:45) propôs uma interação entre os quatro componentes do Modelo de Competência Comunicativa de Canale (1983). Ela desejou em seu diagrama (Figura 4)

mostrar que os quatro componentes interagem constantemente, que em comunicação não existe o fato de se começar por uma competência e as outras virem em seguida “como pérolas em um colar”(1983:45). Ela também sugeriu que as competências sociolinguística e estratégica permitem ao aprendiz uma comunicabilidade, mesmo antes da aquisição de qualquer competência gramatical.

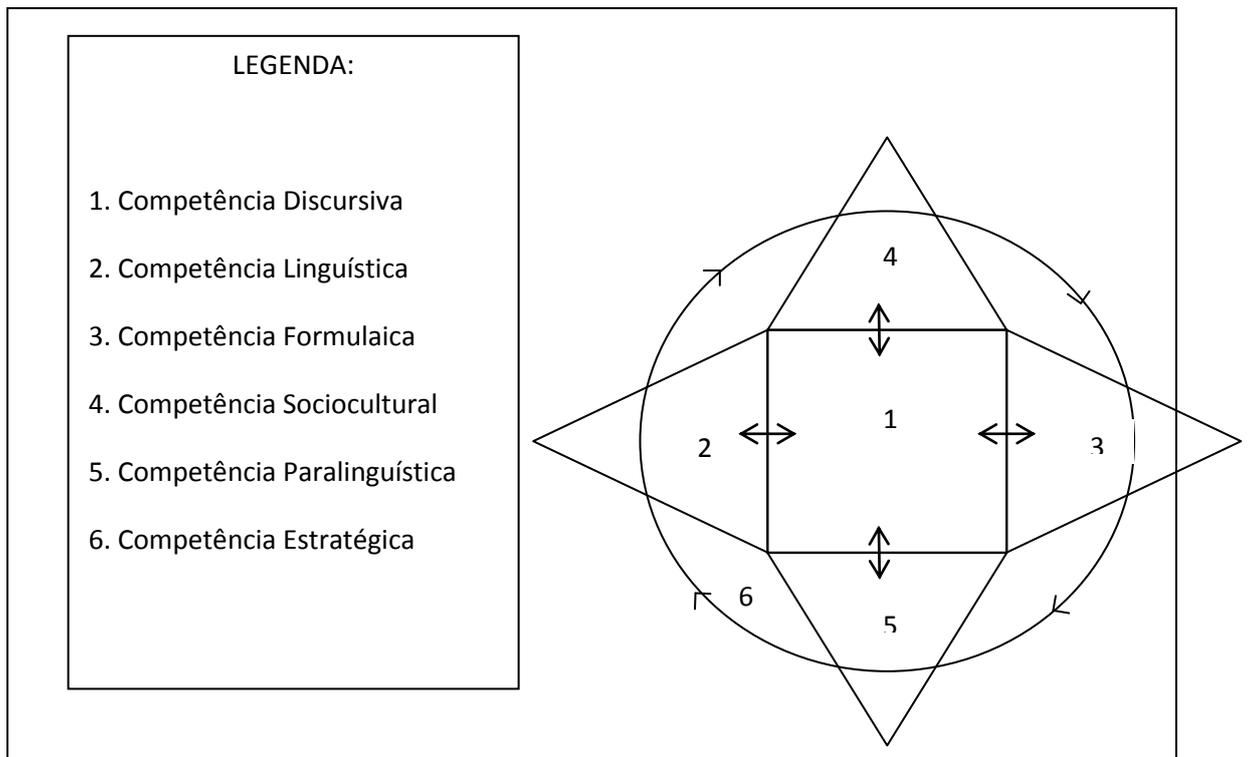


Figura 4: Interações entre os componentes da competência comunicativa
Fonte: Celce-Murcia, 1995:701

Savignon ainda afirmou que a competência estratégica está sempre presente, especialmente no começo da aprendizagem e aos poucos vai perdendo sua importância e cedendo espaço para as outras competências.

A questão das competências serem vistas separadamente, está relacionada com a cultura de aprender de muitos alunos e mesmo professores, que acham que primeiramente é necessário aprender a forma/gramática, apresentada de forma gradual e sequencial para depois usá-la em práticas de conversação, por exemplo.

Em 1982 Lyle Bachman & Adrian Palmer, baseados no modelo de Canale & Swain (1980) organizam seu próprio modelo e atribuem três competências distintas

para o que eles vieram a chamar de proficiência comunicativa (introduzem uma mudança na terminologia). A partir do desenvolvimento de uma bateria de testes que incluíam competência gramatical (morfologia e sintaxe), competência pragmática (vocabulário, coesão e organização) e competência sociolinguística (registro, naturalidade e referências culturais) sendo que o resultado de seus estudos sugere que os componentes que eles chamaram de gramatical e a competência pragmática estão intimamente associados entre si, enquanto os componentes referentes à competência sociolinguística são distintos.

Em 1990 Lyle Bachman propõe um modelo denominado “Competência de Linguagem” (vide quadro 3) . Esse modelo agrupa morfologia, sintaxe, vocabulário, coesão e organização sob uma competência: a *competência organizacional*. A *competência pragmática* desdobra-se em *competência ilocucionária* e *competência sociolinguística*, a última envolvendo os aspectos de polidez, formalidade, metáforas, registro e outros aspectos relativos à cultura.

Para Celce-Murcia (1995:701) o modelo de Bachman (1990) e Bachman & Palmer (1982) são modelos desenvolvidos com o propósito de pesquisa na área de avaliação e não exatamente um modelo para o ensino de línguas. Nesse sentido, a autora propõe um modelo mais dinâmico a partir de um desdobramento do modelo de Canale (1983). Em seu modelo (vide quadro 3) ela mantém três componentes que são:

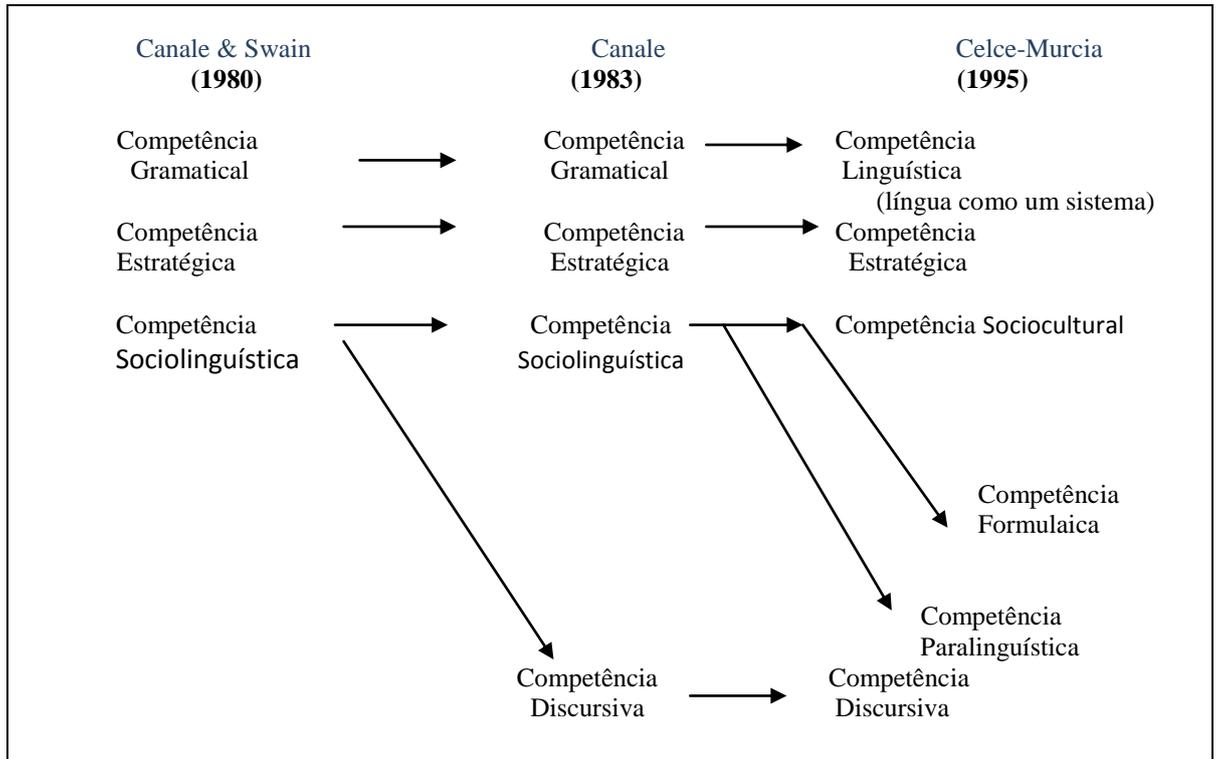
1. *Competência Linguística*, que compreende os elementos básicos da língua necessários para realizar a comunicação oral e escrita.
2. *Competência Estratégica*, que compreende as estratégias e os procedimentos relevantes à aprendizagem de línguas, ao processamento linguístico e à produção linguística. Ela ativa o conhecimento de outras competências e ajuda a compensar problemas de compreensão e deficiências de conhecimento na hora da comunicação.
3. *Competência Discursiva*, que compreende a seleção, a sequência e a organização das palavras, estruturas e sentenças dentro do discurso oral ou escrito. As estratégias de “top down”(conhecimento prévio, intenções) cruzam-se com as de “bottom-up”(léxico, recursos coesivos) e tornam possível ao falante/escritor a criação de textos coerentes.

A autora propõe além disso uma divisão na competência sociolinguística em três competências distintas: a competência sociocultural, competência formulaica e competência paralinguística, que são assim definidas:

1. *Competência Sociocultural*, que representa o conhecimento do aprendiz em se expressar adequadamente em contextos sociais e culturais relacionados à comunicação, de acordo com fatores pragmáticos relacionados a variações de uso da língua como polidez, formalidade, informalidade.
2. *Competência Paralinguística*, que compreende primeiramente os aspectos não verbais usados na interação face a face como gestos, expressões faciais e corporais e em seguida a distância física mantida entre os interlocutores, silêncio, murmúrios, etc.
3. *Competência Formulaica*, que está relacionada ao uso de padrões de sentenças e a ativação de palavras e estruturas discretas. Essa competência é um contraponto necessário à competência gramatical, pois inclui todas as colocações, clichês, palavras multifuncionais, frases lexicais associadas a atos de fala e outras funções pragmáticas.

O quadro 3 mostra o esquema da competência sociolinguística.

Quadro 3: Elaboraões da Competência Sociolinguística



Fonte: Celce-Murcia, 1995

Para Almeida Filho (1993) uma competência comunicativa poderia ser representada sintetizando-se as contribuições teóricas de autores como Chomsky (1965), Hymes (1979), Canale & Swain (1980) e Widdowson (1989). Para ele, comunicação (sempre de forma incompleta e conscientemente provisória) é uma forma de interação social propositada onde se dão demonstrações de apresentação pessoal, combinadas ou não com casos de (re) construção de conhecimento e troca de informações.

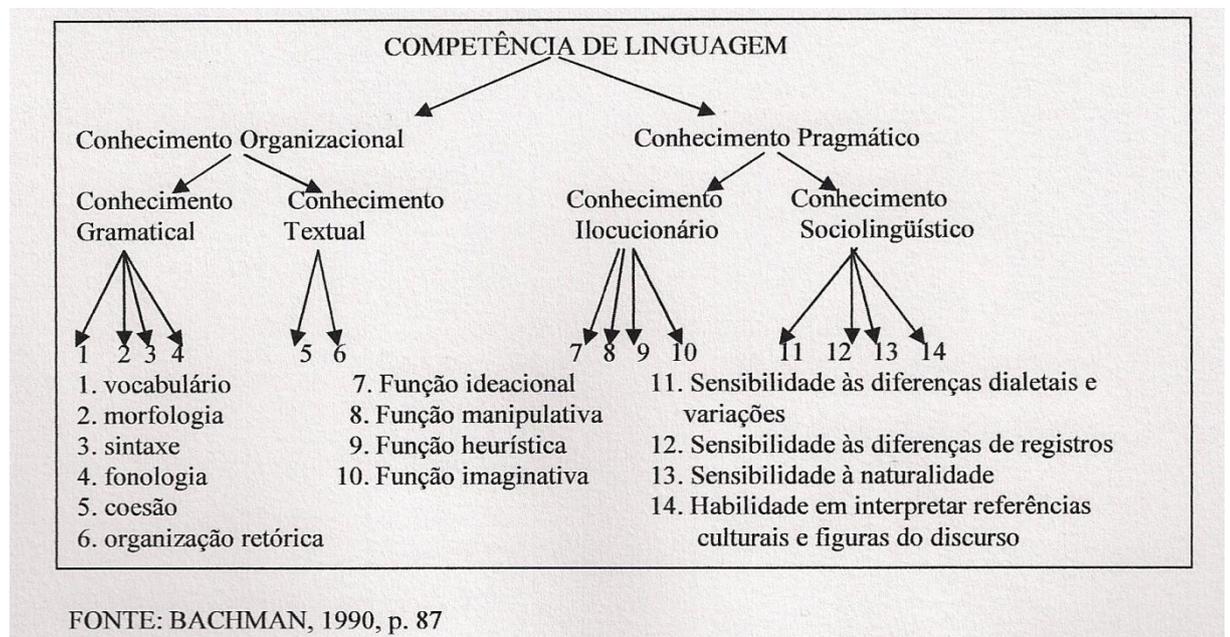
Nesse sentido, o autor concebe a comunicação verbal não como um simples processo linguístico, mas como um processo que exige do locutor outros conhecimentos culturais e sociais disponíveis na sua mente que vão interagir no processo da comunicação. Sendo assim, a comunicação envolve criatividade e imprevisibilidade tanto na forma quanto nos sentidos construídos no discurso.

Na década de 90, Bachman (1990) propôs um modelo, que era uma extensão dos modelos anteriores e o denominou “communicative language ability” ou a capacidade linguístico-comunicativa, constituída de três componentes: a competência linguística, dividida em competência organizacional e competência textual, sendo definida como

um conjunto específico de componentes que são utilizados na comunicação via língua. A competência estratégica, que caracteriza a capacidade mental de implementar os componentes da competência linguística no uso comunicativo e contextualizado da língua, que é subdividida em avaliação, planejamento e execução e os mecanismos psicofisiológicos, que se referem aos processos neurológicos e psicológicos envolvidos na execução real da língua, como luz, som, etc.

O quadro 4 mostra um esquema da competência de linguagem de Bachman.

QUADRO 4 – Competência de Linguagem



1.3.1. PROFICIÊNCIA

Apesar de o termo proficiência aparecer com frequência em pesquisas de ensino/aprendizagem de línguas, não tem sido adequadamente entendido, gerando confusões terminológicas bem como divergências quanto à sua conceituação. Procuraremos, a seguir, trazer algumas conceituações do termo esclarecendo seu uso técnico e não técnico.

Na discussão sobre uma definição para o termo proficiência no contexto de ensino /aprendizagem de línguas estrangeiras, Chastain (1989:48) afirma que “o termo parece fazer parte daquela categoria de palavras que são usadas frequentemente sem

uma atenção em relação ao seu significado exato. O resultado gera confusões que caracteriza sentidos diversos”.

Segundo Stern (1983), em um capítulo devotado à discussão do conceito, proficiência pode ser vista como o resultado da aprendizagem, uma meta definida em termos de objetivos ou padrões. Uma vez definida, a proficiência é relacionada a outras variáveis como contexto, aprendizagem, características, condições de aprendizagem e o próprio processo de aprendizagem.

Os objetivos e padrões definidores da proficiência podem ser usados como critérios para avaliar a proficiência como um fato empírico, ou seja, relativo ao desempenho efetivo de um indivíduo ou grupo de indivíduos.

Diferentes aprendizes em diferentes estágios de aprendizagem de uma segunda língua, a proficiência pode variar do grau zero ao grau semelhante a um falante nativo. O zero nunca é absoluto porque o aprendiz de uma segunda língua como falante de pelo menos uma língua, que seria sua língua materna, conhece a linguagem e como ela funciona. Uma competência completa, qualquer que seja sua definição, é raramente atingida por aprendizes de L2, e é amplamente reconhecido entre praticantes e teóricos que na maioria dos casos que seria um desperdício e talvez até mesmo indesejável tentar alcançá-la. No entanto ela é uma meta ideal para se ter em mente. A ‘competência’, proficiência’ ou ‘conhecimento de língua’ do falante nativo é um ponto de referência necessário para o conceito de proficiência em segunda língua usado na teoria do ensino de línguas. ⁴(STERN,1983:341).

O dicionário *Collins COBUILD* (apud Scaramucci, 2000) define proficiência como capacidade (*ability*) ou habilidade (*skill*). “Ele é proficiente em Inglês” ou “José é proficiente em Matemática”. São exemplos citados de onde resgatamos que o termo proficiência, normalmente relacionado à linguagem, possa ser usado em qualquer área de conhecimento no sentido de conhecimento, domínio, controle, capacidade, habilidade.

⁴ No original: “Among different learners at different stages of learning a second language competence or proficiency ranges from a zero to native-like proficiency. The zero is not absolute because the second language learner as speaker of at least one other language, his first language, knows language and how it functions. Complete competence, whatever its definition, is hardly ever reached by second language learners, and it is widely acknowledged among practitioners and theorists that in most cases it would be wasteful and perhaps even undesirable to attempt to reach it. Nevertheless, it forms an ideal goal to keep in mind. The native speaker’s ‘competence’, ‘proficiency’, or ‘knowledge of the language’ is a necessary point of reference for the second language proficiency concept used in language teaching theory”. (Stern, 1983:341)

Scaramucci (2000) afirma que o termo que se refere ao desempenho de um falante de língua estrangeira com referência ao controle do falante nativo ‘ideal’ é uma acepção ampla, não técnica de proficiência e pode ser considerado sinônimo de fluência. Entretanto, no contexto da avaliação de língua estrangeira, seu uso técnico refere-se ao controle operacional da língua estrangeira, sendo um conceito relativo, que pressupõe vários níveis de proficiências e inclui o propósito da situação de uso futuro da língua.

Para Scaramucci (2000), o termo proficiência tem sido usado de forma não técnica (ou ampla) e técnica (restrita). A utilização do termo ‘proficiência’ do modo como foi empregada na definição do dicionário acima mencionada é típico da utilização não técnica do termo, normalmente utilizada por leigos. Nesse uso, poderíamos dizer que há um ponto de corte entre proficiência e não proficiência e todos os níveis inferiores a esse seriam classificados como não proficientes. O conceito, nesse caso, parece ser absoluto, ou seja, “é proficiente ou não é proficiente”.

Entretanto, o uso mais restrito ou técnico utilizado no contexto de avaliação de segunda língua/língua estrangeira, com o sentido de domínio, funcionamento ou controle operacional da língua em questão, considera uma gradação da proficiência, em que todos os níveis seriam considerados proficientes em graus diferentes e de acordo com a especificidade da situação de uso da língua.

Contudo, este controle operacional da língua está relacionado a certificados reconhecidos internacionalmente, não tendo, portanto, uma ligação direta com a situação de aprendizagem. Sendo assim, outra forma de se definir proficiência seria contrastá-la com o termo rendimento. Enquanto proficiência refere-se ao uso futuro da língua, não se aliando a currículo nenhum, mas sim a um construto teórico, sem ligações com o passado, rendimento é definido como específico e local e tem a função de descrever o programa de ensino.

Taylor (1988:16) propôs que deixássemos a terminologia utilizada por Chomsky (competência X desempenho) e utilizássemos o termo proficiência para designar ‘a habilidade de fazer uso da competência’. Stern (1983) também expôs a ideia de usarmos *proficiência* como substituto para *competência*, especialmente quando se refere à competência do falante não nativo em contextos de ensino aprendizagem de língua estrangeira. Com o uso do termo – proficiência – não haveria a confusão terminológica

sobre a inclusão ou não da noção de habilidade nesse conceito e evitar-se-ia problemas que normalmente aparecem com o uso do termo “competência comunicativa”. Sendo assim, se utilizarmos o termo proficiência como um termo intermediário entre competência e desempenho, poderíamos incluir nele a noção de habilidade.

O que se faz necessário é aceitarmos a definição de competência definida por Chomsky como um termo técnico e a renomeação feita por Bachman (1990) de “competência comunicativa” para “capacidade linguístico-comunicativa” (*communicative language ability*) que incluiria proficiência linguística e proficiência comunicativa, sendo que a primeira teria relação com capacidade de uso da língua e a última incluiria o conhecimento de mundo e as estratégias necessárias para se aplicar à proficiência linguística em situações de uso.

Desta forma seriam distinguidos claramente dois conceitos que poderiam ser descritos separadamente:

- a) competência
- b) capacidade linguístico-comunicativa, que seria subdividida em:
 - proficiência linguística
 - proficiência comunicativa

Assim como a visão de linguagem, o conceito de competência comunicativa tem muita importância neste trabalho, pois, a partir da visão de linguagem do professor, ou de um grupo de especialistas em avaliação, um determinado teste apresentará certas características.

Além disso, com a crescente demanda por cursos de inglês adaptados às necessidades específicas de profissionais inseridos em várias áreas profissionais, novos estudos começaram a surgir no estudo da língua inglesa. Tradicionalmente, o objetivo da Linguística era o de descrever as regras de uso da língua. No entanto, os novos estudos desviaram a atenção para a definição de maneiras pelas quais a linguagem é realmente utilizada na comunicação real (Widdowson, 1978). Esse fenômeno, juntamente com os novos desenvolvimentos da Psicologia Educacional, contribuiu para o estudo de Inglês para Fins Específicos.

1.4. Foco na linguagem - Inglês Geral ou Inglês para Negócios?

Uma das áreas do Inglês para Fins Específicos (English for Specific Purposes, doravante ESP) é o Inglês para Negócios, e deve ser visto no contexto geral de ESP, pois compartilha os elementos importantes da análise de necessidades, plano de ensino, design, projeto do curso, seleção e desenvolvimento de materiais, comuns a todos os campos de trabalho em ESP. No entanto, Ellis & Johnson (2002) afirmam que:

Inglês para Negócios difere de outras variedades em ESP, pois frequentemente há uma mistura de conteúdo específico relacionado a uma profissão em particular ou indústria, e conteúdo geral relacionado à habilidade geral de comunicar mais efetivamente, especialmente em situações de negócios (ELLIS & JOHNSON, 1994: 3).⁵

Alguns autores como Dudley-Evans & St. John (1998) e Robinson (1991) concebem Inglês para Fins Específicos como uma área maior e que dela há ramificação de duas grandes áreas: Inglês para Fins Acadêmicos (EAP - English for Academic Purposes) e Inglês para Fins Ocupacionais (EOP – English for Occupational Purposes).

Os mesmos autores apontam ainda que Inglês para Negócios seria um ramo de Inglês para Fins Específicos, sendo que o mesmo “ exige uma pesquisa cuidadosa e a concepção de materiais e atividades pedagógicas para um grupo definido de aprendizes adultos em um contexto de aprendizagem específica” (Dudley-Evans & St. Johns, 1991), e que “é projetado para atender necessidades específicas do aprendiz” (Stevens, 1988).

Ainda, podemos considerar “Inglês para Negócios” como um termo guarda-chuva, usado para abarcar cursos de inglês geral com um léxico e uma gramática adequada para a comunicação de negócios, além de também descrever um trabalho específico, tanto para um grupo homogêneo, cuidadosamente selecionado de uma empresa, ou pequenos grupos em atividades programadas para desenvolver habilidades específicas como negociações, escrever e-mails, memorandos, relatórios, leitura de gráficos e apresentações.

⁵ ‘No original: “Business English differs from other varieties of ESP in that it is often a mix of specific content relating to a particular job area or industry, and general content relating to general ability to communicate more effectively, especially in the business situations”. (Ellis & Johnson, 1994: 3).

Com a globalização, o livre comércio entre as nações tornou-se uma prática comum. Inúmeras empresas ampliaram seus negócios em outros países, abrindo filiais e contratando funcionários, cuja primeira língua não é o Inglês. A forma de comunicação, seja oral ou escrita, entre a matriz e as filiais precisa ser numa língua comum, sendo o idioma de domínio o Inglês, por ser a língua franca utilizada para esse propósito.

A partir desse cenário, alguns cursos começaram a ser implementados para atender às necessidades específicas das pessoas que necessitavam do Inglês não só para a situação de comunicação geral, mas uma comunicação relacionada a habilidades específicas apropriadas ao seu contexto de atuação.

Ainda, de acordo com Ellis & Johnson (1994: 7) o que caracteriza a linguagem de negócios é o senso de propósito com que as pessoas a usam e o sucesso desse uso é medido pelos resultados de transações de negócios e eventos alcançados.

Alguns profissionais necessitam de inglês para comunicação oral e escrita, com uma grande interação no mundo dos negócios; outros somente precisam desenvolvimento de uma habilidade, para um fim específico, como a leitura de uma planilha, por exemplo.

Essas necessidades levaram ao desenvolvimento de aulas visando o domínio de um vocabulário que normalmente não é trabalhado em cursos de Inglês Geral, fazendo com que novos materiais surgissem no mercado educacional para contemplar essas novas exigências em termos de comunicação em língua estrangeira.

Alguns autores como Ellis & Johnson (1994: 4-5) fazem uma diferenciação entre Inglês para Fins Específicos e para Fins Gerais. Segundo esses autores, no final dos anos 60 e início dos 70, o que diferenciava Inglês para Fins Específicos e para Fins Gerais era o vocabulário específico, sem se preocuparem com o uso dessa linguagem em situações de vida real.

De acordo com Ellis & Johnson (1994:35), uma das principais características de Inglês para Negócios é capacitar o aprendiz para que ele se torne operacionalmente eficaz. Para pessoas da área de negócios, a prioridade é ser capaz de compreender e transmitir sua mensagem, e para a maioria dos aprendizes de Inglês de negócios, muitos dos refinamentos da língua são muito simples e irrelevantes. Para pessoas que trabalham em período integral, o tempo é muito restrito e a aquisição de conhecimento para seu próprio desenvolvimento profissional fica adiada. Segundo os mesmos autores, é

possível dizer que as pessoas que trabalham na área de negócios e têm a necessidade de aprender Inglês para Negócios precisam:

- confiança e fluência;
- habilidades para estruturar e organizar informações;
- precisão suficiente na linguagem;
- estratégias para acompanhar um discurso;
- estratégias para esclarecer e verificar informações que não ficaram claras;
- velocidade de reação às declarações de outras pessoas;
- pronúncia clara;
- consciência da linguagem apropriada e comportamento para as culturas e situações na qual ele atuará;

A partir de meados dos anos 70 e início dos anos 80, o foco esteve na funcionalidade da língua, o uso da língua em um contexto real de interação como pedir esclarecimentos, dar sugestões, fazer recomendações, etc.

Conforme explicam Ellis & Johnson (1994), nos anos 90, algumas variedades de Inglês para Negócios surgiram, sendo que uma delas foi o surgimento de cursos projetados para pessoas geralmente adultas, que se preparavam para ingressar no mercado de trabalho, para aqueles que já estavam (com muita ou pouca experiência) nesse mercado, para alunos que ainda estavam estudando para entrar no mundo dos negócios (pré-serviço), ou ainda para pessoas que já tinham muita experiência (em serviço), porém não tinham o conhecimento da língua inglesa para a comunicação dentro da área. A partir dessa época, a visão de linguagem já não era mais sobre o conhecimento teórico da linguagem, mas sobre a importância do uso da linguagem.

A língua inglesa é utilizada para comunicação em boa parte do planeta. Se um diretor de uma empresa necessita sair de seu país em uma viagem de negócios, ele terá de utilizar como meio de transporte o avião, ele precisará de um táxi e terá de fazer o *check in* em um hotel. Ele provavelmente terá de trocar dinheiro de sua moeda natal pela moeda do país que estará visitando e ao se alimentar terá de entender a linguagem

utilizada no cardápio e fazer o pedido da refeição. Se ele tiver problemas, precisará saber como reclamar. Em todas essas situações, ele utilizará as mesmas estruturas linguísticas de alguém que possa estar na mesma situação em que ele. Porém, se em vez de uma viagem de negócios, ele estiver fazendo uma viagem de férias, será que as exigências linguísticas desse executivo são diferentes de um falante de inglês geral?

Segundo Ellis & Johnson (1994: 4-5), nesse caso o ponto chave é o vocabulário específico. Entretanto, podemos ainda questionar se além do vocabulário, há diferenças entre inglês para negócios e inglês geral. O termo “inglês para negócios” foi cunhado para orientar estudantes de inglês que utilizam a língua estrangeira em um ambiente de negócios. A diferença para muitos parece residir apenas nas atividades desempenhadas.

No entanto, existem algumas diferenças, embora pequenas. O vocabulário pode ser diferente, mas não inédito. Uma pessoa em sua casa geralmente atende ao telefone dizendo “Alô”. No entanto, ao atender ao telefone em um escritório é diferente. Uma forma de comunicação seria falar além do “Alô”, o nome da pessoa, como, por exemplo: “Alô, ou Bom-dia, Eliana Souza falando. Posso ajudar?”. As relações interpessoais são diferentes bem como o nível de formalidade.

O quadro a seguir, baseado em Zhu (2008), apresenta algumas diferenças de vocabulário entre Inglês Geral e Inglês para Negócios.

Quadro 5 - Diferentes significados, a partir de uma mesma palavra, no Inglês Geral e Inglês para Negócios (baseado em: Wenzhong Zhu, (2008) – modificado).

Palavra em Inglês	Significado em Inglês Geral	Significado em Inglês para Negócios
minute	minuto	minuta
liquidate	pagar, liquidar	acabar com um negócio
principal	diretor, principal	capital, rendimento financeiro
composition	composição, redação	pagamento de dívidas
turnover	olhar para outra direção	faturamento

Fonte: Wenzhong Zhu, 2008 – modificado

Em um estudo desenvolvido por (Rosenberg 2004), a autora procura mostrar de que forma as estruturas linguísticas utilizadas no Inglês Geral coincidem com as do Inglês para Negócios. A autora fez uma pesquisa junto a um grupo de professores de

Inglês, questionando quais seriam as semelhanças e as diferenças entre cursos de Inglês Geral e de Inglês para Negócios. Os diagramas a seguir mostram o resultado das respostas fornecidas pelos professores.

Inglês para Negócios	Inglês Geral
<ul style="list-style-type: none"> * vocabulário especializado * motivação em aprender relacionada ao trabalho * ensinar negociação e apresentação <ul style="list-style-type: none"> * técnicas * alunos com objetivos bem definidos <ul style="list-style-type: none"> * alunos mais sérios * correspondência comercial 	<ul style="list-style-type: none"> * atividades de tempo livre * mais liberdade em não cumprir o plano de aula <ul style="list-style-type: none"> * mais tempo para jogos * atmosfera mais relaxante <ul style="list-style-type: none"> * música * literatura * habilidades gerais de escrita <ul style="list-style-type: none"> * eventos sociais

O que o Inglês Geral e Inglês para Negócios têm em comum:
<ul style="list-style-type: none"> * gramática e funções * vocabulário geral * ansiedade em relação à capacidade * estruturas linguísticas usadas no dia a dia <ul style="list-style-type: none"> * conversas breves * vocabulário para viagem * inglês para situações cotidianas <ul style="list-style-type: none"> * eventos comuns

Depois de fazer esse inventário, Rosemberg (2004:36) constatou que os professores ficaram surpresos com a grande quantidade de sobreposições que existem e concluiu que mesmo assim temos uma tendência em separar Inglês para Negócios e Inglês Geral. Uma das razões seria que ao se formarem, os professores são capacitados para trabalhar como professores de Inglês, entendendo que isso seria Inglês Geral, e acreditam que para ensinar Inglês para Negócios é necessária uma especialização.

Além disso, muitas das atividades utilizadas nas aulas de Inglês Geral não são aplicadas nas aulas de Inglês para Negócios e vice-versa. A autora cita, como exemplo, que as atividades mais dinâmicas e divertidas normalmente não são transferidas para as aulas de Inglês para Negócios.

Apesar de não concordamos com essa visão, acreditamos que muitas iniciativas que são retiradas diretamente do mundo dos negócios e são usadas com sucesso para ensinar alunos em aulas de Inglês para Negócios poderiam ser transferidas para as aulas

de Inglês Geral, dando aos alunos uma nova perspectiva de habilidades comunicativas com o propósito explícito de tornar o aluno proficiente em Inglês.

Graddol (2006), em seu livro “The Future of English”, relata que haverá um aumento no uso da língua inglesa por grupos de estudantes, trabalhadores imigrantes, e das populações nas nações que estão emergindo como os poderes econômicos futuros. O que todos esses grupos compartilham é o contexto no qual eles desejam utilizar a língua inglesa: que seja direcionado para sua área de trabalho. Portanto, é possível que nos próximos anos haja uma demanda crescente por testes de Inglês para áreas específicas, como os que já existem para a área de Direito, ou em contextos de Inglês para Negócios.

Nas três últimas décadas várias publicações foram lançadas com foco na linguagem de Inglês para Negócios. Além disso, algumas certificações internacionais que contemplam essa área de estudos também se fazem presentes no mercado. A Universidade de Cambridge oferece vários tipos de exames com foco na área de negócios, que são: BEC – Business English Certificate em três níveis: BEC Preliminary, Vantage e Higher; além de BULLATS – The Business Language Testing Service e ICFE – International Certificate in Financial English. Além desses são oferecidos também o TOEIC – Test of English for International Communication, produzido pela ETS – Educational Testing Service, que tem foco na área de Negócios, o EFB - English for Business, oferecido pela Câmara de Comércio Inglesa e o FTBE (*Foundation Certificate for Teachers of Business English*), certificado para professores de inglês para negócios.

Desenvolver testes para fins específicos como os descritos acima acaba por levantar questões que são distintas em relação aos testes de inglês geral. Bachman e Palmer (1996) afirmam que “qualquer exame de línguas deve ser desenvolvido com um objetivo específico, em um grupo específico de candidatos e em um domínio de uso específico”. Eles definem domínio como “a situação ou o contexto no qual o candidato utilizará a língua fora de um contexto de exame”. Douglas (2000:19) descreve os testes de Inglês para Fins Específicos⁶ como:

⁶ No original: “...one in which test content and methods are derived from an analysis of a specific purpose target language use situation, so that test tasks and content are authentically representative of tasks in the target situation, allowing for an interaction between the test taker’s language ability and specific purpose content knowledge, on the

Um exame para fins específicos é aquele em que o conteúdo do exame e os métodos são derivados de uma análise de uso específico da língua-alvo em determinadas situações, de forma que as tarefas do exame e o conteúdo sejam representantes autênticos de tarefas na situação-alvo, permitindo por um lado, uma interação entre a habilidade linguística do candidato e o conhecimento do conteúdo de uma situação específica do candidato em usar a língua em domínios específicos (DOUGLAS, 2000:19).

Douglas (2000) ainda identifica dois aspectos que normalmente distinguem testes para fins específicos de testes de inglês geral. A primeira diferença está na autenticidade das tarefas, ou seja, as tarefas de um teste compartilham de determinadas características que um examinando pode encontrar numa situação real de uso da língua, de forma que quanto mais próximo o teste apresentar tarefas que estejam presentes na vida real, melhor refletirá o desempenho do examinando para uma situação de uso da língua alvo.

A segunda característica que caracteriza um teste de Inglês para Fins Específicos é a interação entre o conhecimento linguístico e o conhecimento de um conteúdo específico. Este conteúdo é, talvez, a diferença mais importante entre um teste de inglês para fins específicos e um teste de inglês geral. Douglas (2000) aponta que Testes para Fins Específicos são ‘precisos’ uma vez que eles terão características lexicais, semânticas, sintáticas e fonológicas que os distinguem de um teste de línguas em contextos gerais.

De acordo com Davies (2001), não se pode caracterizar que a linguagem para fins específicos é radicalmente diferente da linguagem da vida cotidiana, (embora, como podemos observar no conceito proposto por Douglas (2000), certamente há diferenças quantificáveis entre um teste de inglês geral e de inglês para fins específicos).

Douglas (2000) sugere o uso de escalas desenvolvidas para a avaliação de inglês geral que poderiam ser usadas também para a avaliação de uma linguagem para fins específicos, com a possibilidade de uma expectativa diferente. Isso poderia resultar, por exemplo, que um avaliador encontrasse um critério destinado a focar o vocabulário e depois aplicar esse critério de um propósito específico a ser testado.

A lógica do argumento proposto por Douglas (2000) resultaria em uma incapacidade em avaliar a linguagem sem também avaliar a habilidade do examinando

one hand, and the test tasks on the other. Such a test allows us to make inferences about a test taker's capacity to use language in the specific purpose domain. (Douglas 2000:19)

em desempenhar tarefas dentro de um contexto específico a ser testado. Pode-se destacar aqui que McNamara (1996) descreve esse fenômeno como uma visão ‘forte’ de testes de desempenho, nos quais a tarefa é o foco do teste, com a linguagem sendo vista apenas como uma das habilidades necessárias para se executar com sucesso uma determinada tarefa.

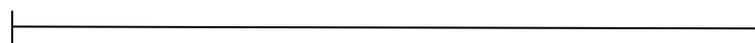
Em termos de testes de inglês para negócios, e, na verdade, de testes para qualquer outro contexto específico, seria extremamente difícil para o elaborador de itens de teste criar um instrumento válido, no qual a definição do construto não distinguisse claramente a linguagem e as habilidades específicas para a realização das tarefas.

A partir do que foi argumentado anteriormente, podemos estabelecer dois pontos de vista diferentes, que oferecem uma expansão que não é incompatível com a definição de testes de inglês para fins específicos apresentada por Douglas (2000:19):

Ponto de Vista 1: como todos os testes são de certa forma “específicos”, o ideal seria pensar em testes de línguas dentro de um contínuo de especificidade, como um teste de inglês geral (como o *Cambridge Proficiency in English*) para um teste mais específico, como o Teste para Controladores de Voo. Assim, esse ponto de vista pode ser explicitado graficamente como:

Propósito Geral

Propósito muito Específico



Ponto de Vista 2: testes muito específicos têm uma tendência a serem muito pobres em termos de generalizações, enquanto que o oposto ocorre em testes não específicos ou gerais. Não há uma escolha binária em operação aqui, e se for aceito que testes podem ser desenvolvidos ao longo de um contínuo específico, então, temos que um teste que esteja posicionado em qualquer lugar que não seja nas extremidades desse contínuo terá o potencial para ser mais ou menos geral.

Com base na análise acima, poderíamos conceber um teste designado para fins específicos apenas na medida em que é colocado dentro do contexto de uma área de emprego/carreira, que utilize um vocabulário específico e que enfoque aspectos específicos dentro do contexto daquela carreira. Caso contrário, o teste seria sempre considerado como um teste geral.

No capítulo 2 será apresentado um breve panorama sobre os estudos da avaliação em línguas estrangeiras, traçando paralelos com concepções de linguagem e também com as abordagens de ensino/aprendizagem de língua estrangeira, para em seguida explorarmos os conceitos de validade, confiabilidade, praticidade, autenticidade, efeito retroativo e impacto, elementos fundamentais para a análise a ser desenvolvida neste estudo.

CAPÍTULO 2 - AVALIAÇÃO E ENSINO/APRENDIZAGEM DE LÍNGUAS

A avaliação na área de Línguas pode ser feita de várias formas e através de diferentes métodos. Para cada período temos abordagens vinculadas aos diversos paradigmas que vão se contextualizando através do tempo. Toda avaliação está associada a um processo interpretativo de dados quantitativos ou qualitativos, pressupondo-se um juízo de valor que tem por finalidade diagnosticar os objetivos propostos no processo de ensino aprendizagem.

O objetivo deste capítulo é primeiramente apresentar um histórico sobre as avaliações de línguas, para em seguida explorar conceitos como validade, confiabilidade, praticidade, autenticidade, efeito retroativo e impacto. Finalmente apresentaremos o conceito de validade e validação de testes.

2.1. Histórico da Avaliação de Línguas

Os testes de línguas evoluíram muito nos últimos sessenta anos. Isso aconteceu ao mesmo tempo em que novos conceitos sobre linguagem e novas abordagens de ensino foram aparecendo. Assim, é importante traçar-se paralelos entre as concepções de linguagem e as concepções de avaliação. Segundo Baker (1989:2), há uma defasagem entre as concepções de linguagem, abordagens de ensino e avaliação.

Mudanças de ênfase no ensino de línguas estrangeiras provocaram inevitavelmente mudanças nos testes de LE. No entanto, métodos e teorias de testes têm sido mais relutantes às mudanças do que teorias de metodologias de ensino e desenvolvimento de cursos. Isto ocorre principalmente porque testes de línguas estrangeiras modernas são baseados em princípios que, como a velha concepção estruturalista, estão pautados na descrição da língua independente de qualquer uso particular que se faça dela. (BAKER, 1989:2).⁷

⁷ No original: 'These shifts in emphasis in language teaching have inevitably had consequences for language testing. Testing techniques and theories, however, have been rather more resistant to change than theories about methodology and course design. This is principally because modern language testing is based on principles which, like the old 'structural' syllabuses, take as their starting point a description of the language independent of any particular use of it.'

Spolsky (1975) em seu artigo seminal “*Language testing: art or science*” divide a trajetória de ensino/aprendizagem/avaliação em três períodos: *o pré-científico, o psicométrico-estruturalista e o sociolinguístico-integrativo*. Morrow (1979), em seu artigo ‘Communicative language testing: revolution or evolution?’, renomeia os estágios de Spolsky em : ‘*Garden of Eden*’ (*Jardim do Éden*) ‘*Vale of Tears*’ (*Vale das Lágrimas*) e ‘*The Promised Land*’ (*A Terra Prometida*).

2.1.1. O período pré-científico ou *The Garden of Eden* (O Jardim do Éden)

Spolsky primeiro identifica o período *pré-científico*, que tem seu início a partir do século XVIII e se estende até o início da década de 50. Como a denominação sugere, a avaliação no período pré-científico não era embasada em conhecimentos científicos, caracterizando-se pelo foco na forma, na tradução e nos exames orais, com seus resultados determinados de forma impressionista. O ensino de línguas não era considerado uma disciplina independente, a avaliação de línguas não era objeto de estudos mais específicos.

2.1.2. Avaliação dos anos 40 aos 60: o período do psicométrico-estruturalista ou *The Vale of Tears* (O Vale de Lágrimas)

A partir da Segunda Guerra Mundial a necessidade de comunicação entre os países fez com que pesquisas na área de ensino-aprendizagem evoluíssem. Surgiu um método de ensino que focalizava a oralidade, o Método Audiolingual, sendo seu maior expoente Lado (1961). Assim, surge também uma nova fase na avaliação, identificada por Spolsky como o período “*psicométrico-estruturalista*”. Os métodos e a terminologia dos testes psicológicos foram adotados no campo dos testes de línguas e o termo mais significativo dessa época é ‘psicométrico’, derivado da palavra psicometria que quer dizer registro e medida de fenômenos psíquicos por meio de métodos experimentais padronizados. Nessa fase dois aspectos dos testes psicológicos são incorporados: as questões de múltipla escolha, verdadeiro ou falso, e preencher lacunas, testes estes de respostas fechadas, onde somente uma única resposta é a correta ou desejada. Tais testes prometiam objetividade na correção, de forma a conduzir a uma maior confiabilidade.

Há, nesse caso, a criação de várias escalas para medir o desempenho do candidato que é avaliado, com a intenção de demonstrar a confiabilidade do teste.

Esse tipo de avaliação recebeu o nome de testes de itens isolados (*discrete point tests*) que se fundamentaram no pressuposto de que a linguagem pode ser fragmentada para viabilizar seu ensino. Os testes de elementos isolados dividem a linguagem em categorias fonológicas, lexicais e sintáticas, mensurando o conhecimento de tais categorias sem levar em conta as correlações entre essas categorias no discurso. Sua preocupação com a forma se sobrepunha à preocupação com o significado e com a comunicação.

Lado (1961:22-23) em seu livro “Language Testing”, define o conceito desse tipo de avaliação:

A teoria da avaliação de línguas pressupõe que a linguagem é um sistema de hábitos de comunicação. Esses hábitos permitem que o comunicador preste atenção consciente ao significado geral que ele comunica ou compreende. Esses hábitos envolvem questões de forma, sentido e distribuição em diversos níveis estruturais: os níveis da sentença, da cláusula, da frase, da palavra, do morfema e do fonema. Dentre esses níveis estão estruturas de modificação, sequência e partes de sentenças. Abaixo deles estão os hábitos de articulação, tipos de sílabas e colocações. Associados a eles, e às vezes, como parte deles, estão os padrões de entoação, acento e ritmo. (LADO, 1961:22-23).⁸

Apesar de Lado admitir que as habilidades ocorreriam de forma integrada na vida real e que as estruturas gramaticais estivessem a serviço das quatro habilidades – ler, falar, ouvir e escrever – tais habilidades, como diz Lado (1961: 25) poderiam ser testadas isoladamente.

(...) os elementos da linguagem discutidos podem ser proveitosamente estudados e descritos, e testados como universos separados, embora eles nunca ocorram separados na linguagem. Eles são integrados em habilidades totais de fala, escuta, leitura e escrita.⁹ (LADO, 1961: 25)

⁸. No original: “The theory of language testing assumes that language is a system of habits of communication. These habits permit the communication to give his conscious attention to the over – all meaning he is conveying or perceiving. These habits involve matters of form, meaning, and distribution at several levels of structure, namely those of the sentence, clause, phrase, word, morpheme and phoneme. Within these levels are structures of modification, sequence, parts of sentences. Below them are habits of articulation, syllable type, and collocations. Associated with them and sometimes part of them are patterns of intonation, stress and rhythm”.

Segundo a citação acima, ser proficiente significa ter automatizado as estruturas linguísticas da língua alvo, sabendo cada um dos seus elementos isoladamente.

A principal vantagem dessa abordagem era a objetividade na correção e a possibilidade de se quantificar os dados com facilidade, direcionando a uma maior confiabilidade nos testes, conforme aponta Weir (1990: 2) “são facilmente quantificáveis; permitem uma vasta cobertura de itens; notas baseadas em testes de itens isolados são confiáveis, pois são objetivas. Porém, não são necessariamente válidas.”

Entretanto Oller (1979) descreve as desvantagens desse tipo de testes:

A análise de itens isolados necessariamente quebra os elementos da língua em pedaços e tenta ensiná-los ou testá-los separadamente com pouca ou nenhuma maneira pela qual esses elementos interagem num contexto de comunicação maior. O que os torna ineficientes como base para o ensino ou testes de línguas é que as propriedades cruciais da língua são perdidas quando seus elementos são separados. O fato é que em qualquer sistema onde as partes interagem para produzir propriedades e qualidades organizacionais se restringem e se tornam propriedades cruciais do sistema que não podem simplesmente ser achadas em partes separadas.¹⁰ (OLLER, 1979:212)

Um fator positivo dessa fase é que foi uma tentativa de dar à avaliação em língua estrangeira um caráter científico, sendo que até hoje utilizamos os testes de itens isolados da mesma forma que foram concebidos nas décadas de 40 a 60. Muitos dos exames de entrada para universidades brasileiras e exames de proficiência reconhecidos internacionalmente como TOEFL, TOEIC e Cambridge apresentam várias questões de itens isolados que têm o objetivo de testar apenas o uso do léxico e itens gramaticais, além de avaliar as habilidades isoladamente.

Um dos motivos do uso dos testes de itens isolados é a praticidade e a confiabilidade da correção (McNamara,1996), pois é possível dar um retorno rápido ao candidato de seu score, sem que haja vários pedidos de revisão das provas, pois a grade

⁹ No original: “(...) the elements of the language discussed can be profitably studied and described, and tested as separate universes, yet they never occur separately in language. They are integrated in total skills of speaking, listening, reading and writing.”

¹⁰ No original: “Discrete point analysis necessarily breaks the elements of language apart and tries to teach them (or test them) separately with little or no attention to the way those elements interact in a larger context of communication. What makes it ineffective as a basis for teaching or testing languages is that crucial properties of language are lost when its elements are separated. The fact is that in any system where the parts interact to produce properties and qualities organizational constraints themselves become crucial properties of the system, which simply cannot be found in the part separately.”

de correção contempla apenas uma resposta. Entretanto, a visão de fragmentação da língua começou a ser questionada, na medida em que começou a ser vista como um todo e não mais como uma somatória de partes. A competência linguística passou a ser vista como uma unidade de capacidade interacional que não podia ser dividida em partes.

2.1.3. O período psicolinguístico-sociolinguístico ou *The Promised Land* (A Terra Prometida)

Na década de 70 novas concepções de linguagem surgem e os testes de elementos isolados já não ofereciam mais uma medida suficiente para se avaliar a competência linguística e a avaliação direcionou-se para o período denominado de *sociolinguístico integrativo* em que a linguagem é vista a partir de um código contextualizado, usado em situações reais de comunicação. Essa visão, que se desenvolveu na década de 80, deu origem ao chamado movimento comunicativo. Nessa nova visão, o termo criado e mais debatido foi, e ainda é, competência, e por consequência, proficiência, sendo que a partir deles outros termos foram cunhados, conforme apresentado na seção anterior.

Nesse período, a linguagem era avaliada em termos da habilidade do usuário de uma língua em lidar com funções linguísticas, bem como usar a língua de maneira específica em situações definidas, visão essa que privilegiaria o desempenho do sujeito que faz uso da língua. A avaliação da linguagem oral, dentro dessa abordagem, possibilita o ensino integrativo (e não de elementos isolados), evitando a compartimentalização da linguagem. Exames desse tipo segundo Scaramucci (1999) apresentam como base comunicativa as seguintes características:

- a) ênfase na comunicação, através de situações reais de interação, sendo que os instrumentos para avaliação buscam coletar amostras de tarefas que apresentam o uso real da língua.
- b) conteúdos autênticos ou contextualizados, (sendo que o termo **autêntico** tem sido usado em oposição à linguagem artificial utilizada em livros didáticos e exames).
- c) os critérios de avaliação utilizados são holísticos, tomando como base o desempenho global do aluno dentro do desempenho do objetivo da comunicação a ser atingido, que serve como parâmetro para correção da prova.
- d) o resultado da avaliação é expresso em faixas de proficiência, com base em descritores de competência e desempenho do candidato.

Bachman (1991) aponta quatro diferenças entre os testes comunicativos e os testes de itens isolados:

Primeiramente tais testes [comunicativos] criam um tipo de lacuna de informação [information gap], o que faz com que o examinando processe informações complementares através do uso de múltiplas fontes de insumo. Pode ser pedido ao candidato que desempenhe uma tarefa escrita que é baseada tanto em uma pequena palestra gravada quanto em um texto sobre o mesmo tópico. A segunda característica dos testes comunicativos diz respeito à interdependência de tarefas: uma é necessária para o desenvolvimento da outra. Em terceiro lugar os testes comunicativos podem ser caracterizados pela integração das tarefas e o conteúdo estabelecido a partir de um contexto. Finalmente, os testes comunicativos visam a medir uma quantidade maior de habilidades da linguagem, tais como coesão, coerência e adequação sociolinguística, dentre outros aspectos, que não faziam parte da testagem indireta de pontos isolados.¹¹ (BACHMANN, 1991:678).

Morrow (1979: 150-51) argumenta que para que um teste possa ser considerado um teste comunicativo ele deve ter as seguintes características:

Os testes deverão ser referenciados em critério para um desempenho operacional de um conjunto de tarefas com linguagem autêntica. Em outras palavras, tais testes mostrarão se (ou o quão bem) o candidato pode desempenhar um conjunto de atividades específicas.

Será primordial estabelecer sua própria validade como uma mensuração das operações que tais testes afirmam avaliar. Portanto, validades de conteúdo, de construto e o efeito retroativo serão importantes, mas a validade concomitante com testes existentes não será necessariamente significativa.

Os testes terão uma ênfase na avaliação qualitativa e não diretamente quantitativa. Poderá ser necessário converter os resultados em escores numéricos, mas o foco na análise qualitativa será sempre evidenciado.

Confiabilidade, claramente importante, será subordinada à validade. (MORROW, 1979: 150-151).

¹¹ No original: "First such tests create an information gap, requiring test takers to process complementary information through the use of multiple sources of input. Test takers, for example, might be required to perform a writing task that is based on input from both a short recorded lecture and a reading passage on the same topic. A second characteristic is that of task dependency, with tasks in one section of the test building upon the content of earlier sections, including the test taker's answers in those sections. Third, communicative tests can be characterized by their integration of test tasks and content within a given domain of discourse. Finally communicative tests attempt to measure a much broader range of language abilities - including knowledge of cohesion, functions and sociolinguistic appropriateness - than did earlier tests, which tended to focus on the formal aspects of language - grammar, vocabulary, and pronunciation."

Almeida Filho (1998) expõe que as avaliações comunicativas têm cinco características teóricas específicas:

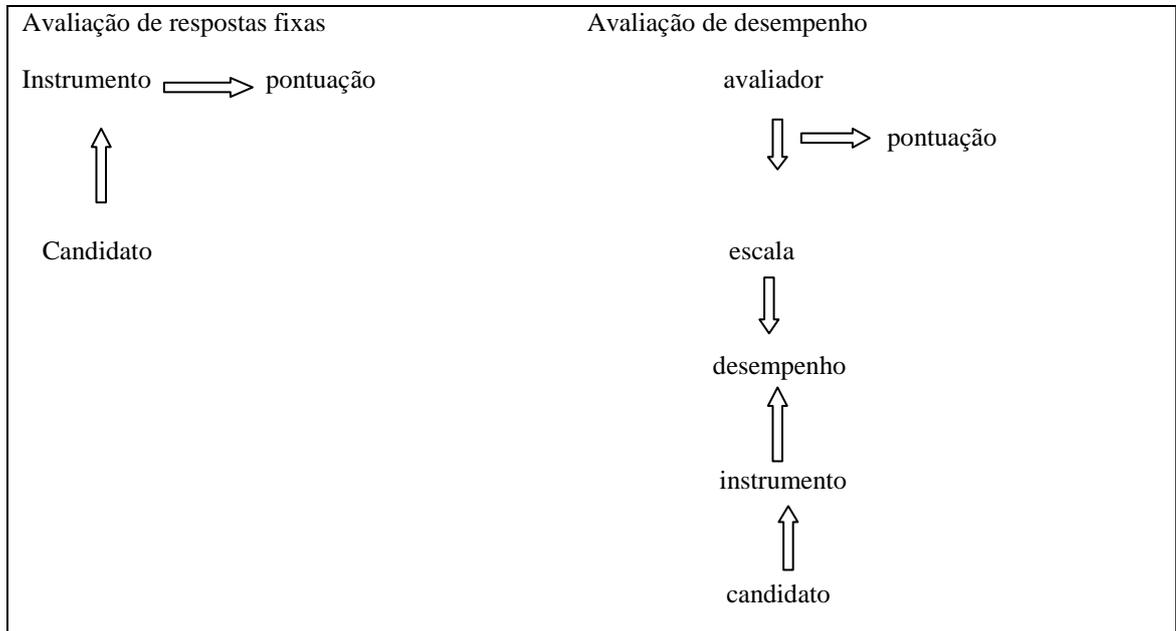
- (a) os instrumentos de avaliação devem ser capazes de produzir ação comunicativa concreta (“O que o aprendiz já pode fazer?”);
- (b) o contexto linguístico e situacional para o desempenho dos alunos deve ser comunicativamente autêntico ou verossímil (“Com que fim ou devido a que necessidade comunicativa a linguagem será criada?”);
- (c) enquanto amostras de desempenho as tarefas e recortes comunicativos devem permitir pelo menos alguma imprevisibilidade no insumo (na produção) em tempo real;
- (d) a avaliação será mais qualitativa do que numérica (quantitativa);
- (e) as validades de conteúdo e externa (plausibilidade) das amostras de desempenho terão precedência sobre a confiabilidade que possam possuir. (ALMEIDA FILHO, 1998:68-69).

Podemos inferir que as características apresentadas acima relacionadas a testes comunicativos estão afinadas com as características apresentadas por Canale & Swain (1980) e Canale (1983), uma vez que envolvem elementos das competências gramaticais, sociolinguísticas, discursivas e estratégicas, pois há uma preocupação em simular situações nas quais o aluno possa usar a língua na vida real, em situações que ocorrem no dia a dia. Segundo Scaramucci (2000), uma das contribuições do modelo de Canale (1983) foi chamar a atenção para a necessidade de integração das competências linguísticas e ressaltar a importância das estratégias comunicativas para o conceito de proficiência.

Todas as fases de avaliação expostas neste trabalho apresentam algum tipo de problema. Podem ter problemas de validade, confiabilidade ou praticidade. Os testes da era psicométrico-estruturalista tendem a ser muito confiáveis, pois são testes objetivos, fáceis de corrigir e os escores raramente sofrem alterações, principalmente porque são corrigidos normalmente por máquinas – leitores ópticos, não havendo interferência de vários corretores. Entretanto a validade de construto é questionável.

Os testes da era psicolinguística sociolinguística têm uma maior validade de construto por utilizarem instrumentos de avaliação em que se tenta mostrar a habilidade linguística do candidato em um contexto ou situação particular. Mas têm uma frágil confiabilidade por serem testes subjetivos e sua correção vulnerável, por ser feita por mais de um avaliador.

McNamara (1996: 120) aponta a diferença entre os testes de base comunicativa e os de base estruturalista, conforme podemos ver na figura a seguir.

Figura 5: Testes tradicionais e testes de desempenho

Fonte: McNamara, 1996:120

De acordo com a figura, podemos perceber que nos testes tradicionais a pontuação do candidato está vinculada apenas a uma resposta única, derivada de uma tarefa objetiva do teste, enquanto que nos testes de desempenho há uma escala de descritores que necessitam ser interpretadas pelos avaliadores de acordo com as respostas do candidato.

2.2. O processo avaliativo e seus componentes

Bachman e Palmer (1996) apontam que três estágios são necessários para o desenvolvimento de um teste: planejamento, operacionalização e administração.

O primeiro estágio, que é o planejamento, está relacionado à definição do que se quer medir, as especificações que vão nortear o desenvolvimento de um teste, de modo que torná-se necessária a definição do construto. Segundo a definição de Ebel e Frisbie (1991:108 apud Scaramucci 2009:32) :

O termo *construto* refere-se a um construto psicológico, uma conceitualização teórica sobre um aspecto do comportamento humano que não pode ser medida ou observada diretamente. Exemplos de construtos são inteligência, motivação para o rendimento, ansiedade, rendimento, atitude, dominância e compreensão em leitura.

O segundo estágio, que é a operacionalização do construto, é o desenvolvimento de itens que sejam referenciados ao construto, normalmente definidos no primeiro estágio; e o terceiro estágio, a administração, que envolve as fases de revisão do próprio teste, o *feedback* resultante, com o intuito de coletar informações necessárias para coletar evidências sobre as habilidades e conhecimentos relevantes para o construto.

Um dos grandes questionamentos que fazemos ao elaborar um instrumento avaliativo é saber se aquele teste é eficiente, se ele realmente avalia o que queremos avaliar, se é um texto “adequado” ou “inadequado”. Entretanto, não existe “o melhor método avaliativo”, mas aquele que se adequará melhor às necessidades do contexto em que ele será aplicado (Bachman & Palmer, 1996).

Brown (2001) afirma que para sabermos se um teste está de acordo com os objetivos propostos para uma avaliação, devemos nos concentrar nos critérios considerados clássicos na literatura de avaliação, que são: praticidade, confiabilidade e validade, além de autenticidade, efeito retroativo e impacto.

Bachman & Palmer (1996:17) afirmam que os três últimos critérios é que definem a qualidade de um teste, pois é por meio deles que podemos avaliar se o instrumento utilizado para avaliação é coerente e adequado. Os autores propõem, então, um modelo para a avaliação de um teste, que compreenderia a soma das seguintes qualidades: confiabilidade + validade de construto + autenticidade + interatividade + impacto + praticidade.

A seguir vamos expor esses critérios fazendo uso de conceituações e ilustrações de como eles podem se apresentar em instrumentos avaliativos.

2.2.1. Interatividade

Bachman & Palmer (1996:25) apontam que a interatividade (*interactiveness*) está relacionada ao tipo de envolvimento das características individuais do examinando ao realizar as tarefas do teste, incluindo-se nessas características o conhecimento da língua, competências estratégicas, conhecimento do tópico e de que forma o

examinando responde emocionalmente ao teste. Segundo os autores, a interatividade é uma qualidade inerente a qualquer tarefa para que o examinando demonstre seu desempenho na língua alvo. Entretanto, ela será diferente dependendo do tipo de habilidade exigida para a realização de cada tarefa.

Toda tarefa dependerá dos objetivos específicos de cada curso, sendo que podemos diferenciar a interatividade em sala de aula, com o auxílio do professor e de colegas para a realização de tarefas e a interatividade em um teste, uma vez que ela poderá se restringir às instruções para a realização do teste ou mesmo das tarefas.

No caso do teste que foi estudado neste projeto de pesquisa, e que será instrumento para validação de seu construto, conteúdo e confiabilidade, conforme veremos mais adiante, todas as instruções para a sua realização aparecem na tela do computador, não havendo qualquer tipo de interatividade com colegas, instrutores ou mesmo com o professor. Assim, espera-se que as instruções para a realização do teste sejam claras, para que não haja qualquer tipo de interferência negativa no desempenho do candidato por falta de entendimento das instruções do teste ou mesmo das tarefas.

2.2.2. Praticidade

Uma das características que se junta à validade e confiabilidade é que um bom teste precisa também ser prático, sendo empregado dentro de limites de tempo e orçamento, proporcionando um alto grau de custo-benefício. É importante que um teste seja fácil de administrar, interpretar e apresentar os escores. A praticidade depende de alguns fatores como tempo disponível para a realização do teste, inclusive de sua duração, número de candidatos, disponibilidade de avaliadores, corretores, funcionários para aplicar o teste, gerenciá-lo e disponibilidade financeira.

Um teste que exija entrevista individual, se o número de candidatos for pequeno, poderá ser considerado prático. Entretanto para um grande número de candidatos, um único avaliador será considerado impraticável.

Nos últimos anos o uso de testes eletrônicos para fins de classificação tem se tornado um método cada vez mais atraente e eficiente de avaliação. Por todo o mundo, muitas faculdades e universidades têm mudado a sua forma de utilizar o modelo convencional, com papel e caneta/lápis para o formato de testes eletrônicos.

O TOEFL, que é um teste de proficiência, hoje é oferecido no modelo eletrônico, mas também ainda no modelo convencional - papel e caneta/lápis. Com o avanço da tecnologia digital, torna-se vital explorar as capacidades e o potencial de uso dessa tecnologia na área de avaliação. Um olhar mais atento indica que algumas instituições de ensino de Inglês como Língua Estrangeira utilizam softwares prontos, feitos por editoras renomadas como, por exemplo, Cambridge, Macmillan e Oxford, ou ainda programas para computador geradores de provas, através do qual o professor escolhe a quantidade de questões a partir de um banco de dados e o formato que cada questão deve ter, como por exemplo, itens de múltipla escolha, preenchimento de lacunas, leitura e compreensão de textos, atividades para recepção oral, entre outras.

Muitos desses testes têm um caráter comercial e utilizam um banco de dados bastante genérico de modo a satisfazer um interessado que procura um teste pronto para ser aplicado. Uma das vantagens para a escolha desse tipo de material é pela praticidade que ele oferece, pois pode ser instalado em qualquer computador ou ser baixado diretamente da internet a qualquer momento e ser utilizado imediatamente, inclusive contando com a correção automática e divulgação de resultados logo após o término da prova. Apesar das vantagens apresentadas, outros profissionais ou instituições optam por desenvolver seus próprios testes de acordo com suas necessidades de ensino, em contextos específicos. Nesses casos há a necessidade de vários tipos de profissionais para a confecção da prova: desde professores proficientes na língua-alvo, até *designers* instrucionais, programadores e técnicos da área de Tecnologia da Informação.

Na aplicação de testes, aqueles mediados pelo computador vêm ganhando popularidade em relação aos tradicionais testes aplicados em formato de papel e caneta/lápis devido a várias vantagens potenciais que ele oferece, como pontuação imediata e comunicação de resultados em tempo real, flexibilidade no agendamento do teste, oportunidade de incluir formatos inovadores, possíveis devido a utilização da tecnologia, com redução de custos de produção e administração mais eficaz.

Além disso, testes mediados pelo computador têm o potencial para oferecer tarefas que reflitam com mais autenticidade as habilidades que possam ser avaliadas muito próximas do mundo real. Exames que tenham, por exemplo, um componente de

fala, podem revelar o comportamento do candidato em detalhes, podem captar os procedimentos que o candidato utiliza para responder os itens. Assim, podem ir além do modelo respostas certas e respostas erradas, de forma a revelar a natureza do entendimento ou a falta de entendimento do candidato, Podendo até trazer contribuições para o aprendizado da língua.

Na próxima seção apresentaremos o resultado de algumas pesquisas sobre a utilização de testes mediados pelo computador e testes que ainda utilizam o formato tradicional de papel e lápis/caneta, a fim de demonstrar os pontos positivos e negativos desse primeiro tipo de teste, uma prática que está se consolidando na avaliação de línguas.

2.2.2.1.Comparação entre exames mediados pelo computador e exames em papel e caneta/lápis: objetivos, abordagens e estado da arte

Com o advento da revolução digital, examinadores da área de línguas procuram utilizar avançadas tecnologias computacionais a fim de satisfazer a crescente necessidade de ferramentas para medir as habilidades comunicativas em língua estrangeira com o máximo de precisão e eficiência.

Graças aos esforços de especialistas nas áreas de linguística computacional, engenharia da computação e psicometria, examinadores da área de línguas desenvolveram testes mediados pelo computador.

A tecnologia possibilita que diversas tarefas sejam feitas com mais eficiência, conveniência e rapidez. Esse é um argumento razoável para se adotar abordagens tecnologicamente avançadas para testes. Além disso, a tecnologia pode permitir que determinadas tarefas sejam feitas de forma diferente e melhor.

Os computadores têm desempenhado um papel fundamental em testes de línguas a partir de 1935. Desde os primeiros dispositivos de pontuação até os testes adaptados ao computador (*Computer Adaptive Tests (CATs)*), os computadores desempenham um papel importante na elaboração e administração de testes, pontuação, análise de dados, geração de relatórios e divulgação de resultados. A partir do momento em que as

máquinas e os programas de computador ficaram mais sofisticados, uma série de questões relacionadas à pesquisa, prática e ética surgiram.

Esta seção tem o objetivo de descrever o papel dos computadores em testes de línguas e discutir algumas questões complexas sobre a utilização dos mesmos na primeira década do século XXI.

O primeiro registro do uso de um computador em testes de línguas remonta a 1935 quando o modelo 805 da IBM tornou-se comercialmente disponível. Essa foi a primeira máquina capaz de corrigir testes objetivos, e passou a ser imediatamente colocada em uso nos Estados Unidos como forma de reduzir gastos, além da trabalhosa tarefa de corrigir milhares de testes realizados todos os anos.

A necessidade prática de avaliar um grande número de alunos de forma barata e eficiente por meio de testes objetivos, e o advento da tecnologia para conseguir isso sempre foi uma das preocupações dos estudiosos e especialistas em avaliação. Entre as décadas de 1920 e 1930 houve um grande debate sobre o valor desse novo tipo de teste e as limitações em termos do que poderia ser testado usando testes de múltipla escolha. O debate ainda continua, embora aqueles que ponderem os prós e os contras em relação ao tipo de testes objetivos frequentemente não estão conscientes do fato de que os argumentos apresentados hoje são muito semelhantes àqueles realizados 80 anos atrás.

O que é importante reconhecer é que testes construídos a partir de itens objetivos não são especificamente mais confiáveis do que qualquer outro tipo de item ou tarefa. Lado (1961) apud Fulcher (1997) fazendo uma distinção importante entre confiabilidade e “pontuação”, pondera que é possível ter uma prova mal construída, não confiável, com itens de múltipla escolha facilmente corrigíveis, em relação a um teste bem construído, confiável, mas que consome muito tempo como um teste de fala. A razão pela qual os testes de múltipla escolha têm sido usados há mais de um século não está relacionada à confiabilidade inata do tipo de item, mas ao simples fato de que pode ser corrigido facilmente por uma máquina.

Desde que o primeiro livro sobre testes mediados por computador foi publicado (Holtzman, 1970), as pesquisas e mesmo a literatura sobre o assunto têm crescido exponencialmente. Desde a primeira máquina para correção de testes objetivos de

1935, até os computadores pessoais e *smartphones*, os pesquisadores da área de avaliação estão continuamente tentando colocar a tecnologia em benefício do aprimoramento dos testes.

Talvez a mudança mais marcante que o computador e a tecnologia da informação trouxeram para a avaliação de línguas seja a possibilidade de se fazer vários testes diferentes em várias partes do mundo ao mesmo tempo, fornecer o resultado desses testes para o examinando imediatamente, 24 (vinte e quatro) horas por dia, 7 (sete) dias por semana. No entanto, juntamente com esses avanços, aparecem preocupações e problemas potenciais. Como o uso de vídeo em uma avaliação de recepção oral afeta o desempenho do examinando? Ouvir uma gravação com vídeo envolve as mesmas habilidades que ouvir uma gravação sem vídeo? De que forma poderiam ser interpretados os resultados de uma prova de produção escrita corrigida por seres humanos e corrigida por um programa de computador? Alguns questionamentos são feitos, muitos deles ainda sem resposta.

Assim, há muitas controvérsias em relação à avaliação de línguas mediada por meios eletrônicos e muitos debates e pesquisas estão em andamento a respeito das vantagens e desvantagens desse tipo de sistema para a avaliação de línguas. No entanto, as vantagens parecem superar as desvantagens (Chapelle & Douglas, 2006).

Críticos argumentam que a avaliação conduzida por meios eletrônicos ainda trata de um elemento conservador, uma vez que os itens do teste são limitados a determinados tipos de programação que as máquinas exigem. De todos os tipos possíveis, o mais popular é o teste de múltipla escolha. Os defensores apontam que os testes por meios eletrônicos permitem a incorporação de recursos multimídia e que a apresentação dos resultados de forma imediata significa uma mais valia para o examinando.

Alderson (1999:39-43) delineou a capacidade do computador em explorar detalhes relevantes como:

O computador tem a habilidade de medir o tempo. O tempo que um aprendiz leva para completar uma tarefa, ou até mesmo o tempo que leva para fazer partes diferentes de uma tarefa pode ser medido, controlado e gravado pelo computador.

O computador tem a habilidade de gravar informações sobre a rotina do aprendiz durante o teste.

O computador pode apresentar informações em uma infinidade de formas diferentes.

O computador pode fornecer acesso fácil e rápido a uma variedade de tipos de informações diferentes.

O computador pode ser ligado a outros equipamentos. Isso permite diferentes tipos de insumo e apresentações.

O computador pode fazer uso de regras de linguagem:

- a) Em um nível relativamente simples, o computador pode verificar erros de ortografia que o aluno cometeu ao redigir um texto;
- b) Analisadores de diferentes graus de sofisticação podem ser usados não somente para verificar erros sintáticos no texto do aluno, mas também para produzir testes comunicativos.” (ALDERSON, 1999: 39-43).

Mazzeo & Harvey (1988) forneceram uma das primeiras revisões sobre o tema relacionado à avaliação intermediada por meios eletrônicos, que incluía cerca de 30 (trinta) estudos sobre a comparabilidade de uma série de testes, com foco na inteligência, aptidão, personalidade e rendimento. Esse estudo revelou evidências mistas sobre a comparabilidade entre testes mediados pelo computador e testes feitos em papel e caneta/lápis. Eles descobriram que o modo de teste parecia não ter efeito sobre o desempenho dos candidatos.

Mais recentemente, com o interesse crescente na realização de testes mediados pelo computador em escolas de ensino fundamental nos Estados Unidos, alguns estudos comparativos com foco na sua aplicação foram conduzidos. Um estudo desenvolvido pela Texas Educational Agency (TEA) foi desenvolvido comparando várias disciplinas do curriculum escolar americano (Matemática, Inglês, incluindo as habilidades de leitura e escrita, Ciências e Estudos Sociais). Em cada área de estudo foram encontradas discrepâncias entre as conclusões de alguns estudos empíricos. Alguns estudos indicaram que a avaliação intermediada pelo computador é mais difícil em relação à avaliação feita em papel e caneta/lápis e vice-versa. (p.ex. Choi & Tinkler, 2002), enquanto outros estudos demonstraram que a avaliação intermediada pelo computador e a avaliação feita em papel e caneta/lápis são passíveis de comparação, apesar de não serem similares.

Paek (2005) também observou que testes que possuíam textos muito longos pareciam trazer mais dificuldades para os examinandos que utilizavam o computador do que no formato em papel, e a autora sugere que isso pode estar relacionado às diferentes estratégias de leitura empregadas pelos examinandos. Por exemplo, no papel o examinando pode grifar certas informações que são relevantes ou utilizar técnicas

visuais para se lembrar de uma informação em particular que está localizada na página. Comenta-se que o uso da rolagem de página para se navegar eletronicamente em oposição a se virar a página inibe o uso dessa “aprendizagem visual” pois o senso de localização na página se perde (Paek, 2005:18). Até certo ponto o argumento parece ter sua veracidade, entretanto parece que quanto mais as pessoas se tornam adaptadas a ler através de meios eletrônicos, mais as estratégias de leitura começam a ser criadas para esses meios.

Na verdade, Paek (2005) verificou que a rolagem de página é um método muito mais popular de navegar por um texto do que o de mudar de página, através do comando “enter”. Entretanto isso também pode variar de indivíduo para indivíduo, pois tudo depende da forma como a pessoa se sente mais a vontade ao lidar com o computador.

Para a avaliação da escrita há certa preocupação não somente pela forma como os candidatos podem interagir com o texto no computador, mas também de que maneira o examinador pode ser influenciado na correção do texto ao visualizar a apresentação do texto. A legibilidade dos textos escritos no computador em relação aos escritos manualmente pode provocar efeitos mais positivos ou negativos na percepção do examinador em relação à escrita. Entretanto, há uma falta de consenso entre alguns autores em relação a essas diferenças, pois alguns argumentam que textos desorganizados, com várias rasuras, respostas ilegíveis, são corrigidos com mais rigor (Shaw, 2003).

Muitos pesquisadores têm tentado fornecer uma justificativa para as diferenças observadas. Leeson (2006) cita alguns fatores potenciais relacionados ao uso do computador a partir de duas perspectivas: o fator participantes e o fator tecnologia.

O fator participantes inclui características demográficas dos examinandos tais como gênero e etnia, processos cognitivos, habilidade dos examinandos, familiaridade com computadores e a ansiedade ao interagir com a máquina. Os problemas tecnológicos envolvem a interface legibilidade (tamanho e resolução da tela, comprimento da linha e tipo de fonte) e a interface usuário (como o rolamento da página para leitura, permissão para revisão de itens e como é a apresentação dos itens da prova – um item de cada vez ou vários ao mesmo tempo).

Entre os fatos relacionados às características dos participantes, a familiaridade com computadores foi bastante examinada por vários pesquisadores. Entretanto os resultados são ainda inconclusivos. Alguns estudos, no entanto, parecem indicar que a maior familiaridade com o computador é um sinal de melhor desempenho do examinando, sugerindo que examinandos que tenham menos experiência com computadores terão uma pontuação mais baixa em testes mediados pelo computador (p.ex. Goldenburg & Pedulla, 2002; Pomplun & Custer, 2005, Pomplun, Ritchie & Custer, 2006, Bennett, Braswell, Oranje, Sandene, Kaplan, & Yan, 2008).

Em relação ao gênero e etnia, alguns estudos afirmam que não há diferenças em relação a esses fatores, principalmente na forma de aplicação dos testes e sua administração, tanto por meios eletrônicos quanto na versão tradicional em papel e caneta/lápis (p.ex. Bennett et al., 2008; Clariana & Wallace, 2002).

Quanto aos participantes e a interface com o computador, uma comparação entre testes mediados pelo computador e avaliações feitas em papel e caneta/lápis pode estar relacionada à rapidez em se fazer as provas. A análise feita por Pomplun, Frey & Becker (2002) revelou que a versão computadorizada de um teste de classificação em leitura e compreensão de textos resultou em uma pontuação média maior em relação ao teste feito em papel e caneta/lápis. Os autores acreditam que a diferença está no procedimento de resposta, pois clicar com o mouse pode ter uma vantagem para o examinando devido a rapidez para percorrer os itens do teste, e que a utilização da caneta/lápis para preencher as respostas em um formulário de respostas levaria mais tempo.

Embora o resultado de vários estudos comparativos não seja consistente, parece haver uma tendência de que os testes mediados pelo computador sejam similares em resultados aos testes em papel e caneta/lápis. Com as constantes mudanças na tecnologia de computadores e a ampla acessibilidade pela população a esse meio, a tendência é de queda em relação às pessoas que não têm acesso aos computadores. Além disso, pesquisas recentes indicam que os examinandos testados se sentem confortáveis em fazer testes no computador e têm uma preferência por esse tipo de teste em relação aos de papel e caneta/lápis (Way, Davis & Fitzpatrick, 2006). Entretanto os resultados ainda são preliminares, de forma que a oferta em se fazer os testes nos dois tipos de versão ainda é possível, e novas pesquisas precisam ser feitas para garantir que

os examinandos não sejam tratados de forma injusta devido ao meio em que o teste é realizado.

Maximizar a comparabilidade e minimizar os efeitos específicos de formato é a melhor garantia de equidade; no entanto, isso é ignorar o fato de que a comparabilidade tem também uma dimensão humana. Existe uma lacuna entre aqueles que preferem escrever em computadores e aqueles que preferem escrever no papel. Essa diferença traz problemas para um princípio tradicional da avaliação: que a equidade seja assegurada pela cuidadosa padronização de procedimentos. Assim, a padronização passa a não ser mais uma garantia de equidade. Oferecer as alternativas de se fazer um teste mediado pelo computador e/ou em papel pode, em algumas circunstâncias, ser uma forma de garantir equidade e validade.

Autores como Alderson, (2000); Buck, (2001); e Douglas, (2000) discutem as diferenças entre os métodos para testagem. Assim, reproduzimos abaixo um quadro comparativo (quadro 6) adaptado de Bachman & Palmer (1996) e Douglas (2000) sobre as vantagens e as limitações do uso de testes mediados por computador, com respeito às características do método de teste.

Quadro 6: As características do método do teste e as vantagens e desvantagens de testes mediados por computador

Características do método de teste	Vantagens		Limitações	
	Instituição	Candidato	Instituição	Candidato
Físicas	Facilidade para armazenamento e revisão dos testes.	Acessibilidade maior para candidatos que se encontram distante dos locais de teste.	Necessidade de investimentos em segurança (invasão ao sistema).	Necessidade de bom equipamento e conexão com internet, fones de ouvido.
			Número limitado de tipos de testes objetivos – múltipla escolha, verdadeiro ou falso, razão e asserção.	

Temporais	Podem ser aplicados em locais e horários diversos, sem a intervenção humana.	Conveniência de tempo e local.	Necessidade de padronização de equipamentos.	Conhecimento sobre o equipamento.
Econômicas	Diminuição de gastos com material impresso e que será em seguida descartado, e número de pessoas envolvidas com o teste.	Conveniência de tempo e local gera menos gastos.	Mão de obra especializada para trabalhar na área de Tecnologia da Informação.	Aumento no preço da inscrição do teste devido a gastos com equipamentos e mão de obra.
Normas (Rubric) e Procedimentos	Uniformidade e consistência na apresentação das instruções.	Tutoriais e telas de ajuda antecedendo o teste, ou mesmo durante o mesmo.	- Problemas com o equipamento, com o programa, velocidade da internet.	Não familiarização com o material e equipamento podem prejudicar o rendimento e desempenho do candidato.
	- Preparação dos candidatos ou examinadores por materiais online ou CD-ROM.		- Criação de tutoriais e exemplos práticos são mais trabalhosos.	
	Apresentação das instruções em várias línguas.			
	Cronometragem do teste por partes ou por completo.			
Insumo e autenticidade	Variedade de tipos de insumo na forma de informações visuais e auditivas, na composição de itens e tarefas.	Contextualização e autenticidade.	Problemas em relação à tecnologia disponível.	Consequências negativas relativas ao efeito do método no desempenho do candidato e na interpretação dos resultados do teste.
Interação entre o insumo e a resposta	Os computadores podem ser programados para atuarem de acordo com as respostas do candidato.	Recebimento rápido do resultado.	Testes mediados pelo computador são caros para serem desenvolvidos.	A interatividade é mais controlada em relação a outros formatos de prova.

	- Apresentação imediata dos resultados.			
Características da Avaliação	Correção automática pelo computador das questões de leitura e escrita e de processamento natural da fala.	Diminuição do estresse do candidato por não estar cara a cara com o entrevistador (intermediação da câmera).	Tecnologia de processamento da fala é cara e limitada, criando problemas potenciais para a definição do construto e validade.	O candidato pode ficar mais preocupado em aprender com o e-rater que corrige o teste ao invés de estudar o conteúdo que será avaliado.
	Diminuição do número de entrevistadores (não há a necessidade da presença física), gerando economia.		Questões indefinidas sobre número de itens e critérios de correção podem afetar o construto e a validade do teste.	
	Possibilidade de mais pesquisas sobre o assunto.			

Embora um teste de classificação possa não ser considerado de alto impacto como um exame vestibular ou um teste de proficiência, seus efeitos são significativos. Para os alunos, ele pode atrasar ou prolongar seus estudos, para o programa, um aluno posicionado de forma inadequada pode comprometer a qualidade e a integridade da sequência do curso. Portanto, um teste de classificação preciso é essencial.

No que se refere ao uso de tecnologia, especificamente aos testes mediados pelo computador, as pesquisas estão evoluindo, principalmente com testes padronizados como o TOEFL-cbt, e seu sucessor, Web-based TOEFL-iBT, o IELTS, e o TOEIC.

Embora as discussões e conclusões sobre esses testes variem, há consenso em alguns aspectos como o reconhecimento potencial das vantagens e das desvantagens de testes de línguas mediados pelo computador. Como vantagens são sempre citadas as formulações possíveis, uma maior padronização, o *feedback* imediato, a pontuação mais precisa, a apresentação multimídia, a segurança no armazenamento dos dados e a capacidade de avaliar uma ampla gama de níveis de habilidades, pois o computador pode ajustar automaticamente o nível do teste de acordo com o desempenho do aluno.

As desvantagens reconhecidas geralmente incluem complicações envolvendo a infraestrutura de acesso a computadores e à Internet, bem como o potencial efeito

negativo sobre os estudantes que não estão familiarizados ou se sentem desconfortáveis com a tecnologia. No entanto, à medida que essa familiaridade aumenta, com as novas gerações e avanços na tecnologia, as desvantagens deixam de existir.

2.2.3. Autenticidade

A autenticidade, segundo Bachman & Palmer (1996) é um dos traços que constrói a utilidade de uma avaliação. Se tivermos como objetivo em uma avaliação fornecer uma amostra representativa do conhecimento ou o desempenho de um determinado indivíduo em relação a certo construto, devemos procurar um tipo de avaliação que reproduza as condições nas quais esse construto é normalmente utilizado, de modo que o teste reproduza uma situação autêntica de uso da língua alvo. A essa situação autêntica, os autores definem essas tarefas de uso da língua alvo como *Target Language Use tasks* (TLU tasks).

Bachman & Palmer¹²(1996:23) definem autenticidade como “o grau de correspondência entre as características de uma tarefa avaliativa de uma língua e as características de uma tarefa de uso da língua alvo”. A figura 6 sintetiza a concepção de autenticidade dos autores. Entretanto, o conceito de autenticidade não é absoluto, uma vez que depende de como a tarefa é realizada, os participantes envolvidos, o nível de proficiência necessário.

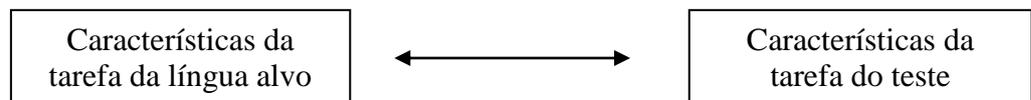


Figura 6: Autenticidade (Bachman & Palmer, 1996, p. 23)

Bachman (1991, p. 690) propõe dois tipos de autenticidade: a situacional e interacional. A autenticidade interacional refere-se ao envolvimento da capacidade linguística do examinando na realização do teste, ou seja, quais as áreas do conhecimento linguístico e quais estratégias estão envolvidas na solução do problema

¹² No original: “We define authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a TLU task.” (Bachman & Palmer, 1996: 23.)

apresentado pelas tarefas do teste. Neste caso temos uma interação entre o candidato e a tarefa.

A autenticidade situacional refere-se ao quanto se percebe das características do método avaliativo como relevantes para uma determinada situação, ou seja, a percepção do examinando em relação ao teste, que está diretamente relacionado ao que chamamos de validade de face, que seria a percepção subjetiva dos examinandos sobre quanto o teste avalia o que deveria avaliar.

Na verdade, a autenticidade está sujeita às percepções que os envolvidos em uma situação de avaliação têm sobre o uso real da língua e os elementos que a compõem. Assim, dois aspectos importantes relacionados à autenticidade em exames de línguas são autenticidade situacional e interacional. A autenticidade situacional refere-se à precisão com que tarefas e itens representam atividades linguísticas baseadas na vida real. Autenticidade interacional refere-se à naturalidade da interação entre os examinandos, as tarefas e os processos mentais que a acompanham.

Apesar da definição de autenticidade ser clara, o seu uso dentro dos vários tipos de teste não é absoluto, uma vez que as características das tarefas de uso da língua alvo variam de acordo com o propósito de uso da língua, os participantes envolvidos nesse processo, o contexto em que a tarefa é realizada e o nível de proficiência necessário.

Cumming & Maxwell (1999) argumentam que interpretações ambíguas em relação ao significado de autenticidade em avaliações podem levar a uma redução de qualidade do produto das mesmas. Em suas pesquisas, os autores verificaram que itens mal formulados, que acabam por não ter uma ligação com o uso real da língua, escolhidos sem critérios claros e coerentes, a falta de compatibilidade entre as noções de autenticidade de examinandos e examinadores são problemas que afetam a autenticidade em avaliações.

2.2.4. Confiabilidade

A confiabilidade de um teste está relacionada à consistência dos critérios utilizados para sua avaliação. Segundo Schoffen (2009:23) “a confiabilidade de um teste reside em minimizar os efeitos dos fatores externos ao teste, fazendo com que a habilidade linguística do candidato seja avaliada de maneira uniforme”.

Há fatores diversos que podem comprometer o desempenho do candidato e provocar desvios de interpretação ao se avaliar as habilidades linguístico-comunicativas de um aluno que está sendo submetido a um teste, como: cansaço, tanto do examinado quanto do avaliador; dúvida ou desconhecimento de um tópico, que pode gerar insegurança e sentimento de fracasso; a preferência por um determinado tipo de teste, ou certo modelo, como questões de múltipla escolha, ou então questões dissertativas; a utilização de provas eletrônicas ou então no modelo convencional com papel e caneta/lápis; conhecimento ou desconhecimento do contexto da prova.

Outros fatores também podem afetar a credibilidade de um teste, como:

- tamanho do teste: testes mais longos são mais confiáveis em relação aos mais curtos.
- unidimensionalidade: quando um teste avalia apenas uma habilidade, a confiabilidade aumenta em relação a testes que avaliam leitura e compreensão oral, por exemplo.
- motivação: um examinando motivado terá um desempenho melhor no teste e conseqüentemente, seu escore será maior.
- limite de tempo: período de tempo razoável para a execução da prova é primordial para o resultado.
- segurança: procurar evitar que respostas do teste possam ser passadas por alunos para outras turmas. Se o teste for *online*, é um exemplo de preocupação com possíveis acessos de alunos em ambientes fora do contexto escolar.

De acordo com Schlatter et al. (2005), quanto menos esses fatores afetarem a avaliação de um teste, e se conseguirmos avaliar no teste o domínio das habilidades linguísticas que foram designadas para serem avaliadas, maior a confiabilidade nos escores do teste.

Como forma de diminuirmos os fatores externos para que o teste tenha uma maior confiabilidade, devemos nos preocupar com itens como: instruções claras para o examinando, *layout* adequado, que a aplicação da prova siga padrões pré-determinados tais como, uso de equipamentos, ambientes adequados para a realização de testes;

treinamento dos aplicadores do teste, treinamento dos corretores, critérios claros de avaliação, dentre outros.

Para Bachman (1990), o que deve ser observado no processo, desenvolvimento e uso de um teste não deve ser se seus escores são confiáveis, porém se a interpretação e o uso que fazemos desses escores são válidos, pois quando aumentamos a confiabilidade de nossas medições, estamos providenciando uma condição necessária para a validade: para o escore de um teste ser válido, é necessário que primeiramente o teste seja confiável.

2.2.5. Efeito Retroativo

Todo tipo de avaliação provoca um efeito e traz consequências para um indivíduo ou grupo de pessoas. Essa influência que um teste ou exame exerce sobre o ensino e aprendizagem é definida na literatura em Linguística Aplicada ou Educação como efeito retroativo.

Hughes (1989: 1) afirma que “o efeito de testes no ensino e na aprendizagem é conhecido como ‘*backwash*’ (sinônimo de *washback*). Segundo Alderson et al (1988) é comumente aceito que exames em geral influenciam atitudes, comportamentos e motivações dos professores, aprendizes e pais. Para Messick (1996:241)¹³ o efeito retroativo trata da “amplitude a que a introdução e o uso de um exame influencia os professores e aprendizes de línguas a fazerem coisas que de outro modo não fariam, que promovem ou inibem a aprendizagem de línguas”.

Para Buck (1988:17)¹⁴ haveria “uma tendência natural que tanto professores como os alunos têm de adaptar as atividades em sala de aula às exigências de um prova, especialmente quando a prova é muito importante para o futuro dos alunos, e as notas de aprovação são usadas para avaliar o sucesso do professor. Essa influência da prova

¹³ No original: “The extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning”.

¹⁴ No original: “There is a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. This influence of the test on the classroom (referred to as *washback* by language testers) is, of course, very important; this washback effect can be either beneficial or harmful

sobre a sala de aula é bastante importante pois esse efeito retroativo pode ser benéfico ou prejudicial”. Neste caso a definição de Buck enfatiza o impacto de um exame nos professores e alunos no cenário da sala de aula.

Alguns fatores podem desencadear o efeito retroativo de um exame. Segundo Messick (1996, p. 243) a “evidência de influência no ensino e na aprendizagem deve ser considerada como efeito retroativo somente se tal evidência pode ser relacionada com a introdução e uso de um teste”. Um exame pode ser positivo para alguns e negativo para outros. Um exame da Ordem dos Advogados do Brasil, por exemplo, é de alto impacto e provoca um efeito retroativo na sociedade e nas faculdades formadoras de profissionais para a área de Direito.

Hughes (1989: 44-47) aponta sete maneiras de se promover efeito positivo ao se preparar um teste:

- 1) Avaliar as habilidades cujo desenvolvimento se quer encorajar.
- 2) Escolher amostras amplas e imprevisíveis.
- 3) Usar exames diretos.
- 4) Fazer o exame referenciado em critérios.
- 5) Basear exames de rendimento em objetivos.
- 6) Assegurar que o exame é conhecido e compreendido pelos alunos e professores.
- 7) Providenciar, quando necessário, assistência aos professores.

Shohamy (1992: 515) afirma que é preciso haver uma conexão entre exame e currículo, baseando-se em seis princípios nos quais um exame deve embasar-se para provocar um efeito positivo. Se um dos princípios for negligenciado, um efeito negativo provavelmente surgirá. Os princípios que devem ser respeitados são:

- 1) Rendimento Escolar e Proficiência: nem sempre o que se aprende na escola (rendimento escolar) é o que é necessário para se usar na vida real (proficiência). Exame, ensino, aprendizagem devem convergir para as necessidades reais da vida.
- 2) Informação diagnóstica: os resultados de exames devem auxiliar para o diagnóstico de problemas, a fim de se achar uma solução.
- 3) Conexão do ensino com aprendizagem: o ensino deve refletir a aprendizagem e vice-versa.

- 4) Envolvimento dos agentes que podem trazer mudanças: todos os que participam do processo de ensino/aprendizagem devem estar unidos para produzirem mudanças.
- 5) Necessidade de informações comparativas: um exame ideal é aquele que é referenciado em norma e em critério para que seus programas possam usar os resultados para avaliar seu sucesso em relação ao seu próprio programa e outros programas.
- 6) Necessidade de exames comunicativos: um exame deve refletir teorias de linguagem atuais.

Ainda, Schllater, Garcez e Scaramucci (2004: 16) apontam fatores que podem ocasionar efeito retroativo benéfico, como:

- a) A qualidade do instrumento de avaliação, o que envolve a validade em seus vários aspectos e a coerência desse instrumento com o construto que busca operacionalizar: critérios de correção devidamente explicitados, com perguntas bem elaboradas e instruções claras, sem ambiguidades, explicitando a natureza e o nível de proficiência almejados;
- b) o conjunto de informações sobre o exame, explicitando seu formato e o nível de exigência necessário, assim como sua abordagem direcionadora;
- c) a língua avaliada deve ter um *status* importante no contexto de uso do exame, que também deve ser relevante para os candidatos;
- d) a boa formação do professor.

Uma forma de potencializar o efeito negativo que testes e exames podem causar em indivíduos é a má formação do professor, ou ainda, quando esse prepara suas aulas somente voltadas aos exames. Assim, vários fatores se agrupam para se ter um bom teste, e é importante que o professor pense nos efeitos que esse teste pode causar nos alunos ou mesmo na sociedade como um todo.

2.2.6. Impacto

Nos últimos vinte anos temos visto uma crescente conscientização de que testes podem ter consequências não somente nas salas de aula. Os resultados de testes têm um

impacto significativo nas carreiras educacionais e profissionais das pessoas. Também têm impacto nos sistemas educacionais e na sociedade em geral, como, por exemplo: no planejamento do currículo escolar, nas políticas de imigração, no exercício profissional (médicos, advogados, por exemplo). Como resultado desse impacto, a partir do momento que o teste se torna de alta relevância começam a surgir materiais e cursinhos preparatórios para o teste.

Alguns pesquisadores da área de avaliação consideram o efeito retroativo como uma dimensão do impacto, descrevendo seus efeitos sobre o contexto educacional (Hamp-Lyons, 1997); outros distinguem o efeito retroativo do conceito de impacto, relacionado-os respectivamente como efeitos “micro” e “macro” na sociedade. O efeito retroativo estaria relacionado-os aos efeitos da avaliação no interior do processo educacional, influenciando no processo de ensino e aprendizagem e nos sujeitos envolvidos nesse processo (alunos e professores), enquanto que o conceito de impacto teria a ver com os efeitos da avaliação na sociedade de maneira mais ampla e abrangente (Alderson e Wall, 1993; McNamara, 2000; Scaramucci, 2004).

Como aponta Spolsky (1997) e outros pesquisadores (McNamara, 2000, McNamara e Shohamy, 2001), os testes não podem ser vistos como instrumentos neutros, mas sim como instrumentos de poder e controle diante da comunidade estudantil.

McNamara (2000) relaciona a avaliação a dois tipos de políticas: sociais e educacionais. Como política social o autor cita um exame desenvolvido na década de 1960 pela Real Polícia Montada Canadense com o propósito de excluir de seus quadros homossexuais. O autor cita ainda a utilização de exames de proficiência pelo Governo Australiano para selecionar candidatos a refugiados durante a Segunda Grande Guerra, muitos deles judeus que fugiam do Nazismo e do Fascismo na Europa, que eram submetidos a exames de proficiência em diferentes línguas (além do inglês) até serem reprovados e, conseqüentemente, impedidos de migrar para a Austrália.

Para ilustrar o seu importante papel na política educacional, McNamara (2000) cita os testes de proficiência em línguas que são usados como forma de controle de acesso a programas internacionais de educação e também a utilização de exames de proficiência na seleção de candidatos ao Ensino Superior.

Para McNamara (2000: 71-72), é possível distinguir duas posições na discussão acerca do papel social e político dos testes. A primeira pressupõe que a prática avaliativa pode tornar-se ética e enfatiza a responsabilidade individual dos avaliadores (*testers*); a segunda posição, contrariamente, vê a avaliação com sendo um construto essencialmente sociopolítico, uma vez que os testes são concebidos como instrumentos de poder e controle, e que deve, portanto, estar sujeitos às mesmas críticas dirigidas a outras estruturas políticas da sociedade. McNamara (2000) denomina essas duas posições de *avaliação ética (ethical language testing)* e *avaliação crítica da linguagem (critical language testing)*, respectivamente.

2.2.7. Validade

A validade é descrita como o conceito mais importante da psicometria, mas suas definições estão ainda um pouco confusas devido a várias ocorrências com o passar dos anos.

Assim, validade

1. não é uma propriedade inerente de um teste,
2. refere-se às interpretações ou ações que são feitas com base em resultados de testes, e
3. deve ser avaliada em relação ao propósito do teste e como o teste é usado.

A validade de um teste refere-se ao grau com que as inferências com base em resultados desse teste são significativas, úteis e apropriadas, trata-se portanto, de um critério para sua aceitabilidade. A validação de um teste refere-se ao agrupamento – reunião de dados empíricos e argumentos lógicos para mostrar que as inferências são realmente apropriadas. Tradicionalmente, os vários meios de acumular evidências de validade foram agrupados em três categorias, relacionadas ao conteúdo, ao critério, e ao construto.

A validade de um teste só pode ser estabelecida por meio de um processo de validação, e isso deve ser feito antes que os resultados possam ser usados para qualquer outra finalidade particular. A fim de realizar essa validação, um estudo de validação deve ser realizado, para se chegar a uma conclusão quanto às interpretações e

utilizações dos resultados do teste. O processo de validação foi pouco realizado no passado (Davies & Elder, 2005:802-803), porém a validação é necessária devido ao impacto que os resultados de testes podem ter sobre os examinandos e as pessoas envolvidas no processo.

Straub (1989) verificou em levantamento em revistas especializadas, que pouca atenção é dada à validação de instrumentos de pesquisa. Um dos erros mais frequentes é o uso de instrumentos utilizados em pesquisas anteriores sem validá-los, ou seja, utilizam-se instrumentos de outras pesquisas, fazendo-se adaptações e não se validam novamente os instrumentos.

Assim, para se alcançar um razoável grau de rigor metodológico devemos nos ater a essas questões, principalmente quando se pretende medir o mesmo construto em diferentes amostras. De acordo com Straub (1989:148), “... com a utilização de instrumentos validados, pesquisadores podem medir o mesmo construto de pesquisa da mesma forma, permitindo melhor a medição de variáveis dependentes e independentes...”.

2.2.7.1. Validade: um conceito plural

Na literatura relacionada à área de avaliação podemos encontrar diferentes concepções sobre validade. Só como exemplos, podemos citar Hughes (1989) que afirma que “um teste é válido se mede precisamente aquilo que deve medir”. Alderson et al. (1995) retoma a definição de Henning (1987:96) e afirma que:

Validade em geral refere-se à adequação de um teste ou de algum de seus componentes como uma medida do que esse teste deve medir. Um teste é válido no grau em que mede o que deve medir. Assim, o termo válido, quando usado para descrever um teste, deve geralmente vir acompanhado pela preposição “para”. Qualquer teste, dessa forma, pode ser válido para alguns propósitos, mas não para outros. (HENNING, 1987 *apud* ALDERSON *ET AL.*, 1995: 89).¹⁵

¹⁵ No original: “Validity in general refers to the appropriateness of a given test or any of its component parts as a measure of what it is purported to measure. A test is said to be valid to the extent that it measures what is supposed to measure. It follows that the term valid when used to describe a test should usually be accompanied by the preposition for. Any test then may be valid for some purposes, but not for others” (Henning, 1987 *in* Alderson *et al*, 1995: 89).

Messick (1989:13) afirma que validade é como "um julgamento avaliativo integrado do grau em que as evidências empíricas e fundamentos teóricos apoiam a adequação e a pertinência de inferências e ações com base em escores de teste e outras formas de avaliação".¹⁶

Assim, podemos entender que :

- a) validade refere-se às interpretações dos escores de teste, não ao teste em si, e
- b) o que validamos são interpretações (e ações) com base nos escores de teste, não um teste de *per se*.

Desde 1952, as opiniões sobre a teoria de validade e validação de testes são apresentadas por várias associações profissionais, particularmente a *American Educational Research Association (doravante AERA)*, a *American Psychological Association (doravante APA)*, e a *The National Council on Measurement in Education (doravante NCME)*. Assim, as definições e o quadro de validade fornecidas por essas associações serão mais destacados nesta revisão histórica.

2.2.7.2. Um Breve Histórico da Teoria da Validade

A teoria da validade e a prática de validação de testes são quase tão antigas quanto a prática de avaliação. Wainer (2000) refere-se a um capítulo da Bíblia, do “Livro dos Juízes” (12:4-6) como, talvez, o primeiro caso de teste de alta relevância. Nessa história da Bíblia, os gileaditas desenvolveram um teste de um único item para identificar os efraimitas inimigos que estavam escondidos entre eles. O teste consistia em pronunciar a palavra “shibboleth”. Embora a relevância associada a passar nessa “prova” de um único item fosse alta (i.e. todos os que falharam em suas respostas foram mortos), a Bíblia não menciona qualquer estudo conduzido sobre validade para apoiar seu uso.

A era moderna da avaliação educacional provavelmente começou com Alfred Binet, que desenvolveu uma bateria de 30 subtestes rápidos, projetados para assegurar que não fosse negado a qualquer criança parisiense a entrada no sistema escolar sem um

¹⁶ No original: “an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of inferences and *actions* based on a test scores and other modes of assessment,”

exame formal. A finalidade de Binet era louvável, uma vez que para o exame os objetivos consistiam na inclusão de crianças que estavam sendo excluídas da escola e da identificação de estudantes que se beneficiariam do sistema escolar. Mais era a característica mais importante, e a característica fundamental do tipo de avaliação utilizada: o uso de testes padronizados. As crianças foram testadas com o mesmo instrumento, nas mesmas condições, e foram pontuadas, portanto, de maneira uniforme.

No século XX a escala Binet-Simon foi publicada (Binet, 1905; Binet & Henry, 1899) de forma que é difícil encontrar estudos de validade hoje em dia para ela. Entretanto, uma revisão das primeiras obras referentes à psicometria indica que as primeiras definições sobre validade eram amplamente pragmáticas, definindo validade em termos da correlação entre os escores de um teste com alguns critério subjetivo. Assim, no início de 1920, os testes eram descritos como válidos para qualquer coisa com que mantinham algum tipo de correlação. Esse ponto de vista foi assegurado até o meio do século XX e está evidenciado pela definição de Guilford (1946, p.429), que assinala que “um teste é válido para qualquer coisa com a qual se correlaciona”.

A visão empírica de validade foi ampliada com o surgimento da análise fatorial desenvolvida por Spearman (1904). Os primeiros psicometristas viam a análise fatorial como uma ferramenta para o entendimento das “características observáveis” ou “construtos” que estão na base dos testes de desempenho. Muitos psicometristas interpretavam os fatores derivados da análise de item e dados de teste como representações latentes de atributos psicológicos, enquanto outros argumentavam que os fatores eram “... apenas um resumo matemático ou simplificação conceitual das relações observadas entre tarefas concretas”. Guilford (1946:428) foi um forte defensor do uso da análise fatorial para definir validade e validar os testes. Ele classificou validade em duas categorias: validade fatorial e validade prática. Validade prática referia-se às correlações entre os escores de teste e os critérios relevantes, enquanto validade fatorial estava relacionada às cargas fatoriais do teste no "sentido comum, dos fatores de referência".

Essa visão pragmática e empírica de validade dominou as pesquisas e a prática no desenvolvimento e validação de testes como fica evidente pelos trabalhos desenvolvidos por militares na Primeira e Segunda Guerras Mundiais. Uma série de

testes foram aplicados a milhões de recrutas para ajudar a classificá-los nas unidades de Serviço. Como é descrito por Jenkins (1946), os testes para o pessoal militar foram adicionados às baterias baseados unicamente sobre a validade incremental, isto é, a capacidade de representar mais variações no desempenho em um critério.

Além da visão empírica de validade com base na correlação e análise fatorial, uma definição teórica de validade também foi proposta no início do século 20. Esse ponto de vista teórico atesta que a validade é o grau em que "um teste mede aquilo que deve medir". Embora amplamente refutada hoje e criticada como uma definição incompleta, essa visão simplista ainda é vista nos livros sobre o assunto e algumas literaturas psicométricas, por ser um requisito importante para apoiar o uso de um teste para um propósito particular, ou seja, não se pode interpretar corretamente uma pontuação de um teste se aquilo que o teste avalia não for conhecido.

Para resumir as primeiras concepções de validade, a teoria da validade surgiu de uma forma congruente com a filosofia da ciência e a estatística popular na época em que testes educacionais e psicológicos surgiram pela primeira vez. Teorias de atributos psicológicos não observáveis existiam e podiam ser medidas, promovendo a ideia de que validade é o grau que um teste mede aquilo que é suposto medir. Os novos métodos estatísticos de correlação e análise de fatores proporcionaram uma estrutura para avaliar se os resultados dos testes e os itens de respostas observadas são congruentes com noções prevalecentes sobre o que o teste devia "supostamente medir". Embora as teorias e estratégias de validação iniciais fossem simples, não demorou muito para a insatisfação com a teoria e a prática de validação de testes emergirem.

2.2.7.3 A teoria da validade evolui além das relações teste/critérios - 1940

Embora muitos dos primeiros psicometristas do século 20 tenham ficado animados com os procedimentos estatísticos de correlação e de regressão para a validação de testes, outros estavam completamente insatisfeitos com a noção de que a validade referia-se apenas a "o que é um teste podia medir" e que os testes poderiam ser validados somente por meio de estudos de fatores analíticos e correlacionais.

Muitos pesquisadores apontaram que os critérios com os quais os testes eram correlacionados eram muitas vezes imperfeitos na melhor das hipóteses, e tinham confiabilidade e validade questionáveis.

Rulon (1946) é um exemplo de pesquisador que clamava por um tratamento mais abrangente da teoria da validade e validação de testes. Um de seus primeiros trabalhos nessa área foi: *Technical Recommendations for Psychological Tests and Diagnostic Techniques: A Preliminary Proposal* (APA, 1952), que resultou de um painel encomendado pela APA para oferecer uma proposta formal de padrões de teste para serem usada em construções, uso e interpretações de testes psicológicos. Esse comite mudou a concepção e a terminologia de validade e descreveu quatro categorias para a mesma: validade preditiva, validade de status, validade de conteúdo e validade congruente. Essa proposta preliminar levou a formação de um comitê misto reunindo o a APA, AERA e NCME, que publicou o trabalho intitulado de *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (1954). Essa publicação representa a primeira versão do que é hoje o Standards for Educational and Psychological Testing (AERA et al, 1999) e que mudou as quatro “categorias” de validade para “tipos” ou “atributos” de validade. A “validade congruente” foi renomeada de “validade de construto” e a “validade de status” como “validade concorrente”.

Com a publicação da *Technical Recommendations* em 1954, o conceito de “validade de construto” passou a existir. Entretanto, o conceito foi, de fato, descrito e elaborado por dois membros do Comitê Misto – Lee Cronbach e Paul Meehl (1955) que afirmam em seu artigo seminal:

a validação de construto acontece sempre que um teste deve ser interpretado como uma medida de algum atributo ou qualidade que não é “operacionalmente” definido. O problema enfrentado pelo pesquisador é “O que é relevante para construções de variância no desempenho do teste?”¹⁷ (Tradução nossa).

Uma leitura mais atenta ao trabalho de Cronbach e Meehl (1955) sugere que eles inicialmente propuseram o conceito de validade de construto primeiramente para testes

¹⁷ No original: “construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not ‘operationally’ defined. The problem faced by the investigator is ‘What constructs account for variance in test performance?’ (Cronbach & Meehl, 1955:282).

de natureza mais psicológica em oposição àqueles de progresso educacional. Os dois excertos a seguir ilustram esse ponto.

1. A validade de construto deve ser investigada sempre que nenhum critério ou universo de conteúdo é aceito como inteiramente adequado para definir a qualidade a ser medida (p.282).¹⁸
2. A validade de construto foi introduzida a fim de especificar os tipos de investigação necessários no desenvolvimento de testes para os quais as visões convencionais de validação são inapropriadas.¹⁹

Dado que as *Recommendations* de 1954 listaram 4 (quatro) tipos de validade, é plausível que com essas duas características apresentadas para os dois tipos de teste a validade de construto seja descrita como a mais pertinente aos exames psicológicos e a validade de conteúdo seria descrita como a melhor para os testes educacionais. Em todo caso, parece claro que Cronbach e Meehl (1955) hesitaram em atribuir validade de construto para todos os testes educacionais e psicológicos. Logo depois, no entanto, vários teóricos de validade concluíram que a validade de construto poderia ser aplicável a todos os testes educacionais e psicológicos.

Messick e outros teóricos argumentaram que nunca há um critério universal aceitável de conteúdo, o que significava que toda validade é validade de construto. Essa noção de validade ficou conhecida como a conceituação unitária de validade e continua popular até hoje.

2.2.7.4. Conceitos contemporâneos de validade

A conclusão de que a validade não é uma propriedade inerente de um teste, mas sim um conceito que se refere à adequação das inferências derivadas de um teste é apoiada na versão atual dos Standards (AERA et al., 1999) bem como em outros artigos seminais sobre validade desde 1974 incluindo o de Messick (1975, 1989); Kane (1992,

¹⁸ No original: “Construct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured (p.282).

¹⁹ Construct validity was introduced in order to specify types of research required in developing tests for which the conventional views on validation are inappropriate (p.299).

2006), e Shepard (1993). Na verdade, a definição de Messick (1989) de validade como "um julgamento avaliativo integrado do grau em que as evidências empíricas e fundamentos teóricos apoiam a adequação e a pertinência de inferências e ações com base em escores de teste e outras formas de avaliação".²⁰

Essa definição está muito próxima da definição constante na versão mais recente dos *Standards*, que afirmam que "A validade se refere ao grau em que evidências e teorias apoiam as interpretações nos escores dos testes decorrentes dos usos propostos para os testes (AERA et al., 1999, p.9).²¹

Messick (1989, 1996a, 1996b) argumenta que a concepção tradicional de validade é fragmentada e incompleta, especialmente porque não leva em conta tanto a evidência das implicações de valor dos significados da pontuação como base para a ação quanto as consequências sociais do uso de pontuação. Sua abordagem moderna visualiza validade como um conceito unificado que coloca uma ênfase mais pesada em como um teste é usado. Seis aspectos distintos de validade são destacados: evidências relativas ao conteúdo, substantiva, estrutural, passível de generalização, externa e consequencial, como meio de abordar questões centrais implícitas na noção de validade como um conceito unificado.

Assim, os seis aspectos de validade para Messick (1989:20) são:

- 1) Evidências relativas ao conteúdo: Uma questão fundamental para o aspecto da validade de conteúdo é determinar o conhecimento, habilidades e outros atributos para ser revelado pelas tarefas de avaliação. Os padrões de conteúdo devem ser relevantes e representativos do domínio do constuto.
- 2) Substantiva: inclui análises quantitativas para reunir evidências para dar sustento a notas de um exame e suas possíveis interpretações. Em primeiro lugar, as tarefas de avaliação devem ter a capacidade de fornecer uma amostragem adequada de processos de domínio, além da cobertura tradicional de conteúdo de domínio

²⁰ No original: "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of inferences and *actions* based on a test scores and other modes of assessment,"

²¹ No original "Validity refers to the degree to which evidence and theory support the interpretations of test and entailed by proposed uses of tests" (AERA et al., 1999:9).

- 3) Estrutural: investiga a estrutura interna do exame avaliando até que ponto uma dimensionalidade observada das respostas coletadas é consistente com o construto que foi considerado. A maneira pela qual a execução de tarefas são avaliadas e contabilizadas devem basear-se em como os processos implícitos das ações do entrevistado combinam dinamicamente para produzir efeitos.
- 4) Passível de Generalização: As avaliações devem fornecer cobertura representativa do conteúdo e do construto. Isso permite que as interpretações das notas possam ser generalizadas para um dado grupo e além dele, assim como nos cenários e tarefas.
- 5) Externa; inclui evidências provenientes de comparações que sinalizam convergências entre exames bem como aspectos que os discriminam.
- 6) Consequencial: O aspecto consequencial da validade inclui provas e justificativas para avaliar as consequências intencionais e não intencionais de interpretação da pontuação e uso. Esse tipo de investigação é especialmente importante quando se trata de consequências adversas para os indivíduos e grupos que estão associados com viés na pontuação e interpretação

Com efeito, esses seis aspectos conjuntamente funcionam como critérios de validade geral ou padrões para todas as medições educacionais e psicológicas. Esses seis aspectos devem ser vistos como formas interdependentes e complementares de evidência de validade e não vistos como tipos de validade separados e substituíveis.

Messick (1989:20) apresenta um modelo unificado, conhecido como “matriz progressiva de Messick” e coloca a validade de construto como um componente central, sendo as validades de conteúdo e de critério como aspectos da validade de construto. Além disso, essa matriz oferece uma diretiva no sentido de orientar de que forma as evidências podem ser produzidas para ajudar nos métodos de validação e apresenta consequências e impactos sociais. Entretanto autores como McNamara e Roever (2006); Shepard (1993), acham essa matriz complexa e difícil de ser operacionalizada.

Sua matriz progressiva está representada na tabela abaixo:

Quadro 7: Matriz progressiva de Messick (1989:20)

	INFERÊNCIAS	USOS
Base evidencial	Validade de construto	Validade de construto + Relevância/utilidade
Base consequencial	Validade de construto + Implicação de valor	Validade de construto + Implicações de valor + Relevância/utilidade + Consequências sociais.

Baseando-nos nessa revisão histórica de validade e suas definições atuais, fica claro que o consenso para validade é que validade não deve ser atribuída a um teste, porém ao uso de um teste para um propósito particular e seu uso defensável. Portanto, a validade não é uma propriedade do teste ou da avaliação, mas do significado dos seus resultados. Assim, a alegação de que a validade refere-se simplesmente para demonstrar que “um teste mede aquilo que é proposto medir”, ou que é uma propriedade inerente de um teste é ignorar no mínimo 70 anos de pesquisa em torno da teoria da validade e validação de testes, bem como todo o material publicado como *Technical Recommendations and Standards* que pesquisam o assunto desde 1954.

A tabela abaixo extraída de Chapelle (1999: 258) apud Scaramucci (2009:34) é ilustrativa para demonstrar as mudanças em termos de validade tradicional e moderna.

Quadro 8: Diferenças de concepções entre a validade tradicional e a contemporânea

Tradicional	Contemporânea
Considerada uma característica de um teste: o grau em que um teste mede aquilo que pretende medir.	Validade é um argumento relativo à interpretação e uso: o grau em que as interpretações e usos de um teste podem ser justificados.
Confiabilidade era vista como distinta e uma condição necessária para validade.	Confiabilidade pode ser vista como um tipo de evidência de validade.

<p>A validade era frequentemente estabelecida através de correlações de um teste com outros.</p>	<p>Validade é argumentada com base em um número de tipos de justificativas e evidências, incluindo as consequências da avaliação.</p>
<p>Validade de construto era vista como um dos três tipos de validade (conteúdo, relacionada a critério e construto).</p>	<p>Validade é um conceito unitário, em que a validade de construto ocupa uma posição central: validade de conteúdo e relativa a critério podem ser usadas como evidência sobre validade de construto.</p>
<p>O estabelecimento da validade era uma tarefa de responsabilidade de pesquisadores da avaliação, responsáveis pelo desenvolvimento de testes de grande escala e alta relevância.</p>	<p>A justificativa de validade de um teste é de responsabilidade de todos os usuários de um teste.</p>

2.2.7.5. Ameaças à Validade

Duas grandes ameaças para testar a validade são dignos de nota, especialmente com a ênfase atual sobre os testes de alta relevância, que são:

- "Sub-representação do construto" que indica que as tarefas que são medidas na avaliação deixam de incluir importantes dimensões ou facetas do construto. Portanto, os resultados do teste não são susceptíveis de revelar a verdadeira capacidades de um aluno dentro do construto que foi indicado para mensuração do teste.

- "Variância irrelevante ao construto" significa que o teste mede muitas variáveis, muitas das quais são irrelevantes para o construto interpretado. Este tipo de invalidação pode assumir duas formas, "facilidade irrelevante ao construto" e "dificuldade irrelevante ao construto".

A "facilidade irrelevante ao construto" ocorre quando há pistas estranhas no item ou no formato das tarefas que permitam que alguns indivíduos respondam corretamente ou adequadamente em formas que são irrelevantes para que o construto seja avaliado,

A "dificuldade irrelevante ao construto" ocorre quando os aspectos externos da tarefa tornam a tarefa difícil para alguns indivíduos ou grupos. Enquanto o primeiro tipo de variância irrelevante ao construto leva alguém a pontuações mais elevadas do que seria em circunstâncias normais, esse último faz com a pontuação seja bastante

inferior. Compromete a característica do teste de ser direto (*directness*). Como há uma dependência relativa de respostas e tarefas sobre os processos, estratégias e conhecimentos que estão implicadas no desempenho das tarefas, a pessoa deve ser capaz de identificar por meio da análise cognitiva, os mecanismos teóricos subjacentes ao desempenho da tarefa.

2.3. Validação de Testes

Validar é um processo de investigação, é examinar uma determinada predição ou inferência realizada a partir dos escores de um teste. Não se valida o teste em si, mas a interpretação dos dados decorrentes de um procedimento específico.

“A validade, segundo o consenso dos especialistas, não é uma característica geral, antes de tudo ela é específica. Um instrumento de medida não é válido em tese, pode ser válido para um curso, mas não para outro. Pode ser válido para um currículo, mas não para outro; para um professor, mas não para outro, inclusive, pode ser válido para uma escola, mas não o ser para outra instituição.” (VIANA, (2003: 53).

Para o autor, a validação, além de obter escores numéricos, é uma forma de examinar situações relativas à aprendizagem. Há três aspectos da validade que correspondem aos objetivos de um teste e que são a validade de construto, a validade de conteúdo e a validade de critério.

Assim, o que orienta o trabalho aqui proposto é uma indagação sobre como se constitui a validade de construto, validade de conteúdo e validade de face do teste de classificação aplicado aos alunos ingressantes dos três cursos superiores de tecnologia.

2.3.1. Validade de construto

A validade de construto é um dos temas mais complexos na área de avaliação, mas de grande relevância na área educacional. Várias definições são postuladas para o termo construto e brevemente serão apresentadas, de forma cronológica, algumas dessas definições.

Na verdade, o que é um construto? A partir de uma perspectiva de desenvolvimento de um, teste construto é um conceito (ou um grupo de conceitos)

usado para explicar o que os testes avaliam. Anastasi²² (1986:4-5) define o construto relacionado a testes como “conceitos teóricos de diferentes graus de abstração e generalizações que facilitam a compreensão de dados empíricos”.

Chapelle (1998:33) apresenta uma definição muito mais simplificada e expõe que são “interpretações significativas de comportamento observado”. Os construtos são expressos em palavras e as palavras usadas para se dizer o que os escores de um teste significam identificam o construto usado na interpretação de uma tarefa de recepção oral ou em um exame de proficiência.

Para Hughes (1989:26) a definição de construto no contexto de aprendizagem de línguas é “qualquer habilidade (ou traço) subjacente que é hipotetizado em uma teoria de habilidade linguística”.

Bachman (1990:41) apresenta sua definição de construto da seguinte maneira:

Características físicas como altura, peso e cor dos olhos podem ser experienciadas diretamente por meio dos sentidos, e podem, portanto, ser definidas por comparação direta com um padrão observável. Habilidades mentais como proficiência linguística, contudo, não podem ser observadas diretamente. Não podemos experienciar competência gramatical, por exemplo, da mesma forma que experienciamos a cor dos olhos. Nós inferimos a habilidade gramatical por meio da observação do comportamento que presumimos ser influenciado pela habilidade gramatical. O primeiro passo para a mensuração de uma dada habilidade linguística é, portanto, distinguir o construto que desejamos medir de outros construtos semelhantes, definindo-o de forma clara, precisa e não ambígua. Isso pode ser conseguido por meio da determinação de quais características específicas são relevantes para o dado construto. (BACHMAN, 1990, p. 41).

Portanto, a validade de construto refere-se à prova de que o instrumento realmente mede aquilo que se propõe medir. Para se chegar a uma validação do construto são necessários estudos relacionados a testes estatísticos e a construções teóricas sobre a relação entre as variáveis a serem medidas. (Raymundo, 2009).

Tradicionalmente, os especialistas em avaliação distinguem diferentes tipos de validade: validade de conteúdo, preditiva, concorrente, de construto e de face. Em várias publicações, Messick (1989, 1994, 1996) argumentou que a validade de construto é um

²² No original: “theoretical concepts of varying degrees of abstraction and generalizability which facilitate the understanding of empirical data.”

conceito multifacetado, porém unificado e abrangente, e que pode ser pesquisado a partir de diferentes perspectivas.

Uma propriedade chave do construto, de acordo com Chapelle (1998:33-34), é que ele deve refletir a consistência do desempenho. Isso é necessário porque só através da consistência do desempenho os pesquisadores e especialistas em avaliação podem ir além da observação de casos discretos para generalizações sobre o que as observações indicam.

Para detectar consistência, várias observações da mesma categoria de desempenho devem ser feitas. Juntas, as categorias observadas formam o construto que será avaliado. É possível que o desempenho em algum nível de habilidade seja consistentemente inconsistente. Entretanto, se nenhuma consistência for detectada em qualquer nível de habilidade, será difícil afirmar que as realizações de um construto foram observadas. Nesse caso, a relação entre a teoria do construto e a sua operacionalização nas tarefas e nas categorias de observação devem ser verificadas e, se a operacionalização for aceita, a teoria, provavelmente requer uma revisão ou pelo menos não tem apoio em estudos empíricos.

A principal finalidade de um teste de línguas é medir a habilidade linguística do examinando, embora testes possam ser usados para outras funções também, como, por exemplo, promover aprendizagem (Bachman & Palmer, 1996:19). Para um teste ter a finalidade de medir a capacidade linguística do examinando, uma das bases do teste deve apresentar um contínuo de habilidades. Diferentes níveis de pontuação no contínuo devem denotar diferentes graus ou valores da capacidade avaliada. Essas pontuações indicam o construto medido em um teste e, uma vez que essas medições são representadas em pontos ou categorias nesse contínuo, a precisão e o significado das pontuações numéricas são dependentes da qualidade de medição do teste. Isso é determinado pelo grau de consistência dos resultados e pela extensão dos diferentes escores que podem demonstrar quantidades diferentes da capacidade avaliada.

Um teste de línguas é um instrumento usado para expressar aspectos específicos da habilidade linguística do examinando através de escores. A habilidade linguística geralmente é o construto de mais alto nível em testes de línguas, mas nenhum teste pode

abarcando tudo isso. A razão é que a habilidade linguística é um conceito bastante amplo. Por outro lado, os testes são compostos de pequenas amostragens de tarefas que devem ser desempenhadas em um curto período de tempo. Além disso, a partir da definição de competência linguístico-comunicativa, postulada por alguns especialistas, considerar todos os detalhes da teoria para serem empregados em testes é muito difícil, por não dizer, quase impossível. As definições do construto relacionadas a um teste devem especificar quais aspectos o teste pretende verificar.

Aqueles que desenvolvem testes de línguas precisam ser capazes de dizer o que os escores do teste representam. Nesse sentido, a definição do construto é motivo de preocupação, o que significa, na prática, a necessidade de detalhamento das definições e descrições para públicos e propósitos diferentes; sem falar que evidências para a análise dos escores são necessárias ou suficientes para se responder a questionamentos que não têm uma resposta padrão.

Hughes (1989: 26), aponta que as especificações das habilidades ou das estruturas a serem cobradas no teste são essenciais para sua validação. As especificações devem ser feitas nos primeiros estágios do teste. Segundo o autor, nem tudo o que foi especificado pode aparecer no teste; entretanto esse estágio fornecerá a base para a elaboração de um teste preciso e confiável.

Ao tratar de validação de construto, Raymundo (2009: 88) aponta que:

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 do teste, com vistas a ter elementos que possam esclarecer o significado do instrumento (RAYMUNDO, 2009:88).

Portanto, ao se validar o construto de um teste, validamos também a teoria pela qual o instrumento se fundamenta. No caso deste trabalho de pesquisa, há claramente alguns problemas para se estabelecer a validade de construto. O teste de classificação, que é o foco de validação, somente avalia as habilidades de recepção escrita e de recepção oral. A parte de produção oral somente é avaliada quando o examinando

atingir o nível 3 (três) no teste escrito e não há uma avaliação escrita. Isso pode implicar na percepção de que o teste, além de não valorizar a produção oral e escrita dos alunos, é construído com base na teoria de que habilidades linguísticas de produção do aluno possam ser proporcionais às suas habilidades receptivas (Miyata, 2004: 61), de forma que se avaliarmos uma das habilidades, estaremos avaliando todas elas. Para ilustrar essa observação, podemos citar que alguns exames de proficiência, como o TOEIC, avalia somente a recepção oral e escrita. Entretanto, alguns estudos sobre o fato são preliminares, e novas pesquisas se fazem necessárias sobre o assunto.

2.3.2. Validade de conteúdo

Segundo Hughes (1994:22), um teste tem validade de conteúdo se seu conteúdo constitui uma amostra representativa das estruturas, habilidades e sub-habilidades que pretende avaliar. Weir (2005: 19) utiliza o termo validade de contexto para se referir à validade de conteúdo, devido à dimensão social da linguagem que o termo abrange. Entretanto, neste trabalho, utilizaremos o termo validade de conteúdo, admitindo o termo validade de contexto como sinônimo.

Assim, de acordo com Weir,²³ (2005: 19) validade de contexto ou conteúdo:

... refere-se à extensão a que a escolha das tarefas em um teste é representativa de um universo maior de tarefas no qual assume-se que o teste é uma amostra. Está relacionado com as exigências linguísticas e do interlocutor feitas pela tarefa (s), bem como das condições em que a tarefa é executada resultante tanto da tarefa em si quanto do cenário administrativo. (WEIR, 2005, pág. 19)

Além disso, Anastasi²⁴ (1986, p. 132) prevê um conjunto de diretrizes úteis para estabelecer a validade de conteúdo:

1) O domínio do comportamento a ser testado deve ser sistematicamente analisado para nos certificarmos de que todos os aspectos principais estão cobertos no teste, e proporcionalmente corretos;

²³ No original: “Context validity is concerned with the extend to which the choice of tasks in a test is representative of the larger universe of tasks of which the test is assumed to be a sample. This coverage relates to linguistic and interlocutor demands made by the task(s) as well as the conditions under which the task is performed arising from both the task itself and its administrative setting.”

²⁴ No original: “1. The behaviour domain to be tested must be systematically analysed to make certain that all major aspects are covered by the test items, and in the correct proportions; 2. The domain under consideration should be fully described in advance, rather than being defined after the test has been prepared; 3. Content validity depends on the relevance of the individual’s test responses to the behavior area under consideration, rather than on the apparent relevance of item content.”

- 2) O domínio em consideração deve ser prévia e totalmente descrito, em vez de fazê-lo somente após o teste ter sido elaborado;
- 3) A validade de conteúdo depende da relevância das respostas individuais dos testes em relação à área de comportamento que está sendo considerada, em vez da aparente relevância do conteúdo do item. (tradução minha)

Para analisarmos se um teste tem validade de conteúdo ou não, todo esforço deve ser feito dentro dos limites da situação do teste para que o mesmo se aproxime de uma autenticidade situacional. Não é no ambiente de sala de aula ou em um momento de avaliação que se atinge uma autenticidade plena. Entretanto, o cenário de ensino ou de uma avaliação deve ser o mais realista possível.

A autenticidade é uma forma de se construir a utilidade de uma avaliação. Se tivermos como objetivo em uma avaliação fornecer uma amostra representativa do conhecimento ou do desempenho de um indivíduo em relação a um determinado construto, devemos procurar um tipo de avaliação que reproduza as condições nas quais esse construto é normalmente utilizado, de modo que o teste busque reproduzir, da melhor forma, uma situação autêntica de uso da língua alvo.

Na verdade, a autenticidade está sujeita às percepções que os envolvidos em uma situação de avaliação têm sobre o uso real da língua, e os elementos que a compõem. Daí, em uma situação de ocorrência de um teste, começamos primeiramente a observar e buscar compreender as instruções que são dadas para a realização das tarefas, ou para os itens dos testes.

2.3.3. Validade de face

Apesar de ser considerado como um aspecto menos científico e mais subjetivo em relação aos outros tipos de validade, a validade de face, também denominada de validade aparente, apresenta aspectos importantes num exame.

De acordo com Hughes (1994, p. 27), considera-se a validade de face de um teste quando ele parece mensurar o que deveria estar mensurando, e isso, inevitavelmente, envolve o julgamento, tanto dos elaboradores de testes quanto dos examinandos. Quanto mais validade de face tiver um teste, mais motivados os examinandos ficam, não só por se submeterem a ele, mas também por acreditarem nos resultados com um instrumento verdadeiro de avaliação.

pode-se dizer que a validade de face funciona como uma medida da aceitação de um exame e, se investigada adequadamente, pode demonstrar se os objetivos dos formuladores estão sendo atingidos junto aos examinandos e, também, o que os examinandos acreditam ser importante na realização do teste.

Muitas críticas sobre a validade de face versam sobre aceitar seu aspecto “atrativo” como evidência suficiente do valor e qualidade da avaliação, sem a devida verificação se o teste realmente dá conta de avaliar os construtos e conteúdos a que se propõe. Para esses teóricos, a validade de face se refere a um tipo de credibilidade apenas superficial, e estaria mais ligada a aspectos de aceitação social do teste do que a características técnicas que permitissem dizer se a avaliação é adequada ou não. (ARAÚJO, 2007:58-59).

Questionar o examinando sobre suas percepções sobre um teste é uma forma de ajudar examinadores a identificar pontos problemáticos em itens dentro de um teste, nos moldes em que esse teste está sendo administrado.

Nevo (1985:288) aponta que a validade de face pode trazer fatores que promovam uma atitude positiva em relação a um teste, pois há um envolvimento e motivação dos examinandos, atraindo-os para incentivar futuros candidatos ao teste. Além disso, a validade de face colabora para que os examinandos sintam-se menos insatisfeitos se tiverem um desempenho não tão satisfatório no teste. Desse modo, a validade de face seria um elemento de ligação entre os protagonistas.

Quando se questiona a opinião de pessoas que estão diretamente envolvidas com o teste, como os examinandos, e não somente a opinião de especialistas e pessoas envolvidas com a administração do processo avaliativo, desde a elaboração e aplicação da prova até a sua correção, pode-se dizer que esse tipo de validade também constitui uma característica importante do teste.

É lógico que não podemos utilizar somente a opinião dos examinandos ou examinadores e especialistas sobre a qualidade de uma prova. A aparência que um item do teste tem é apenas um dos aspectos avaliados no processo de validação de um teste, pois podemos também utilizar outros aspectos como a validade de construto e a validade de conteúdo, por exemplo.

Obviamente, deve-se concordar com os críticos da validade de face no que se refere a tomá-la como evidência única da qualidade do teste. A “aparência” do item é apenas um dos aspectos de sua validade, e é necessário estar atento a outras características no processo de validação de um teste. Avaliar adequadamente um

construto ou um conteúdo, complementado pela validade de face, é uma garantia a mais de que o processo de validação está sendo feito. Esse é o motivo pelo qual, neste trabalho, foram consideradas as validades de construto, de conteúdo e de face no processo de validação do teste classificatório.

Como base de pesquisa para este trabalho reunimos os dados resultantes dos testes de classificação aplicados aos alunos ingressantes desde o 2º semestre de 2010 ao 2º semestre de 2012. Além disso, foi pedido no 2º semestre de 2012, que os alunos ingressantes respondessem um questionário, denominado *Questionário II*, sobre a percepção do conteúdo do teste, que pode ser visualizado no Apêndice 1. Além desse questionário, foi também aplicado o *Questionário III*, respondido por dois docentes da área de inglês da instituição, onde o teste de classificação foi aplicado, e também apresentado no Apêndice 2, o questionário, e as respostas desses docentes estão apresentadas no Anexo 2. Os resultados desses questionários irão fornecer dados, indicadores da validade de face, que podem enriquecer a análise do teste para sua validação. Esses resultados serão discutidos no capítulo V, mais adiante.

CAPÍTULO 3 - AVALIAÇÃO EM LÍNGUA ESTRANGEIRA: O PROCESSO AVALIATIVO

3.1.O processo avaliativo: os testes e sua caracterização

A avaliação na área de Línguas pode ser feita de várias formas e por meio de diferentes métodos. Para cada período temos abordagens vinculadas a paradigmas diversos, contextualizados e definidos ao longo de cada um desses momentos. Toda avaliação, como já apontamos, está associada a um processo interpretativo de dados quantitativos e/ou qualitativos, pressupondo-se um juízo de valor que tem por finalidade diagnosticar os objetivos propostos no processo de ensino-aprendizagem.

Alderson²⁵ (2000: 27) argumenta que “(...) a sociedade precisa de avaliações. A avaliação é um evento diferente de muitas outras atividades, mas é valorizada pela sociedade por seus próprios méritos, e é por isso justificada”.

Na última década do século XX, com a busca da globalização, grandes transformações ocorreram no campo da avaliação e da aprendizagem de línguas, fazendo com que essa avaliação adquirisse grande importância no cenário educacional mundial e, também, no cenário brasileiro. Assim, de acordo com McNamara (2000), é de vital importância que estudos sobre testes e avaliações sejam realizados, pois testes são mecanismos para o controle institucional dos indivíduos e devem, portanto, estar comprometidos com a qualidade de seus itens.

Para a elaboração de um teste que seja coerente com os seus objetivos, é importante que tenhamos um bom conhecimento quanto às funções e tipos de avaliação. Na maioria das vezes, busca-se considerar em um teste todos os componentes da aprendizagem de línguas, como por exemplo: as habilidades específicas (ouvir, falar, ler, escrever); o conhecimento e uso de vocabulário; aspectos de pronúncia e gramática; aspectos culturais em situações realistas, contextualizadas e nas quais se promova a interação do examinando, entre outros. Apesar dessa pretensão em abranger todos os componentes da língua como elemento de comunicação e de interação social, em qualquer teste há sempre uma escolha e prioridade dada a certos aspectos da linguagem.

²⁵ Citação original em inglês: “(...) society needs assessment. Assessment is different from many other activities, but it is valued by society for its own sake, and is therefore justified”.(Alderson, 2000:27).

No entanto, dependendo do que se pretende avaliar, podem ser elaborados testes específicos, voltados para avaliar uma habilidade (oral ou escrita) ou testes para medir o conhecimento de vocabulário especializado de uma área, por exemplo.

Assim os exames normalmente são classificados em categorias como: entrada/classificação, rendimento, progresso/domínio, diagnóstico, proficiência.

- 1. Classificação** (*placement test*) – são utilizados para avaliar o nível de habilidade linguística do aluno de forma que possa ser colocado em um estágio de um programa de ensino apropriado as suas habilidades. É normalmente utilizado em universidades, escolas ou instituto de idiomas e a prova geralmente consiste de itens normalmente referentes ao material utilizado no curso, ou ao conteúdo programático que será desenvolvido. Muitos desses testes estão disponíveis no mercado, sendo que alguns acompanham o material didático que será escolhido pela unidade de ensino, e outros são elaborados pela própria instituição.
- 2. Rendimento** (*achievement test*) – são empregados para avaliar o nível de rendimento do aluno durante um curso. De acordo com Hughes (2003: 13), há dois tipos de testes de rendimento: os testes finais e os parciais ou progressivos. Os testes finais são aplicados ao término de um período letivo e têm por objetivo avaliar o desempenho do aluno de acordo com o que foi trabalhado durante o curso e, conseqüentemente, acabam por avaliar o aluno, o curso e a instituição. Os testes parciais ou progressivos avaliam o progresso que o aluno está tendo em determinados períodos do curso,
- 3. Diagnóstico** (*diagnostic test*) – tem por objetivo identificar áreas em que o aluno necessita ajuda extra. Podem ser gerais, para saber se os problemas do aluno concentram-se em uma ou mais habilidades, ou mais específicos, ou seja, relacionado a dificuldades com uma determinada habilidade, por exemplo.
- 4. Entrada** (*entrance test*) – buscam avaliar, a capacidade de uso da língua independentemente de como, onde ou em que circunstâncias ela foi

aprendida ou adquirida. São normalmente utilizados para o acesso em cursos cuja demanda é muito alta, como os vestibulares.

5. **Proficiência** (*proficiency test*) - estão baseados no uso futuro de determinada habilidade. Como já comentado anteriormente, o conceito de proficiência está relacionado a várias e diferentes concepções teóricas que determinam, por vezes, uma visão de linguagem para cada concepção. Podemos dizer que os testes de proficiência são elaborados para se medir o quanto um determinado indivíduo sabe sobre um assunto, no nosso caso, uma língua estrangeira, levando-se em conta sua capacidade de comunicação diante de uma determinada situação. Ao analisarmos um exame de proficiência, veremos que o mesmo tem como objetivo avaliar, para uso futuro, a capacidade de uso da língua independentemente de *como*, *onde* ou *em que circunstâncias* ela foi aprendida ou adquirida. Alguns exames de proficiência, a maioria deles produzidos na Inglaterra, têm a intenção de verificar se o aluno atingiu um determinado nível de habilidade geral de compreensão da língua; outros são elaborados com a intenção de verificar se o aluno tem competência suficiente para usar a língua em determinadas áreas profissionais como medicina, turismo, entre outras, Trata-se, no caso, de um tipo de Exame para Fins Específicos (ESP). Uma avaliação de proficiência precisa criar seu próprio conteúdo programático, presumidamente através de uma análise de uso da língua na situação alvo.

No caso específico deste estudo, conforme já salientado, foi analisado um teste de classificação, aplicado aos alunos ingressantes nos cursos de tecnologia. O teste de classificação analisado foi especialmente desenvolvido com o objetivo de fornecer informações que ajudem a posicionar o aluno em um determinado nível do programa de ensino da instituição, considerando-se o desempenho do aluno e as características do curso. Há, porém, vários tipos de testes de classificação já prontos no mercado. Entretanto, para Hughes (1989), os melhores testes de classificação são aqueles organizados pelas próprias instituições de ensino onde eles serão aplicados, pois o esperado é que sejam elaborados de acordo com as especificações desenvolvidas e exigidas pela instituição.

Spolsky (1978), em relação à avaliação, relaciona três períodos distintos, mas que se sobrepõem no campo da avaliação: o pré-científico, que está relacionado às tarefas relacionadas à gramática e tradução; o psicométrico-estruturalista, que surgiu sob a influência da abordagem áudio lingual e da linguística estrutural, com a utilização de testes de itens isolados (*discrete point items*) e, o sociolinguístico integrativo, com ênfase na comunicação, autenticidade e contexto.

Sob a influência da abordagem áudio lingual e a linguística estrutural, os testes foram modificados e potencializou-se o desenvolvimento de testes de linguagem objetiva com um trato estatístico, que ainda é um dos testes mais utilizados até hoje. Esses testes concentram-se em um determinado item isolado (gramatical, vocabulário, fonético, etc.) e são normalmente constituídos de uma alternativa que apresenta uma única resposta correta, e de outras com *distratores* que, na verdade, são alternativas com opções erradas para um determinado item.

Os testes objetivos são aqueles que exigem que o usuário escolha ou forneça uma resposta a uma pergunta cuja resposta é pré-determinada. O tipo de pergunta pode exigir que o aluno:

- selecione uma solução de um conjunto de opções (múltipla escolha, verdadeiro ou falso, correspondência);
- identifique um objeto ou posição;
- forneça breves respostas de um texto.

Os testes objetivos são especialmente adequados para determinados tipos de tarefas, principalmente pela sua praticidade de aplicação, uma vez que as perguntas podem ser projetadas para serem respondidas rapidamente; permitem também que os professores testem os alunos em uma ampla gama de conteúdos e forneçam aos estudantes *feedback* automático (com critérios variados como pontuação, sugestões, elogios e orientação). Sua correção é confiável e rápida; as respostas podem ser verificadas por pessoas sem nenhum conhecimento específico, a partir de um gabarito montado pelo examinador com as respostas corretas do teste. Não é necessária, nesse tipo de teste, a “mediação” de um ou vários corretores, como é o caso de testes com itens que exijam produção escrita, por exemplo. Além disso, para esses testes é possível

utilizar a tecnologia para a elaboração e correção, através de computadores e leitores ópticos. Mais ainda: esses testes permitem uma análise estatística do desempenho de cada aluno e as perguntas podem ser pré-testadas, a fim de se avaliar a sua eficácia e grau de dificuldade.

Uma das desvantagens de testes objetivos é que, diferentes dos testes subjetivos, não exigem a presença de um examinador, que emite um juízo de valor sobre a resposta dada pelo examinando, tanto escrita quanto oral. Esse fato pode, em certos casos, diminuir a confiabilidade de um teste. Nos testes objetivos os problemas existentes recaem sempre na formulação do item. Acredita-se que uma vez elaborado o item e as alternativas (a correta e os distratores) temos, na verdade, a interpretação da pessoa que elaborou o item. Nesse caso, devemos nos ater que em provas de interpretação de textos, não existe uma única leitura possível, além de que na construção de itens de múltipla escolha a construção de respostas corretas ou mesmo dos distratores podem tornar as questões muito fáceis ou muito complicadas, ambíguas, ou até mesmo de difícil solução. Por isso, durante o processo de elaboração do teste é importante e necessário um ou mais revisores para a apreciação dos itens.

Os testes objetivos podem ser de:

A) Resposta única

O item de resposta única enuncia o problema ou a situação-problema na forma de pergunta e apresenta as alternativas de resposta.

B) Afirmação completa

O item de afirmação incompleta apresenta o enunciado do problema ou situação-problema como uma afirmação a ser completada por uma das alternativas.

C) Resposta múltipla

O item de resposta múltipla apresenta uma situação contextualizada com afirmativas pertinentes a ela. Na sequência, enuncia o problema ou situação-problema na forma de pergunta ou afirmação incompleta e apresenta uma chave de respostas. Recomenda-se esse tipo de item quando se quer avaliar vários tópicos de conteúdo em um único enunciado.

D) Foco negativo

O item de foco negativo apresenta várias respostas corretas e apenas uma incorreta, que é a solicitada. Esse tipo de item deve ser evitado, segundo vários especialistas, a não ser que a negativa seja realmente o conhecimento a ser avaliado. Isso porque, a solicitação da resposta incorreta orienta o aluno para a busca do erro, e não do acerto.

E) Asserção ou razão

O item de asserção e razão apresenta duas afirmativas ou asserções que podem ou não ser proposições verdadeiras ou corretas, assim como podem ou não estabelecer relações entre si (causa e efeito, proposição e justificativa, princípio e justificativa, asserção e razão). Esse tipo de item é indicado para a avaliação de habilidades complexas.

F) Lacuna

O item de lacuna apresenta uma sentença com partes suprimidas para serem completadas com palavras ou expressões constantes das alternativas.

G) Interpretação

O item de interpretação é construído com base em textos, gráficos, tabelas, gravuras, fotografias e outros materiais para que o examinando faça interpretações, inferências, generalizações, conclusões ou críticas.

H) Associação

O item de associação (correspondência, emparelhamento, combinação ou acasalamento) apresenta elementos com alguma relação entre si e, por isso, podem ser associados de alguma maneira.

I) Ordenação ou seriação

O item de ordenação ou seriação apresenta elementos para serem ordenados segundo uma determinada lógica ou critério.

J) Alternativas constantes

O item de alternativas constantes (certo/errado; verdadeiro/falso; fato/opinião; sim/não) é bastante utilizado quando se quer avaliar um número significativo de conceitos, fatos e características.

A capacidade dos testes objetivos para avaliar uma ampla gama de conteúdos e, conseqüentemente, o seu aprendizado, é muitas vezes subestimada. Os testes objetivos são muito eficientes para examinar recordação de fatos, de conhecimento e aplicação de termos e, leitura de pequenos textos simples ou respostas numéricas. Entretanto, uma preocupação comum entre muitos avaliadores é a de que os testes objetivos não podem avaliar a aprendizagem além da compreensão básica. Alguns avaliadores, pelo contrário, acreditam que itens de testes objetivos, quando elaborados com imaginação, podem ser bastante desafiadores para os alunos e testar níveis mais elevados de aprendizagem.

Uma das grandes vantagens desse formato é permitir uma interpretação precisa, que pode levar a uma evidência importante da validade de conteúdo de uma prova. Há, no entanto, limites para o que as provas objetivas podem avaliar. Elas não podem, por exemplo, testar as habilidades do aluno para se comunicar, nem tampouco para construir argumentos ou oferecer respostas originais.

Todo e qualquer teste deve ser cuidadosamente elaborado de modo a evitar a “descontextualização” do conhecimento. Para assegurar que os testes apresentem a qualidade esperada, que sejam realmente instrumentos de medição de conhecimentos dos examinandos e, que possam também ser utilizados para tomada de decisões quanto às políticas públicas, determinados critérios precisam ser observados na elaboração e administração dos mesmos.

Fini (2005, p.104) apresenta alguns critérios que devem ser observados para a adequação de um instrumento de avaliação, começando pelo enunciado dos itens que deve:

- apresentar claramente um único problema proposto para o examinando;

- conter as informações essenciais para a solução do problema apresentado, evitando elementos supérfluos;
- ser adequado em relação à dificuldade pretendida;
- ser adequado em relação ao tempo disponível para a prova;
- ser adequado em relação à quantidade de tarefas a serem executadas para a escolha da alternativa;
- não conter informações preconceituosas, além de possibilitar a inclusão de elementos que se repetem nas alternativas, visando diminuir o tamanho da questão e tornar mais evidente o elemento variante que aparece nas alternativas.

Em relação às alternativas dos itens, Fini (2005, p.104) comenta que é necessário verificar se:

- a alternativa correta é sem dúvida alguma a única;
- as alternativas incorretas (distratores) representam relações possíveis de serem estabelecidas pelo examinando, mas não são condições suficientes para a resolução do item.
- são adequadas em relação ao tempo disponível para a prova;
- estão colocadas em ordem lógica, crescente ou decrescente, sempre que envolvem valores numéricos;
- são homogêneas no conteúdo, integrando ideias e conhecimentos;
- são homogêneas na forma;
- são independentes, sem subentendidos ou referências às alternativas anteriores;
- não contém “pistas” que possam ajudar o participante na resolução da questão;
- não contém elementos (pegadinhas) que possam induzir o examinando a erros;
- não contém certas palavras que induzem a afirmações falsas ou verdadeiras, utilizando palavras como “sempre”, “nunca”, “tudo” ou “todo”, “só” ou “somente” que são, em sua grande maioria, falsas. As que contêm “alguns” ou “geralmente” são quase sempre verdadeiras.

Ainda, quanto à redação dos itens, Fini (2005, pp.104–105) aponta que é preciso:

- estar redigida de forma clara e correta (ortografia, pontuação e gramática) evitando coloquialismos;
- ter textos que contenham informações pertinentes, coerentes. A escolha dos autores deve ser bastante criteriosa, uma vez que toda avaliação sinaliza para uma desejável apropriação do conteúdo;
- ter uma desejável apropriação de conteúdo;
- apresentar representações gráficas e/ou pictóricas na proporção correta, pertinentes e necessárias, com informação completa e boa visualização de legendas, incluindo a fonte original dessas representações;
- apresentar resposta a uma questão que não dependa da (s) resposta(s) de outra(s), para evitar a propagação de erros;
- que o conjunto das partes (situação-problema, questão e alternativas) apresente o nível de dificuldade (alto, médio, baixo) pretendido; e
- que a habilidade que se pretende avaliar com a questão esteja de fato contemplada.

Antes de tecer os comentários sobre o teste classificatório analisado neste estudo, alguns termos e conceitos básicos precisam ser esclarecidos. Primeiramente, a diferença entre o teste referenciado à norma (*norm-referenced test*) e o teste referenciado a critério (*criterion-referenced test*).

Em termos gerais, o teste referenciado à norma tem o propósito de medir as habilidades ou competências linguísticas gerais do examinando. Sua intenção não é a de verificar objetivos e determinar deficiências individuais, mas a de coletar informações que possam levar à tomada de decisões como a de posicionar alunos de acordo com sua pontuação em termos de um contínuo, de forma que aqueles que apresentem uma capacidade mais baixa fiquem em uma extremidade, enquanto os que possuem uma capacidade mais elevada fiquem na outra extremidade. Além disso, embora os examinandos possam saber, de antemão, o formato geral do teste (múltipla escolha, associação, preencher lacunas, falso ou verdadeiro, etc.), eles não sabem exatamente o conteúdo que será avaliado.

O teste referenciado a critério procura informar o examinando sobre seu resultado e se ele conseguiu atingir os objetivos propostos pelo teste, a partir de um nível mínimo de competência, que é denominado de critério, de forma que o aluno consiga visualizar seu progresso e o programa possa reestruturar pontos que não estejam funcionando bem.

3.2.O teste analisado e seus atributos

No teste de classificação analisado nesta pesquisa temos o que chamamos de teste indireto, no qual os examinandos são avaliados por componentes isolados da língua-alvo por meio de exercícios de múltipla escolha, para preencher lacunas e afins. Pelo fato de o teste de classificação ter sido desenvolvido com o objetivo de fornecer informações que ajudem a posicionar o aluno em um determinado nível do programa de ensino da instituição, considerando-se o desempenho do aluno e as características do curso, temos, também, no caso, um teste referenciado à norma.

Antes de procedermos à análise geral dos itens do teste, faremos um comentário sobre três aspectos que estão presentes no teste analisado, que dizem respeito à

avaliação de recepção escrita - leitura, à avaliação de recepção aural e à avaliação de produção escrita.

3.3. A Avaliação de recepção escrita e a habilidade de leitura

A avaliação da habilidade de ler textos em língua estrangeira tem despertado interesse de muitos pesquisadores na área de línguas. Muitos autores colocam a seguinte questão: a leitura em língua estrangeira é um problema mais próximo do escopo da leitura ou do escopo da língua? Segundo Alderson, (1984:24) “fica cada vez mais evidente que a leitura em L2/LE é um problema de língua para os níveis mais básicos de competência linguística na língua estrangeira”. Em pesquisas relacionadas ao termo, surgem duas tendências no modo de conceber o processamento da leitura: uma aborda a leitura como uma atividade subdividida em *habilidades e conhecimento*, levando em consideração os processos cognitivos subjacentes à atividade, enquanto a outra vê a leitura como *uso de estratégias*.

Bernhart & Kamil (1995), a partir da afirmação de Alderson (1984), desenvolveram pesquisas para investigar qual dos dois aspectos seria o mais adequado em relação ao desempenho em leitura em língua estrangeira, a partir de uma primeira hipótese, inicialmente denominada de “hipótese do curto-circuito” e posteriormente denominada de “hipótese do limiar linguístico”, que postulava ser necessário atingir um determinado nível de competência linguística na língua estrangeira para que se aprenda a ler em língua estrangeira. Entretanto, a grande dificuldade dessa perspectiva é que não é simples determinar qual seria esse nível limiar de proficiência em leitura. A segunda hipótese denominada de “Hipótese da Interdependência Linguística” – previa que o desempenho em leitura na língua estrangeira é compartilhado em grande escala com o desempenho em leitura na língua materna.

A leitura, seja em língua materna ou língua estrangeira, envolve o leitor, o texto, a interação entre leitor e texto, o conhecimento prévio – enciclopédico e linguístico – do leitor e o processamento cognitivo da informação linguística em vários níveis: ortográfico, fonológico, sintático e semântico, entre outros aspectos.

Geralmente, um leitor em língua materna tem um vocabulário vasto e conhece muitas palavras antes de começar o processo de letramento, enquanto um leitor de

língua estrangeira, na maioria dos casos, possui um vocabulário bastante limitado ao iniciar sua tarefa de leitura.

Apesar dessas dificuldades, as habilidades linguísticas de um leitor em língua materna podem perfeitamente ser transferidas para a leitura de um texto que exija um baixo nível de conhecimento linguístico por parte do leitor, em um assunto que seja familiar, escrito com palavras e estrutura simples, com certa semelhança às da língua materna do leitor.

Um dos objetivos da pesquisa em avaliação em leitura é explorar a natureza da dificuldade dos itens em um teste e as causas dessas dificuldades. No entanto, podemos dizer que as dificuldades em relação à leitura são relativas, uma vez que temos estudantes com habilidades e cognições diferentes. Perkins et al. (1995) desenvolveram pesquisa com dois grupos de estudantes. No primeiro, havia alunos que possuíam um bom nível de conhecimento da língua estrangeira e que conseguiam compreender informações implícitas no texto. O segundo grupo era composto por estudantes com pouco conhecimento da língua estrangeira e que não conseguiam compreender as informações implícitas no texto. Entretanto, os dois grupos de estudantes demonstraram competência em estruturas linguísticas que relacionavam partes do texto, independente de sua proficiência linguística. Esse resultado nos dirige para a “hipótese do curto-circuito”, já apresentada neste capítulo, que estabelece a relação entre habilidade de leitura e proficiência linguística.

3.4.A avaliação da compreensão oral ou auditiva

Especialmente durante as três últimas décadas, muitas pesquisas foram feitas em relação ao papel da compreensão oral ou auditiva na aquisição da linguagem. Ao mesmo tempo, essa habilidade começou a receber uma atenção mais sistemática nas salas de aula de línguas e uma grande variedade de livros, artigos e materiais destinados a ajudar os professores a desenvolver essa habilidade nos alunos, bem como uma variedade de metodologias baseadas em compreensão têm sido propostas (Anderson e Lynch, 1988; Buck, 2001). No entanto, apesar das tarefas de compreensão oral, especialmente desenvolvidas para o ensino de LE, estarem cada vez mais estudadas e incentivadas, as que aparecem nos testes de avaliação de compreensão auditiva recebem uma cobertura limitada na literatura da área de avaliação.

Buck (2001) explica que as atividades de compreensão oral devem ser vistas principalmente como um processo de construção de significado de um insumo auditivo, e que esse processo é mais do que somente um sinal auditivo. A compreensão oral é vista como um processo de inferência envolvendo a interação entre conhecimentos linguísticos e não linguísticos. O autor reforça ainda que a compreensão oral envolve o conhecimento de elementos essenciais como fonologia, vocabulário e sintaxe, porém vai além desses elementos porque envolve também interpretação.

A compreensão oral é feita automaticamente e em tempo real (os ouvintes raramente têm uma segunda chance de ouvir exatamente o mesmo texto), envolve conhecimento anterior sobre o assunto e algumas variáveis específicas do ouvinte (como um propósito para o ato de ouvir determinada fala, por exemplo) e, além disso, é um processo individual, sugerindo que quanto mais complexo é um texto, mais interpretações podem ser feitas sobre ele.

Na verdade, um dos problemas da avaliação de compreensão oral é que é muito difícil se ter um construto de um teste unicamente de compreensão oral que não exija o uso de outras habilidades linguísticas. Adicionalmente ao insumo auditivo, os examinandos precisam ler as instruções e questões do teste e, fornecer respostas auditivas e visuais para as questões. Conseqüentemente, o que era para ser um teste de compreensão oral acaba por avaliar outras habilidades linguísticas.

Para complicar, outros fatores, alguns deles ligados às características pessoais do avaliando, como a sua capacidade de memória auditiva, por exemplo, também influenciam e podem, assim, contribuir para um melhor ou pior desempenho no teste.

Sherman (1997) realizou uma pesquisa considerando algumas formas para se avaliar a compreensão oral de alunos. Na primeira experiência, os alunos liam as questões antes de ouvir o áudio; na segunda, eles ouviam o áudio, liam as questões e ouviam o áudio novamente; na terceira, eles ouviam duas vezes o áudio antes de ler as questões. Como resultado, a versão que teve um efeito significativo na produção de respostas corretas foi aquela em que os alunos ouviram o áudio pela primeira vez, apresentava-se as questões e eles ouviam o áudio pela segunda vez. Entretanto, não

ficou claro se a pontuação maior ocorreu por eles terem ouvido o áudio duas vezes ou pelo fato de as questões estarem intercaladas ao áudio.

Questionamentos sobre uma definição do que é autêntico ou válido para se avaliar a compreensão auditiva têm sido frequentes e, em diferentes períodos, posições diferentes têm sido tomadas dependendo da abordagem vigente em cada período.

Até os anos 60 do último século, um importante componente definidor da proficiência linguística era o conhecimento da pronúncia das palavras “com perfeição”, ou seja, igual à de um falante nativo da língua. A partir dessa década, no entanto, o ensino/aprendizagem e avaliação de línguas presenciou uma mudança de foco marcada pela ênfase na capacidade de usar a língua e conseguir ser compreendido. Na década de 70, com o paradigma do movimento comunicativo, o foco já estava centrado no uso da língua para a comunicação e essa visão acabou por sugerir novas formas de avaliação da compreensão oral em LE, voltadas principalmente para a recepção e compreensão global da mensagem.

Geralmente entendemos a proficiência em recepção oral de LE envolvendo a habilidade de processar acusticamente (e, também visualmente) o insumo e, usá-lo para construir um tipo de representação mental que pode ser a base para algum tipo de resposta oral ou escrita.

A partir de uma perspectiva sociocognitiva, Weir (2005:45) coloca o insumo acústico/visual como um dos processamentos fundamentais para a recepção oral (e que a literatura atual sugere como essencial). Outro processamento possível é o cognitivo, além de aspectos contextuais como o objetivo da tarefa, as restrições de tempo e as condições nas quais o teste é feito. Outros fatores como a velocidade de produção do discurso, as variações de sotaque, o uso de fone de ouvido durante o teste ou estar em uma sala com caixas de som são exemplos de interferências que podem afetar o processo de recepção e, portanto, nos resultados do teste.

Uma das discussões mais frequentes em termos de recepção oral é quantas vezes o insumo oral deve ser ouvido pela pessoa que fará um teste. A gravação deve ser tocada uma vez só, como tentativa de se demonstrar a autenticidade da ação, ou duas ou

mais vezes, sabendo-se que em um teste de recepção oral há um componente de artificialidade.

Normalmente em uma interação entre pessoas podemos ter um estímulo visual como gestos ou a possibilidade de pedir para que o falante repita a frase e até mesmo reformular alguns trechos que não foram bem entendidos, de forma que pode-se admitir a ideia de estar implícita “uma segunda chance” para se entender aquilo que foi dito.

Em alguns exames de larga escala como os de Cambridge, há a possibilidade, dependendo do tipo de tarefa, de se ouvir somente uma vez o áudio ou até, no máximo, duas vezes. Essas formas podem ser usadas dependendo de fatores tais como: a finalidade do teste, a consistência das tarefas, a amostragem e praticidade do teste, sendo que todos eles refletem a necessidade de se equilibrar considerações, às vezes incompatíveis, tais como a elaboração o design e a aplicação do teste.

Em alguns tipos de provas, como o IELTS ou o TOEIC, as questões de compreensão auditiva são reproduzidas uma única vez, porém há, pelo menos, uma intenção clara para isso: a possibilidade de um número maior de itens em um mesmo teste, e, portanto, mais dados de resposta para serem avaliados. Assim, a amostragem das questões e o tempo de duração da prova, por exemplo, são fatores determinantes para definir se o examinando deve ou não ouvir cada questão mais de uma vez.

Ao permitir que o examinando ouça duas vezes o áudio, por exemplo, temos como uma possível justificativa o fato de que ouvir um texto gravado no âmbito de um exame é claramente diferente da experiência de se ouvir algo no dia a dia. Há, no caso, fatores relevantes no contexto de uma prova que incluem a ausência de informações paralinguísticas e contextuais, o tempo necessário para se acostumar com o sotaque, a velocidade de fala do falante, muitas vezes diferente de uma situação face a face, bem como a sequência natural de audição a ser realizada, condicionada à tarefa pedida. Outra razão diz respeito à necessidade de garantir a consistência das tarefas e o nível da prova. Reproduzir o texto duas vezes é uma forma de minimizar problemas (como ruído, por exemplo) ou outras eventuais interferências ou transtornos que possam ocorrer durante a administração do teste.

Para além da questão de se reproduzir o áudio uma ou duas vezes, há, naturalmente, outras questões que têm impacto sobre a interação do examinando e a validade de contexto, ou de conteúdo, do teste. Essas questões incluem alguns aspectos já citados como a velocidade de fala e a variedade de sotaques, além do grau de convivência com a língua alvo, o número de falantes participantes e que podem ser distinguidos no áudio e o gênero a que pertencem, pois gêneros diferentes podem, quase sempre, trazer implicações para a concepção e formato nos testes de compreensão auditiva.

Nos últimos anos tem havido uma gradual mudança de opinião em relação à inclusão de diferentes sotaques em testes de compreensão auditiva. Há pelo menos duas décadas atrás, nos testes de compreensão auditiva, os sotaques encontrados eram predominantemente britânicos (em RP = Received Pronunciation)²⁶, ou o sotaque americano padrão (no caso de exame como o TOEFL).

Hoje em dia, com o mundo globalizado, e com o uso generalizado da língua inglesa, há uma maior disposição (e até incentivo) para a aceitação de sotaques provenientes de outros países que possuem o inglês como língua materna ou segunda língua, como a Austrália, Nova Zelândia, Canadá, País de Gales, Escócia, bem como o sotaque de falantes estrangeiros dessa língua. Tudo isso, sem esquecer padrões gerais, tais como o chamado *international english* também conhecido como *Global English*, *World English*, *Common English* ou *General English*, nomes utilizados para definir a língua inglesa que não é a língua nativa de ninguém, mas sim uma língua aprendida, conhecida e usada internacionalmente para a comunicação por todas as pessoas do mundo.

3.4.1 O impacto da tecnologia na avaliação da compreensão auditiva

A natureza e a qualidade do insumo acústico em testes de compreensão auditiva é um aspecto muito suscetível a mudanças, devido aos avanços da tecnologia nos últimos 50 anos. Até a década de 1950, o insumo acústico geralmente era feito através

²⁶ O termo “Received Pronunciation” tem sido usado há mais de 80 anos em referência ao sotaque do inglês falado na Inglaterra. Frequentemente é recomendado como a forma mais adequada do inglês britânico para a transmissão jornalística e como modelo tanto para instrução em língua materna quanto em segunda língua ou língua estrangeira. Os argumentos utilizados para seu uso como modelo estão fundamentados em duas alegações: (1) que foi mais cuidadosamente descrito do que qualquer outro sotaque e (2) que é amplamente entendido.

de uma pessoa lendo um texto para um grupo de alunos, submetidos a uma avaliação. Já nos anos 60, os gravadores começaram a ser usados para reproduzir diálogos, trechos de entrevistas, jornais falados, bate-papos, etc.

Na década de 1990, os avanços na tecnologia de áudio promoveram sensíveis melhorias nos testes de compreensão auditiva. As fitas de rolo utilizadas na década de 60 foram substituídas por fitas cassete nos anos 70 e alguns anos mais tarde por discos compactos (CDs) e, atualmente, por *flash drives* e fones de ouvido, o que ajudou muito na qualidade da recepção do som e nas formas de armazenamento.

Tais avanços tecnológicos têm implicações diretas para questões de validade de um teste. O uso de um material gravado em fita, CD ou no hardware de um computador apesar de padronizar a administração de uma prova e ser de grande auxílio para a confiabilidade do teste, não deixa de evitar outros problemas acústicos do ambiente durante a aplicação do teste. É bastante razoável especular que um candidato que ouve uma gravação através de fones de ouvido é susceptível a um padrão de escuta melhor do que aquele que ouve através de alto-falantes em um grande salão.

O uso da tecnologia, principalmente em testes por meios eletrônicos, também permite explorar e desenvolver novos itens em testes. Por sua vez, esses meios podem nos levar a rever e ampliar a compreensão do construto de testes de compreensão auditiva, por meio da inclusão de tarefas integrativas. Douglas (2007) defende o uso de Podcasts nos testes de compreensão auditiva em provas de Inglês para Fins Específicos, e argumenta que esses recursos estão se tornando cada vez mais populares entre os estudantes norte-americanos como um meio de interação social e estudos acadêmicos nos campi, e são, portanto, dignos de consideração.

3.5. A avaliação da produção oral

Questões sobre a natureza da proficiência oral e a melhor maneira de avaliá-la têm motivado muitas investigações na última década.

Acredita-se que os testes de produção oral são dos mais difíceis para elaborar e para responder. Apesar das grandes melhorias na área, ainda temos algumas questões que precisam ser melhores definidas quando falamos de avaliação da produção oral,

como por exemplo, a definição do construto, a descrição das tarefas, os efeitos do interlocutor sobre o examinando. A maioria dos problemas nesse caso tem a ver com problemas relacionados à validade e confiabilidade dos testes orais.

Shohamy (1988) aponta que somente a partir da década de 60 a avaliação oral começou a receber mais atenção. Os primeiros testes orais eram constituídos de repetição de palavras e frases, perguntas e respostas padrão e executados em laboratórios equipados com gravadores. Entretanto, a partir da década de 70 várias críticas começaram a ser lançadas sobre a artificialidade dos métodos de avaliação oral, uma vez que os examinandos repetiam frases como se fossem máquinas, não havendo interação com outras pessoas. É nessa época que foi introduzida a entrevista oral, com um examinador que fazia perguntas sobre tópicos diferentes a um examinando. As respostas dadas eram classificadas e inseridas em uma escala de notas. Assim era processada a avaliação.

Do mesmo modo que outros procedimentos de avaliação, a entrevista oral também tem sofrido várias críticas com o passar dos anos. Uma delas é que ela não se assemelha a um evento comunicativo real, possível de acontecer em situações naturais de interação no nosso dia a dia. Além disso, muitas das perguntas feitas ao examinando são planejadas com o objetivo de gerar amostras de competência do examinando e avaliar apenas a habilidade ou falta de habilidade do examinando em cumprir as tarefas de uso da língua propostas no teste.

Outro problema levantado por pesquisadores é que a entrevista oral procura medir a proficiência do examinando com base no falante proficiente em situações reais de comunicação, e a avaliação é feita baseada em grades de proficiência, que, na verdade, não estão ligadas à descrição de como se dá a comunicação natural. Paralelamente, são propostas listas de características do que um falante ideal utilizaria. A definição de falante ideal proposta nessas situações, por outro lado, ainda está por merecer maior consistência e embasamento teórico.

Pesquisadores como Shohamy (1988) e Bachman (1990) apresentam várias restrições à entrevista oral, concluindo que ela não é um instrumento válido de medição de habilidade oral, uma vez que no dia a dia outras formas de exposição oral são efetivamente utilizadas, tais como a narração, a conversa e a exposição oral.

Entretanto, há autores que defendem o uso de entrevistas para a avaliação da proficiência oral em língua estrangeira, como Molder e Halleck (1998) e apontam "que as entrevistas constituem eventos de fala autênticos e que ilustram comportamentos de comunicação verbal, mesmo que esses comportamentos não sejam exatamente iguais àqueles que ocorrem em conversas mais informais".

McNamara (1996:117), ao tratar desse tipo de testagem, assinala que toda entrevista oral apresenta julgamentos subjetivos, que envolvem atos de interpretação por parte do examinador e que inevitavelmente geram discordâncias. Como forma de evitar discordâncias, o autor propõe três procedimentos: o uso de grades de avaliação formuladas para cada nível de proficiência, incluindo exemplos ilustrativos das características de desempenho de cada nível; a avaliação de cada candidato mais de uma vez, com procedimentos para lidar com possíveis discrepâncias que possam surgir durante o teste; o treinamento cuidadoso dos examinadores no uso dos procedimentos de avaliação.

Entretanto, apesar do treinamento ser um dos aspectos importantes para o bom andamento do processo avaliativo, ele não é suficiente, especialmente se o sistema de pontuação também não for monitorado. Em muitos casos, a expertise do examinador parece não ter impacto na avaliação. Comparando-se examinandos que fizeram o teste por meio de gravações e aqueles que fizeram ao vivo verificou-se que os examinadores acabam por subestimar o desempenho de alunos mais proficientes quando somente o contato com o examinando é de forma gravada.

Além disso, embora gravações que não ofereçam uma boa qualidade de áudio tenham uma tendência para serem avaliadas de forma mais criteriosa, o desempenho daqueles examinandos considerados menos competentes são julgados com maior leniência.

Por isso, a grande preocupação tem recaído na grade de avaliação, descrevendo diferentes níveis de língua em forma de um contínuo, a partir de um nível de proficiência de nível limiar até o mais alto, com descritores precisos e claros, que não deixem dúvidas quando o examinador for fazer a avaliação do candidato. Para Fulcher (1996:228) "até que os pesquisadores estudem seriamente a validação das grades de avaliação dos testes, o problema da indeterminação e da não interpretação dessas grades vai permanecer".

Apesar da atividade de investigação que a avaliação oral tem recebido, o contínuo volume de trabalho na área indica que o construto para a avaliação de produção oral ainda não foi completa e devidamente entendido e estudado. Novas pesquisas, realizadas em conjunto com as áreas de linguística aplicada, análise do discurso e aquisição de segunda língua, têm se preocupado com o assunto.

CAPÍTULO 4 - METODOLOGIA DE PESQUISA

4.1. A pesquisa: um estudo de caso com propósito descritivo e exploratório

Normalmente as abordagens escolhidas para uma análise da pesquisa são um reflexo de valores culturais sobre o mundo e a época em que vivemos. As abordagens refletem, nesse sentido, o questionamento sobre as crenças e expectativas que temos e o método escolhido procura ser aquele que consideramos o mais correto e com garantia de melhores resultados. As abordagens escolhidas refletem, assim, uma conscientização da complexidade e contingências sócio históricas por meio das quais o conhecimento sobre o mundo e sobre nós é construído. Elas enfatizam a necessidade da ciência de ser auto reflexiva sobre suas limitações. Elas reconhecem, também, os contextos socioculturais, a inseparabilidade entre aquele que conhece e o objeto do conhecimento, o conhecimento e os meios de saber, a hipótese de um mundo objetivo e a existência da complexidade, a incerteza, a diferença e a heterogeneidade.

Sobre a escolha de um determinado método de pesquisa está implícita a escolha de um método adequado para responder ao questionamento que está sendo feito, o grau de controle que o investigador tem sobre o evento e o foco temporal. Assim, nesta pesquisa, dentre as possibilidades de escolha de estratégias para serem utilizadas, escolhemos primeiramente realizar uma pesquisa bibliográfica na área de Linguística Aplicada e na de Avaliação de Línguas, que resultou em levantamentos sobre os temas *testes de classificação e validação de testes*, ambos relacionados ao foco principal deste estudo. Além disso, feitas as escolhas da instituição e dos grupos a serem pesquisados, bem como a definição dos objetivos e perguntas da pesquisa, os procedimentos de pesquisa foram dirigidos para o levantamento e caracterização dos programas de ensino de língua inglesa da instituição e a apresentação do trabalho como um estudo de caso, com propósito descritivo e exploratório, desenvolvido a partir de análise qualitativa e quantitativa dos dados.

Segundo Yin (2005), o estudo de caso enfoca um fenômeno contemporâneo em contexto real, na qual o pesquisador e o objeto de seu estudo interagem de modo cooperativo e participativo para a resolução de um problema ou seu esclarecimento ou, ainda, propor soluções que ampliem o conhecimento sobre os fatos observados.

Creswell (1997, p.61) define o estudo de caso como “a exploração de um sistema limitado ou um caso (ou múltiplos casos) que envolve coleta de dados em profundidade e múltiplas fontes de informação em um contexto.” O autor ainda aponta que a noção de sistema limitado relaciona-se com a definição de tempo e espaço e que podemos entender o caso como um evento, uma atividade ou indivíduos. Ainda, segundo Yin (2005) as fontes de informação ou evidências são constituídas por entrevistas, observações, documentos e reportagens.

Stake (1995) sugere uma reflexão sobre as diferentes formas de se fazer um estudo de caso. Ao se buscar causas, há a predominância da abordagem quantitativa, sendo que a posição do pesquisador é descrever para inferir possibilidades de relações causais, ou seja, explanação e controle. Entretanto, a abordagem quantitativa busca compreender as inter-relações entre tudo o que existe.

O caso pode ser escolhido pelo seu caráter único, em que se requer um estudo intrínseco ao se buscar determinada proposição. Howe e Eisenhart (1990) sugerem 5 (cinco) passos possíveis de observação para julgar um estudo: (1) fidelidade à proposta realizada; (2) coerência e aplicação da técnica de coleta; (3) verificação de perguntas relacionadas aos pressupostos feitos diretamente ao entrevistado; (4) verificação dos resultados e confronto com as teorias utilizadas; (5) verificação do valor do estudo sob aspectos de informação e melhoria da prática.

Chizzoti (2006a p.102) apresenta três fases que delimitam o pesquisador em relação ao estudo de caso. A *primeira* compreende a seleção e delimitação do caso que será estudado. Para esta pesquisa a seleção e delimitação do caso foram determinadas pela escolha do tema, com foco na análise e validação de um teste de classificação/proficiência aplicado a alunos ingressantes em cursos superiores de tecnologia. A *segunda*, o trabalho de campo do pesquisador, que, neste estudo, refere-se ao processo de coleta e organização dos dados que nortearão as inferências e o relatório do caso, e, em *terceiro lugar*, a organização e redação formal dos resultados do estudo, que nesta pesquisa correspondem à análise dos dados obtidos, sua relação com as perguntas de pesquisa e pesquisas futuras sobre os temas estudados e a apresentação do trabalho final.

4.2.Os Instrumentos para a Coleta de Dados

Na primeira fase desta pesquisa, foi selecionado e delimitado o caso a ser estudado, ou seja, a validação do teste de classificação em língua inglesa aplicado aos alunos ingressantes de cursos superiores em tecnologia de uma determinada instituição de ensino superior. Para a validação, foram escolhidos o período de coleta dos dados e os participantes da pesquisa, descritos mais adiante. Além disso, foi também prevista para a validação uma revisão documental referente ao conteúdo apresentado nas ementas unificadas dos seis semestres da disciplina Inglês, oferecida para os cursos de tecnologia. Essa revisão teve como objetivo garantir a análise posteriormente desenvolvida com os itens dos testes cujo objetivo foi o de verificar se o que está previsto nas ementas se faz presente no teste e, em caso positivo, de que maneira é apresentado. Para uma visão abrangente dos temas *avaliação, ensino/aprendizagem de língua estrangeira e validação de testes* foram feitos levantamentos e leituras de trabalhos na área, de modo a garantir uma revisão da literatura pertinente a esses assuntos e a escolha de um aporte teórico para fundamentar o estudo da validação dos testes. A apresentação do referencial teórico utilizado no trabalho já foi desenvolvida nos capítulos 1 e 2.

Na segunda fase, que compreende a coleta e organização dos dados, foram selecionados alguns dados de um questionário *online* aplicado aos alunos pela instituição de ensino, juntamente com o teste *online*, denominado Questionário I, antes de eles iniciarem o teste de classificação. O nosso foco nesse levantamento foi buscar traçar um breve perfil escolar do aluno da instituição e dos cursos escolhidos para a coleta de dados: saber em que tipo de escola estudou (particular, pública), seu histórico escolar (o que foi cursado ao longo da sua vida escolar) e que contato(s) teve com a língua inglesa (onde estudou e se teve outro tipo de contato). Os resultados desse levantamento e a consequente caracterização do perfil escolar dos ingressantes na instituição de ensino estão detalhados e comentados na seção 4.4 do capítulo 4.

Para a complementação dos dados sobre a avaliação dos testes e sua validação, foram também aplicados dois questionários: um deles, em agosto de 2012, para os 141 (cento e quarenta e um) alunos ingressantes no 2º semestre de 2012, dos três cursos escolhidos, denominado Questionário II, e outro, para duas professoras de Inglês da unidade de ensino escolhida para a pesquisa, denominado Questionário III.

O Questionário II, que foi aplicado aos alunos que ingressaram em agosto de 2012, e foi aplicado logo quando os alunos concluíram o teste de classificação, é composto de 10 (dez) questões (vide Apêndice 1), detalhando aspectos da prova e de seu conteúdo para os alunos avaliarem, de modo a coletar dados para a validade de face e servir de exemplo para o desempenho dos alunos, conforme veremos mais adiante.

O Questionário III, composto de 25 (vinte e cinco) perguntas fechadas e 5 (cinco) perguntas abertas, foi respondido por duas professoras da instituição, que aplicam o teste de classificação aos alunos, buscando as considerações sobre o teste em geral e, os pontos positivos e negativos do teste percebidos pelas duas profissionais. Essa avaliação do teste tem também como objetivo institucional a sua análise e revisão, resultando em eventuais propostas de alterações consideradas como relevantes pelos profissionais de ensino.

A escolha dos múltiplos meios para o levantamento do corpus que são: o teste de classificação aplicado aos ingressantes nos cursos de Eventos, Logística (manhã e noite) e Gestão da Tecnologia da Informação no 2º semestre de 2010 ao 2º semestre de 2012, portanto 5 (cinco) amostras de resultados do teste, que compreende as habilidades de recepção oral e escrita, o questionário I respondido pelos examinandos pelo computador, em uma primeira etapa, antes de iniciarem o teste de classificação propriamente dito, que tem como objetivo identificar o perfil escolar desses alunos; o questionário II sobre a percepção do teste, que foi respondido por 141 examinandos no 2º semestre de 2012, após a realização da etapa feita no computador, e o questionário III respondido por duas docentes da área de inglês da instituição, deve-se à orientação de autores como Stake (1998) ao observar que diversas percepções para o levantamento de dados podem trazer mais clareza aos resultados e reduzir o risco de interpretações erradas, de modo a permitir olhar o caso de diversas maneiras.

Na terceira e última fase, os dados foram organizados, analisados e descritos segundo orientação e pressupostos pertinentes aos objetivos e perguntas de pesquisa e serão relatados em capítulo próprio, mais adiante.

4.2.1. Considerações éticas

Para esta pesquisa, todos os participantes - alunos e professores - receberam informações sobre o conteúdo deste trabalho e foi pedido o consentimento dos mesmos para a utilização dos dados dos questionários e das entrevistas.

Quando da aplicação do Questionário I, aplicado juntamente com o teste de classificação, conforme já apontado, consta um termo de consentimento de uso de voz e de uso de imagem, reproduzido a seguir:

<p>ATENÇÃO!</p> <p>ANTES DE CONCLUIR ESTA ETAPA LEIA OS TEXTOS ABAIXO:</p> <p>Termo de ciência e de consentimento livre e esclarecido Estou ciente de que este processo avaliativo tem como objetivo fazer uma classificação inicial do meu nível de conhecimento da língua inglesa. Como em qualquer avaliação, variações no resultado podem ocorrer, o que poderá implicar reclassificação.</p> <p><input type="checkbox"/> Ciente</p> <p>Autorização para utilização de imagem e voz Autorizo a utilização de minha voz e imagem, respectivamente, gravada e fotografada pela equipe de professores do NEPLE - Núcleo de Ensino-Aprendizagem e Pesquisa em Línguas Estrangeiras - em publicações e eventos científicos e acadêmicos de qualquer natureza, desde que sua veiculação não tenha fins lucrativos.</p> <p><input type="checkbox"/> Autorizo</p>

Quando da aplicação do questionário II, respondido após o teste pelos 141 participantes do 1º semestre de 2012, foi comunicado por escrito e verbalmente o motivo dos questionários serem aplicados e indicado que nos mesmos não iriam constar nomes, sobrenomes ou qualquer indicação do ingressante. Foi também apontado que a resposta e entrega do questionário II significava a ciência e aceitação de divulgação dos dados e resultados para fins acadêmicos dos questionários I (respondido antes do teste) e II (após o teste). Para nossa surpresa e satisfação, todos os ingressantes prontificaram-se a responder ao questionário II, entregando-os após o seu preenchimento. Nos questionários III, como eles seriam respondidos de qualquer forma, a entrega de cópia dos questionários respondidos pelas duas professoras foi considerada como aceitação da

divulgação dos dados e resultados para fins acadêmicos, uma vez que os mesmos esclarecimentos passados aos alunos foram também informados às duas docentes.

A seguir, descreveremos o contexto da pesquisa.

4.2.2. A coleta e os procedimentos de pesquisa: a análise quantitativa e qualitativa de dados

Ao se discutir a forma como podemos coletar dados para uma pesquisa, podemos destacar as pesquisas de caráter quantitativo e qualitativo na análise dos dados. Stake (1983a apud Vianna, 2000: 124) mostra que a partir da percepção de um fenômeno educacional, a pesquisa de caráter quantitativo:

...identifica algumas poucas variáveis ou propriedades; - seleciona as variáveis intuitivamente (redução e delimitação que são necessárias à pesquisa); - concentra-se nas variáveis ou propriedades que mostram grande variação; - realiza medidas e estabelece correlações procurando identificar padrões de covariação; - explica genericamente fenômenos, a partir da ideia que tem a respeito da população, selecionar uma amostra para estudo, compara medidas, faz análises estatísticas, procurando explicar a variação e covariação; - prepara tabelas e quadros para explicar essa variação covariação; - generaliza, formalmente, sobre a relação entre as variáveis estudadas, elaborando relatório. O produto é uma explicação que destaca as características das variáveis para a população. A generalização, desse modo, assemelha-se a leis. (STAKE, 1983a apud VIANNA, 2000: 124).

Para este estudo, conforme já apontado anteriormente, dois tipos de questionários foram aplicados aos alunos ingressantes. O primeiro deles foi aplicado a todos os alunos ingressantes nos cinco semestres de coleta de dados. Esse primeiro questionário, denominado Questionário I, foi feito *online*, antes dos examinandos fazerem o teste de classificação. Composto por 6 (seis) itens com cinco alternativas em cada um deles, o Questionário I procura verificar a aprendizagem do ingressante em relação à língua inglesa, com proposições relacionadas ao tipo de escola que o aluno frequentou durante o ensino fundamental e médio; de que forma o aluno entrou em contato com a língua inglesa; sua experiência com a mesma; como o aluno classifica as habilidades de ler, falar, ouvir e escrever em língua inglesa; se o aluno tem contato com a língua e se ele utiliza a língua para a comunicação oral e de que forma isso é feito. Entretanto, para este estudo, não utilizamos as respostas referentes aos 6 (seis) itens propostos no questionário. Utilizamos somente as respostas às perguntas 01 e 02 do

questionário, que permitem identificar o tipo de escola que o aluno frequentou durante o ensino fundamental e médio e de que forma o aluno entrou em contato com a língua inglesa, para traçar um breve perfil escolar dos alunos ingressantes. Os dados analisados neste trabalho referem-se aos testes realizados para os ingressantes no 2º semestre de 2010 até o 2º semestre de 2012, computando no total 5 (cinco) aplicações de exames e um corpus com aproximadamente 718 (setecentos e dezoito) participantes.

Para o Questionário II – Questionário de percepção da prova - composto de 10 itens e aplicado aos alunos ingressantes no 1º semestre de 2012, após eles terem feito o teste, o nosso interesse voltou-se para a análise de todas as respostas, uma vez que todas elas versam sobre a avaliação do teste e o seu conteúdo. Foi escolhida para avaliar o teste, no questionário II, a Escala de Likert, utilizada para avaliar produtos/serviços, na qual o entrevistado assinala um único item de acordo com seu grau de satisfação. Os alunos participantes indicaram em que graus concordam ou discordam das afirmações ali expressas, assinalando as respostas escolhidas com cinco opções que variavam de *discordo totalmente*, *discordo parcialmente*, *não tenho opinião*, *concordo parcialmente* a *concordo totalmente*. Os testes desse grupo também serviram de exemplo para indicar o desempenho dos alunos no teste.

O Questionário III – “Experts Post-Test Questionnaire”- foi aplicado a duas professoras de Inglês da instituição pesquisada que participaram como docentes da avaliação dos alunos ingressantes na instituição. O Questionário III é composto de 25 (vinte e cinco) itens fechados, utilizando também a escala Likert como modelo, com as mesmas cinco opções (*discordo totalmente*, *discordo parcialmente*, *não tenho opinião*, *concordo parcialmente*, *concordo totalmente*) e 5 (cinco) indagações abertas, que pretendem registrar o nível de concordância ou discordância com uma declaração dada, além de fazer uma análise do teste pelos seus aplicadores.

Os dados dos Questionários I, II e III estiveram sujeitos à análise de caráter quantitativo. O mesmo ocorreu em relação às respostas para os testes de nível 1 a 6 por todos os candidatos ingressantes. Os dados computados prestaram-se ao desenvolvimento da descrição e caracterização do corpus de pesquisa, dos seus participantes e de aspectos relacionados ao contexto analisado, especialmente no que diz respeito aos conteúdos de ensino avaliados nos testes. O computo dos dados e seus resultados serão apresentados e discutidos, prioritariamente, no próximo capítulo.

A pesquisa de caráter qualitativo na análise dos dados, por outro lado, compõe grande parte deste estudo. A partir da percepção de um fenômeno educacional, segundo Stake (1983a apud Vianna, 2000: 124) temos que esse tipo de investigação:

identifica eventos ou casos que sejam de interesse (pessoas, grupos, produtos, etc.); - considera propriedades ou variáveis, que, entretanto, não são o ponto nuclear do estudo, na verdade, momentos, lugares, pessoas são o centro do enfoque; - procura a singularidade individual, base para a compreensão do que e é típico e, depois, do geral; - identifica sequências importantes de eventos, testemunhas para eventos passados, e observa como os eventos são determinados; - registra e interpreta as observações, procura padrões de covariação, selecionando-os intuitivamente; - verifica com rigor as observações, validando seu significado via triangulação, ou seja, as observações são repetidas; - por múltiplos observadores, usando múltiplos métodos e múltiplas interpretações teóricas; - valida as observações em diferentes fases e momentos, à medida que o processo se desdobra; - elabora relatório final que é um retrato descritivo dos fenômenos observados, apresentando generalizações naturalistas (informais, explicações pessoais do mundo externo) e propicia aos usuários da pesquisa condições para que cheguem às suas próprias generalizações. (STAKE, 1983a apud VIANNA, 2000: 124).

Os dados de caráter quantitativo foram também considerados para auxiliar na composição e maior solidez do corpo descritivo e exploratório deste estudo. Além da investigação e análise em materiais institucionais, como as ementas de cursos e o regimento da instituição, do contexto institucional, foram também processadas as inferências e interpretações dos dados referentes a esse contexto, sua relação com o processo avaliativo para a classificação nos níveis de ensino de inglês desenvolvido pela instituição e a validação dos testes aplicados. Foi feita extensa revisão da literatura (em Linguística Aplicada e Avaliação, em especial), com o objetivo de elucidar questões a respeito do tema central da pesquisa *validação de testes classificatórios em língua inglesa* e outros temas paralelos, mas necessários para a composição de uma fundamentação teórica consistente, dada a importância e complexidade do tema avaliação. Assim, um apanhado histórico sobre as visões de linguagem no ensino de línguas, sobre a avaliação de línguas e sobre competência linguístico-comunicativa, acrescido da discussão de conceitos relacionados às práticas avaliativas, como: praticidade, confiabilidade, autenticidade, efeito retroativo e impacto, e de validade e validação de testes, já apresentados no capítulo 2, possibilitaram o registro e as interpretações feitas ao longo de todo este trabalho.

4.3. Descrição do cenário

Como dissemos anteriormente, para este estudo utilizamos os dados de uma única unidade pública de ensino superior tecnológico, distante 60 km da capital de São Paulo. Os cursos disponíveis nessa unidade são: Análise e Desenvolvimento de Sistemas, Eventos, Gestão Ambiental, Gestão da Tecnologia da Informação e Logística. Entretanto, os cursos de Análise e Desenvolvimento de Sistemas e Gestão Ambiental começaram a funcionar a partir do 1º semestre de 2011. O curso Informática para Gestão de Negócios, recebeu uma nova nomenclatura a partir de 2011, transformando-se em Gestão da Tecnologia da Informação.

Em todos os semestres são oferecidas 40 vagas para cada um desses cursos. A seguir, apresentamos a demanda de interesses para cada um deles, a partir do 2º semestre de 2010 até o 2º semestre de 2012. O interesse da comunidade em geral pelos cursos dessa unidade de ensino está representado pelo número de inscritos e pela porcentagem para cada curso (demanda) nas tabelas a seguir.

Tabela 1: Demanda dos Cursos – 2º semestre de 2010

Curso	2º sem. 2010	Período	Vagas	Inscritos	Demanda
Eventos		Manhã	40	105	2,63
*Informática Gestão de Negócios		Noite	40	255	6,38
Informática e Gestão de Negócios		Tarde	40	80	2,00
Logística		Manhã	40	166	4,15
Logística		Noite	40	305	7,63

Fonte: CEETEPS

Tabela 2: Demanda dos Cursos – 1º semestre de 2011

Curso	1º sem. 2011	Período	Vagas	Inscritos	Demanda
Análise e Desenvolvimento de Sistemas		Tarde	40	135	3,38
Eventos		Manhã	40	158	3,95
Gestão Ambiental		Tarde	40	185	4,63
*Gestão da Tecnologia da Informação		Noite	40	320	8,00
Logística		Manhã	40	190	4,75
Logística		Noite	40	353	8,83

Fonte: CEETEPS

*Mudança no nome do curso de: Informática e Gestão de Negócios para Gestão da Tecnologia da Informação.

Tabela 3: Demanda dos Cursos – 2º semestre de 2011

Curso	2º sem. 2011	Período	Vagas	Inscritos	Demanda
Análise e Desenvolvimento de Sistemas		Tarde	40	94	2,35
Eventos		Manhã	40	73	1,83
Gestão Ambiental		Tarde	40	66	1,65
Gestão da Tecnologia da Informação		Noite	40	281	7,03
Logística		Manhã	40	135	3,38
Logística		Noite	40	271	6,78

Fonte: CEETEPS

Tabela 4: Demanda dos Cursos – 1º semestre de 2012

Curso	1º sem. 2012	Período	Vagas	Inscritos	Demanda
Análise e Desenvolvimento de Sistemas		Tarde	40	147	3,68
Eventos		Manhã	40	113	2,83
Gestão Ambiental		Tarde	40	100	2,50
Gestão da Tecnologia da Informação		Noite	40	294	7,35
Logística		Manhã	40	173	4,33

Fonte: CEETEPS

Tabela 5: Demanda dos Cursos – 2º semestre de 2012

Curso	2º sem. 2012	Período	Vagas	Inscritos	Demanda
Análise e Desenvolvimento de Sistemas		Tarde	40	68	1,70
Eventos		Manhã	40	66	1,65
Gestão Ambiental		Tarde	40	49	1,23
Gestão da Tecnologia da Informação		Noite	40	223	5,58
Logística		Manhã	40	94	2,35
Logística		Noite	40	200	5,00

Fonte: CEETEPS

Segundo os dados apresentados acima, podemos perceber um aumento progressivo na procura pelos cursos noturnos nessa faculdade superior de ensino tecnológico no interior de São Paulo, o que vai ao encontro das estatísticas do Censo do Ensino Superior 2010.

Para fins de dados de análise neste trabalho, utilizaremos somente três cursos: Eventos - manhã, Gestão da Tecnologia da Informação (GTI) - noite e Logística – manhã e noite. Essa escolha dos cursos não foi aleatória e deu-se principalmente pelo fato desses cursos no período analisado não terem sofrido alterações no currículo durante os 5 (cinco) semestres em que se deu a coleta de dados.

O número de alunos ingressantes que fizeram o exame a partir do 2º semestre de 2010 até o 1º semestre de 2012 estão relacionados nas tabelas a seguir:

Tabela 6: Quantidade de alunos que fizeram o exame/ano – 2º semestre de 2010

2º sem. 2010 Curso	Quantidade de alunos
Eventos	29
GTI (noturno)	46
Logística (manhã)	41
Logística (noite)	42
Total	158

Fonte: CEETEPS

Tabela 7: Quantidade de alunos que fizeram o exame/ano – 1º semestre de 2011

2º sem. 2010 Curso	Quantidade de alunos
Eventos	69
GTI (noturno)	25
Logística (manhã)	35
Logística (noite)	37
Total	166

Fonte: CEETEPS

Tabela 08: Quantidade de alunos que fizeram o exame/ano – 2º semestre de 2011

2º sem. 2011	Quantidade
Curso	de alunos
Eventos	44
GTI (noturno)	34
Logística (manhã)	39
Logística (noite)	32
Total	149

Fonte: CEETEPS

Tabela 09: Quantidade de alunos que fizeram o exame/ano – 1º semestre de 2012

1º sem. 2012	Quantidade
Curso	de alunos
Eventos	38
GTI (noturno)	39
Logística (manhã)	26
Logística (noite)	38
Total	141

Fonte: CEETEPS

Tabela 10: Quantidade de alunos que fizeram o exame/ano – 2º semestre de 2012

2º sem. 2012	Quantidade
Curso	de alunos
Eventos	32
GTI (noturno)	25
Logística (manhã)	27
Logística (noite)	30
Total	104

Fonte: CEETEPS

Podemos perceber que apesar da faculdade oferecer 40 vagas para alunos ingressantes, em alguns cursos há mais de 40 (quarenta) alunos submetendo-se ao teste e, em outros cursos, há um número menor de alunos do que o número de vagas oferecidas. Essa variação explica-se por diferentes motivos. Em alguns casos, há alunos que já fizeram um curso superior e trazem para a faculdade seu histórico escolar, podendo aproveitar disciplinas que já foram cursadas e serem dispensados, inclusive em

inglês; em outros casos, há alunos que foram reprovados no semestre 1 e farão todas as disciplinas novamente, inclusive a disciplina Inglês; há também alunos transferidos de outras unidades, que fazem o exame para serem posicionados no semestre adequado; há ainda alunos que faltaram nos dias de aplicação do exame de classificação, ou mesmo vagas nos cursos que não foram preenchidas.

Surpreendentemente, no 1º semestre de 2012, escolhido como exemplo para a análise do desempenho dos alunos no teste de classificação, as quarenta vagas que estavam à disposição de cada curso não foram preenchidas. Nesse semestre, 141 (cento e quarenta e um) alunos participaram de teste para os cursos descritos no estudo de que trata este trabalho e responderam ao questionário II, como já assinalado anteriormente.

4.4. Descrição dos participantes

Os alunos que fazem parte desta pesquisa são oriundos de uma faculdade de tecnologia pública e que nela ingressaram através de um processo seletivo: o exame vestibular. Cinco cursos são oferecidos nessa unidade: Logística, Eventos, Gestão Ambiental, Análise e Desenvolvimento de Sistemas e Gestão da Tecnologia da Informação. Entretanto, para este estudo não utilizaremos os dados referentes a Gestão Ambiental e Análise e Desenvolvimento de Sistemas, pois esses cursos começaram a funcionar em 2011.

Sabemos que dados como gênero e faixa socioeconômica são importantes para uma análise do perfil dos participantes. Entretanto, para fins de dados estatísticos para esta pesquisa, nos fixamos apenas na escolaridade desses alunos e de que forma estão ou estiveram em contato com a língua inglesa, uma vez que optamos por delinear apenas um breve perfil escolar dos participantes e sua relação com a língua inglesa.

Os resultados obtidos com o Questionário I estão demonstrados nos gráficos a seguir, a partir do 2º semestre de 2010 até o 2º semestre de 2012, como amostragem para esta pesquisa. Em seguida, apresentamos os resultados de cada semestre e o correspondente às duas perguntas aplicadas aos alunos antes de iniciarem o teste *online*.

Tabela 11 – Tipo de escola que frequentou no Ensino Fundamental e Médio

No ensino fundamental e médio, você estudou em:					
	2º sem.	1º sem.	2º sem.	1º sem.	2º sem.
	2010	2011	2011	2012	2012
(1)Escola Pública	61%	61%	66%	56%	63%
(2)Escola Particular	17%	15%	19%	23%	18%
(3)Escola Pública e particular	17%	21%	12%	20%	18%
(4)Supletivos e/ou Educação de Jovens e Adultos (EJA)	1%	1%	1%	1%	1%

Fonte: Teste de proficiência

Como podemos observar pelos dados apresentados nesse levantamento, os alunos que fazem parte dessa faculdade são em sua grande maioria oriundos de escolas públicas, pois as porcentagens atingem aproximadamente 60% dos alunos (no 2º semestre de 2010 tivemos o equivalente a 61%; no 1º semestre de 2011 o equivalente a 61%; no 2º semestre de 2011 o equivalente a 66%; no 1º semestre de 2012 o equivalente a 56%; e no 2º semestre de 2012 o equivalente a 63%).

Tabela 12 - Contato e/ou experiência com a língua inglesa

Experiência com inglês					
	2º sem.	1º sem.	2º sem.	1º sem.	2º sem.
	2010	2011	2011	2012	2012
(1) Estudei somente na escola regular e/ou cursinho pré-vestibular.	72%	58%	65%	44%	48%
(2) Estudei em escolas de idioma.	55%	37%	31%	32%	29%
(3) Estudei fundamentalmente sozinho.	29%	16%	8%	10%	10%
(4) Aprendi no meu convívio com amigos ou no trabalho.	15%	9%	7%	9%	8%
(5)Morei/estudei em um país de língua inglesa.	0%	2%	1%	1%	1%
(6)Nunca estudei inglês.	7%	3%	4%	4%	4%

Fonte: Teste de proficiência

Em relação ao contato que esses alunos têm ou tiveram com a língua inglesa, o que imperou foi seu estudo na escola regular e/ou cursinho pré-vestibular. Ao questionarmos se os alunos já haviam frequentado cursos em escolas de idiomas, pudemos perceber que houve uma queda drástica sobre essa opção do 2º semestre de 2010 para o 2º semestre de 2012. No 2º semestre de 2010, 55% dos alunos responderam que tinham feito um curso de idiomas, enquanto que no 1º semestre de 2011 o número baixou para 37%, no 2º semestre de 2011 foi para 31%, no 1º semestre de 2012 para 32 e no 2º semestre de 2012 para 29%. Infelizmente não temos dados para justificar essa queda nos índices de interesse para fazer um curso de Inglês em um

instituto de idiomas. Podemos, entretanto, lançar algumas hipóteses, como: queda no poder financeiro da parcela da população que é oriunda de escolas públicas e que procura um curso superior em tecnologia ou dificuldades na aprendizagem na escola regular, o que levaria a uma falta de interesse em aprender um segundo idioma.

Os resultados do Questionário II denominado de “Questionário de percepção da prova” serão apresentados e discutidos no Capítulo 5 para ratificar as conclusões referentes aos dados da validade de conteúdo, bem como o Questionário III, denominado de “Expert Post-Test Questionnaire” que apresenta a visão de duas professoras que administram a disciplina Inglês na unidade de ensino.

A seguir, discorreremos sobre a aplicação dessa metodologia, a partir dos objetivos estabelecidos na pesquisa, seguindo os fundamentos discutidos neste capítulo, descrevendo a coleta e apresentando a análise desses dados.

CAPÍTULO 5 - CORPO DA PESQUISA E ANÁLISE DOS DADOS

No capítulo anterior estabelecemos a metodologia utilizada para esta pesquisa, o contexto da mesma, o cenário e os participantes. Apresentaremos neste capítulo, estruturado em cinco seções principais, a análise e discussão dos dados coletados.

Na seção 5.1 serão apresentados os documentos de pesquisa que serviram de base para a discussão dos dados, ou seja, as ementas das disciplinas de Inglês unificadas para todos os cursos, reunidas de 1 a 6 (cada qual correspondente a um semestre do curso de Inglês), o teste de classificação aplicado aos alunos ingressantes de todos os cursos envolvidos na pesquisa e os questionários I, II e III, os dois primeiros aplicados aos alunos e os dois últimos aplicados aos alunos ingressantes do 1º semestre de 2012 e a dois professores, respectivamente, elaborados para colher dados sobre a percepção desses indivíduos a respeito do teste. Na seção 5.2 será analisado o construto que delinea os passos para a elaboração do teste, a partir das ementas do curso de inglês. Na seção 5.3 estarão expostos os itens que compõem o teste e a análise dos mesmos. Na seção 5.4 uma breve análise da avaliação da produção oral dos alunos ingressantes será apresentada, apesar dessa habilidade não ser o foco principal deste estudo. Finalmente, no item 5.5, será desenvolvida a apresentação e análise dos dados coletados nos questionários respondidos pelos examinandos e professores, de modo a caracterizar a validade de face no processo avaliativo.

5.1. Documentos de Pesquisa

Para respondermos às perguntas de pesquisa propostas, foram selecionados dados de fontes que pudessem ajudar a investigar o objetivo principal deste estudo: a validação do teste de classificação de inglês em três cursos superiores de tecnologia. As fontes para essa investigação foram as ementas das disciplinas de Inglês 1 a 6, compostas do objetivo da disciplina nos 6 (seis) semestres do curso, além do teste de classificação que é operacionalizado na faculdade de tecnologia escolhida para este estudo. Para a construção dos itens do teste de classificação não foi utilizado como base nenhum livro didático, uma vez que não há um critério de adoção de material apesar de constarem sugestões de bibliografias básica e complementar nas ementas, O professor tem a opção de adotar qualquer um dos livros sugeridos nas bibliografias, ou até mesmo

outro material de sua preferência, contanto que contemple o que está estabelecido nos objetivos e no conteúdo programático dos cursos. A investigação sobre o processo de avaliação foi complementada pelos dados coletados em questionário respondido pelos alunos e por professores, apresentados no item 5.5 deste capítulo.

5.1.1. As ementas das disciplinas Inglês 1 a 6

As ementas das disciplinas Inglês 1 a 6 enfatizam o emprego da oralidade, porém compreendem as habilidades de compreensão oral e leitura e, expressão oral e escrita. Contemplam também as especificidades acadêmico-profissionais do aluno para cada nível de Inglês e aspectos socioculturais da língua inglesa. O conteúdo programático não é descrito por meio de tópicos gramaticais, mas sim por tarefas que o aluno deve ter a capacidade de realizar, sempre contemplando as habilidades de recepção/produção orais e escritas.

5.1.2. O teste de classificação

Como mencionado anteriormente, a visão de linguagem que norteia o teste de classificação, objeto de estudo deste trabalho, está baseada em Clark (2000: 49), que focaliza o ensino e a aprendizagem numa perspectiva sociocultural afirmando que “a linguagem é usada para se fazer coisas”. Portanto, um falante de uma língua estrangeira, para ser proficiente precisa mais do que conhecer a estrutura da língua; precisa saber usá-la em situações reais de uso dessa língua, como apresentar-se, pedir e fornecer informações, anotar recados. Dessa forma, é necessário que o teste avalie esse conhecimento.

Nesse estudo foi feita a comparação de cada item que compõe o teste com o conteúdo especificado nas ementas. Dessa forma, foi verificada a relevância do conteúdo que está sendo utilizado no teste. Entretanto, nem tudo que estava contido nas ementas foi contemplado no teste, bem como alguns itens não previstos nas ementas foram testados, desviando-se, assim, da intenção prevista na avaliação.

5.2. O construto do teste

Conforme já salientado, para esta pesquisa foi avaliado um teste de classificação para alunos ingressantes em três cursos superiores de tecnologia e para a definição do construto do teste a análise das ementas dos 6 (seis) semestres de aprendizagem de língua inglesa foi desenvolvida. No quadro 7 (sete) estão apresentadas as ementas que compõem o curso e em seguida os comentários sobre o construto que delinea o teste.

Quadro 9 – Ementas do Curso de Inglês

	Ementa
Inglês 1	Introdução às habilidades de compreensão e produção oral e escrita por meio de funções sociais e estruturas simples da língua. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos sócio-culturais da língua inglesa.
Inglês 2	Consolidação da compreensão e produção oral e escrita com a utilização de funções sociais e estruturas simples da língua desenvolvidas na disciplina Inglês 1. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos sócio-culturais da língua inglesa.
Inglês 3	Expansão da compreensão e produção oral e escrita com a utilização de funções sociais e estruturas básicas da língua. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa.
Inglês 4	Consolidação da compreensão e produção oral e escrita com a utilização de funções sociais e estruturas básicas da língua desenvolvidas na disciplina Inglês 3. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa.
Inglês 5	Aprofundamento da compreensão e produção oral e escrita com a utilização de funções sociais e estruturas mais complexas da língua. Ênfase na escrita e na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa.
Inglês 6	Aprimoramento da compreensão e produção oral e escrita com a utilização

	de funções sociais e estruturas mais complexas da língua desenvolvidas na disciplina Inglês 5. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa.
--	---

Em todas as ementas, com pequenas variações como o uso das palavras *introduzir, consolidar, expandir, aprofundar* e *aprimorar*, temos como objetivo principal do ensino de língua inglesa “ a oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa”. A escolha de uma abordagem comunicativa promovendo a interação entre os alunos para se engajarem em atividades contextualizadas, integrando as quatro habilidades.

No entanto, para fins avaliativos, apresenta um instrumento de avaliação, que é o teste de classificação, cujo construto de uso da linguagem está limitado ao reconhecimento de regras gramaticais. Com base em itens isolados, o teste algumas vezes apresenta itens que procuram avaliar a compreensão de textos, porém utilizando-se do texto como pretexto para a verificação do conhecimento de tópicos gramaticais. Nesse instrumento revela-se um conceito de proficiência cuja característica principal é ser capaz de memorizar significados de palavras, sem se levar em conta o uso efetivo real da linguagem.

Scaramucci (2000) relaciona o termo proficiência aos objetivos e padrões definidores utilizados para avaliar um candidato em um dado contexto, correspondendo, portanto, às especificações do teste sobre o que se espera que o candidato faça na língua alvo. Assim, o teste tem o propósito de medir a habilidade linguístico-comunicativa em língua inglesa, e sua capacidade de comunicação em situações de uso específico da LE, baseando-se apenas nas ementas dos 6 semestres do curso de inglês, sem levar em conta como ou onde esse examinando a aprendeu.

A partir das ementas apresentadas podemos apreender a definição do construto do curso fundamentada na recepção e produção oral e escrita da linguagem, com ênfase na oralidade, na próxima seção faremos a análise do conteúdo do teste, comparando os

itens do teste com os tópicos das ementas para verificar se há convergência do construto das ementas no teste.

5.3. ANÁLISE DOS ITENS DO TESTE

A ementa é um documento curto, que contempla os objetivos do curso de Inglês com foco nas quatro habilidades: ouvir e compreender, falar, ler e escrever em língua inglesa, seguindo uma abordagem comunicativa. A indicação dessa abordagem foi por nós inferida a partir da menção de *funções sociais e aspectos socioculturais da língua inglesa*, uma vez que essa informação não está explicitada nas ementas.

Os tipos de itens abordados no teste são: itens gramaticais, situacionais (que compreendem questões relativas às situações reais contextualizadas), identificação de erros/acertos, diálogos incompletos, itens relativos à tradução da língua materna para a língua alvo e itens relacionados a entendimento de textos, tanto escritos como aurais.

Como mencionado anteriormente, para fazermos a validação de conteúdo, utilizaremos a ementa e os itens que compõem o teste para cada nível, sendo que as questões propostas compreendem questões de leitura e de recepção oral. Começaremos com a ementa relacionada a Inglês 1, conforme o quadro 8 (oito) e itens retirados do teste. Analisaremos primeiramente as questões de recepção escrita de cada nível da prova, para na seção 5.4 analisarmos as de recepção oral, para em seguida, na seção 5.5 analisarmos as de produção oral. Todas as 60 questões que compõem o teste de classificação mediado pelo computador podem ser consultadas no Anexo I.

5.3.1. Nível 1

O nível 1 do teste é composto de 10 (dez) itens, sendo 8 (oito) itens de leitura e 2 (dois) itens de recepção oral.

No quadro abaixo descrevemos a ementa e seus objetivos, e os tópicos da ementa que estão contemplados no teste.

Quadro 10: Ementa da Disciplina Inglês 1

INGLÊS 1		
EMENTA	OBJETIVOS	ITENS
Introdução às habilidades de compreensão e produção oral e escrita por meio de funções sociais e estruturas simples da língua. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos sócio-culturais da língua inglesa.	O aluno deverá ser capaz de:	
	- Compreender instruções, informações, avisos, textos curtos e descrições de produtos.	X*
	- Apresentar-se, fornecendo informações pessoais, cotidianas e corporativas.	1, 4, 5, 6
	- Descrever locais e pessoas.	X
	- Preencher formulários com informações pessoais e profissionais.	X
	- Dar e anotar recados.	X
	- Utilizar números em contextos diversos para anotações de horários, datas e locais.	9
- Entender diferenças básicas de pronúncia.	9, 10	

*X representa objetivo não avaliado pelos itens.

Observando-se o quadro acima, que faz uma comparação entre os tópicos listados na ementa para o ensino de Inglês 1 e os itens escolhidos para compor o teste de nível 1 podemos observar que alguns tópicos da ementa não foram contemplados no teste, como: compreender instruções, informações, avisos, textos curtos e descrições de produtos; descrever locais e pessoas; preencher formulários com informações pessoais e profissionais; e dar e anotar recados. Esses tópicos estão marcados no quadro com um X.

Analisando-se agora o teste verificamos que nos 10 itens do nível 1, composto pelos itens de números 1 a 10, há 6 itens gramaticais, 2 itens com textos incompletos,

cujas respostas precisam ser escolhidas nas alternativas apresentadas e 2 itens de recepção oral.

Na verdade, a ementa não prevê o ensino de itens gramaticais. Apesar disso temos 6 itens que focam no assunto. Os itens 1 e 7 estão relacionados ao reconhecimento do uso do *Simple Present Tense*; o item 2 ao uso do verbo *there to be*; o item 3 ao uso do verbo *to be* e o item 8 ao uso do modal *can*, sendo que nesse item é apresentado em português, é também exigida a versão do português para o Inglês.

Embora a avaliação deva estar orientada de acordo com a abordagem de ensino de um curso, infelizmente, em muitos casos, não há uma correlação entre elas. No caso específico deste teste de classificação apesar da ementa de Língua Inglesa sugerir que o curso está ancorado em uma abordagem comunicativa, promovendo a interação entre os alunos para o engajamento em atividades contextualizadas, relacionadas à área de trabalho e integrando as quatro habilidades, o teste apresenta um construto de linguagem que prioriza um (re)conhecimento de regras gramaticais. No entanto os enunciados dos itens utilizam situações nas quais se pressupõe um vocabulário que tangencia os tópicos discriminados na ementa. Por esse instrumento podemos revela-se um conceito de proficiência cuja característica principal é dominar regras gramaticais, sem se levar em conta o uso real da linguagem.

A seguir apresentaremos cada um dos itens do teste acrescidos dos comentários que subsidiam a análise.

Item 1

Em cada questão há somente uma resposta certa.

1) Escolha a melhor alternativa para completar o diálogo entre um professor e seus alunos, no primeiro dia de aula.

Teacher - GOOD MORNING, CLASS. NICE TO MEET YOU. I'M YOUR ENGLISH

TEACHER. AND YOU, WHAT DO YOU DO?

Student A – I AM A FULL-TIME STUDENT.

Student B - I _____ WITH INFORMATION TECHNOLOGY.

Teacher – WHAT ABOUT YOU? ...

- a. not work
- b. are work
- c. work not
- d. work

O item 1, apesar de referir-se a uma situação real de sala de aula, na qual o professor se apresenta aos alunos, prioriza um conhecimento essencialmente gramatical. Poderíamos dizer até que o texto representa uma situação autêntica, que reproduz o uso real da língua, porém, o item utiliza como meio para a orientação da tarefa um contexto autêntico, para avaliar, na verdade, o conhecimento do presente do verbo. As alternativas apresentam uma forma afirmativas e três negativas do verbo que funcionam como distratores.

Esse tipo de teste, chamado de teste de item isolado aparece no período psicométrico-estruturalista, e era muito utilizado por levar a uma maior objetividade na correção e confiabilidade, e dificilmente apresenta alteração dos escores. Entretanto, ao se fragmentar a língua em pequenas unidades, as características da língua em uso para comunicação podem ficar comprometidas e não serem efetivamente avaliadas, nesse sentido o examinando necessariamente não precisa saber comunicar-se na língua alvo para responder ao que é pedido.

Além disso, o item pode ser respondida por temos três alternativas negativas e somente uma afirmativa. Um aluno mais atento, sem ter um grande domínio da língua, poderia optar pela letra **d**, por ser ela a única alternativa diferente das demais. A letra **d** poderia ser a escolha mais óbvia para esse aluno.

Outra possibilidade de resposta ao item seria optar pela alternativa **a**, que garante por meio do processo da tradução um resultado satisfatório. Porém estaria incoerente com a estrutura gramatical da língua inglesa, diante da exigência de um verbo auxiliar. As alternativas **b** e **c** também poderiam ser traduzidas, e para aqueles que não conhecem a estrutura gramatical da língua inglesa aceitas como possíveis. Por outro lado, o aluno precisaria prestar atenção ao conteúdo do texto e observar que a lacuna a ser preenchida, por coerência, deveria estar na forma afirmativa.

Como exemplo do desempenho dos alunos no teste apresentamos uma tabela com as respostas fornecidas pelos examinandos ao fazerem o teste no 1º semestre de 2012. Podemos observar pela tabela 15 (quinze) , a quantidade de alunos que fizeram a opção por cada alternativa do item 1.

Tabela 13: Amostragem das respostas ao item1 – Nível 1

Item 1	Alternativas			
	A	B	C	D
Curso				
Eventos	4	6	3	25
Logística(manhã)	2	2	5	17
Logística (noite)	2	6	1	29
GTI	2	X	X	37
Total	10	14	9	108

Podemos observar que dos 141 alunos que participaram da prova, 108 alunos responderam corretamente ao item 1, apontando a alternativa **d** como correta. Temos portanto aproximadamente 77% dos alunos, ou $\frac{3}{4}$ deles sem dificuldade para resolver o item. Somente 23% deles, ou $\frac{1}{4}$ não souberam respondê-la. A grande incidência de acertos parece dever-se ao fato do item envolver um conhecimento gramatical, frequentemente trabalhado nas escola de ensino fundamental e médio.

A seguir, os itens 2, 3, 5, 7 e 8.

Item 2

Assinale a melhor alternativa para completar a frase abaixo.

Currently _____ 46,000 students at the 49 FATECs in São Paulo State.

- a. there was
- b. are
- c. there is
- d. there are

Item 3

Escolha a melhor alternativa para completar a frase abaixo.

Felipe and Mariana _____ classmates. They both study Computer Database.

- a. are
- b. is
- c. was
- d. am

Item 5

Escolha a resposta mais apropriada para a seguinte pergunta.

What do you do in your free time?

- a. I watched TV.
- b. I travel on business.
- c. Sometimes I read magazines.
- d. I always work late.

Item 7

Qual seria a pergunta para a seguinte frase?

Eduardo works with International Relations.

- a. What does Eduardo do?
- b. When does Eduardo work with International Relations?
- c. Does Eduardo work overtime?
- d. Why does Eduardo work?

Item 8

Como você diria que sabe falar inglês?

- a. I can speak English.
- b. I speaking English.
- c. I am speak English.
- d. I can talk English.

Nos exemplos 2, 3, 5, 7 e 8 observa-se que o teste também avalia itens gramaticais isolados, sendo que o conteúdo dos mesmos não está contemplado na ementa. Percebe-se que nesses itens a preocupação maior não está na compreensão da frase, mas prioritariamente de que forma a frase pode ser completada para que haja concordância entre os seus . Temos neste caso uma visão estruturalista da língua que pressupõe prioritariamente a composição do texto ou da frase por pequenas partes.

A seguir os itens 4 e 6.

Item 4

Situação: O coordenador está apresentando o professor de Inglês aos alunos.

Como os alunos o cumprimentariam?

Coordinator – This is Antonio, your new English teacher.

Students - _____.

- a. See you soon.

- b. Good to see you again.
- c. It's nice to meet you.
- d. Very good.

Item 6

Em uma entrevista de emprego, qual seria a melhor resposta para a pergunta abaixo.

- Do you have any experience in finance?

- a. Yes, I do. I can answer the phone.
- b. Unfortunately I don't.
- c. No, I am sorry but I can't.
- d. No, I'm not.

Os itens 4 e 6 apresentam uma determinada situação, na qual o examinando deve assinalar qual seria a melhor resposta. Esse tipo de item tem o objetivo de avaliar o entendimento do examinando sobre o uso de determinadas expressões que são usadas em conversas habituais.

Nos itens 4 e 6, as habilidades de ler o texto e inferir a melhor resposta para a pergunta é uma tarefa que tem um propósito comunicativo. A melhor resposta para o item é a alternativa **b**, pois o aluno precisaria estar atento ao contexto da pergunta. O aluno para resolver este item de forma satisfatória deve ter o domínio da competência linguística, isto é, gramatical e lexical, bem como da competência discursiva ambas já apresentadas no capítulo 2.

Podemos observar pela tabela 16 a resposta dos alunos a cada uma das alternativas do item 6.

Tabela 14: Amostragem das respostas ao item 6 – Nível 1

Item 6	Alternativas			
Curso	A	B	C	D
Eventos	5	7	13	13
Logística(manhã)	5	4	7	10
Logística (noite)	3	4	18	13
GTI	5	2	24	8
Total	18	17	62	44

Pelos resultados referentes às respostas assinaladas pelos alunos na tabela 16, observa-se que dos 141 alunos que fizeram o teste, somente 17 optaram pela alternativa **b**, que é a correta, representando 12% de acertos para esse item. As alternativas mais selecionadas foram respectivamente a alternativa **c** com 62 respostas ou 43% de escolha e a alternativa **d** com 44 respostas ou 31% de escolha. A alternativa **a**, que também teve mais escolhas do que a alternativa **b** que é a correta, foi assinalada por 18 alunos, que corresponde a 13% das opções feitas pelos examinandos.

Podemos concluir que esse tipo de item apresenta certa dificuldade para os alunos, uma vez que estão acostumados a fazer testes que verificam o conhecimento da gramática da língua. Muitos deles ao concluírem o ensino médio parecem ainda não possuir a habilidade de reunir várias competências para a resolução de uma mesma tarefa.

A seguir temos exemplos dos itens 9 e 10 do nível 1 do teste de classificação que têm como propósito a avaliação da compreensão auditiva do examinando.

Item 9

As questões seguintes referem-se a uma gravação.

Áudio: audio1.wma

THE GIRL IS FREE AFTER _____.

- a. 9:00
- b. 8:30
- c. 5:00
- d. 12:30

Item 10

Ainda em relação ao áudio acima:

In the end, the boy and girl decide to _____.

- a. attend a concert.
- b. attend a show.
- c. watch a movie.
- d. stay in the park near school.

Audio script:

A: Are you free on Friday? There's a concert in the park near school.

B: Really? What time?

A: It's at 2:30.

B: I'm sorry. I'd love to go, but I'm busy until 5 o'clock.

A: Oh, that's too bad. Maybe we can go to the movie on Saturday night, instead.

B: Great! What do you want to see?

A: How about the new Tom Cruise movie?

B: Sure! We can go to the 7:30 show.

Para resolver os itens 9 e 10 o examinando deve ser capaz de entender o essencial da mensagem. Os detalhes relevantes são referentes ao uso de números em contextos diversos para anotações de horários, datas e locais, um assunto que está contemplado na ementa. Além de ser uma situação possível e corrente no dia a dia. Pode ser considerada uma questão autêntica, pois além de contextualizada, é um tópico significativo para o aluno em situações de uso da língua.

Para que o examinando entenda o diálogo na questão, precisa prestar atenção à sequência de fatos que estão sendo expostos e ter conhecimento de números. Trata-se um item que envolve a organização de ideias, e exige o conhecimento da preposição “*after*”, pois é a partir dela que o item será respondido.

Uma grande porcentagem dos examinados que fazem um curso de Inglês em institutos de idiomas devem ter se sentido muito a vontade com esse tipo de item. Isso acontece principalmente por essas escolas priorizarem a recepção e produção oral em língua inglesa, o que gera a crença de que comunicar-se em inglês significa entender o que está sendo falado e falar a língua, com fluência e precisão. Assim por priorizarem a oralidade o desempenho das habilidades de leitura e escrita não é considerados por eles como o conhecimento efetivo de uma língua.

A seguir apresentamos uma amostragem do desempenho dos examinandos no item 9 do teste nos cursos analisados.

Tabela 15: Amostragem das respostas ao item 9 – Nível 1

Item 9		Alternativas			
Curso	A	B	C	D	
Eventos	3	X	34	1	
Logística(manhã)	6	4	16	X	
Logística (noite)	3	13	22	X	
GTI	2	2	35	X	
Total	14	19	107	1	

Pelos resultados demonstrados na tabela 17 (dezessete), referentes às opções feitas pelos alunos dos 3 (três) cursos que compõem o corpo desta pesquisa, podemos verificar que dos 141 alunos que fizeram o teste, 107 alunos, ou 75% deles optaram pela alternativa **c** como resposta ao item, sendo esta a resposta correta. Verificamos também que somente 1(um) aluno escolheu a alternativa **d** como correta, sendo que a hora apresentada na alternativa não foi mencionada no áudio. A alternativa **b** foi escolhida por 19 alunos, o equivalente a 13% do total, assim como 14 (quatorze) alunos ou 10% do total optaram pela alternativa **a**, escolhas totalmente aleatórias, uma vez que os dois horários não foram mencionados no áudio.

Ao se analisar a ementa visualizamos que a mesma enfatiza o uso da oralidade, atendendo as especificidades de cada curso, entretanto o teste, neste primeiro nível, só apresenta dois itens de compreensão aural, sendo que em nenhum momento o aluno é convidado a interagir com o teste, uma vez que neste primeiro nível não há uma prova de produção oral, e nem mesmo de produção escrita.

Podemos perceber que essa diferença demonstra quão distinta é a avaliação e o ensino, ou pelo menos, o que está descrito na ementa e o que está sendo feito em forma de avaliação. Encontramos, na verdade, um desencontro do que seja ensinar uma língua e de como avaliá-la.

Ao indiretamente postular um ensino embasado no uso das quatro habilidades, podemos deduzir que o conteúdo do programa envolve a competência gramatical, a sociolinguística, a discursiva e a estratégica; todas elas relacionadas às teorias desenvolvidas por Canale & Swain (1984), e ampliadas por Bachman & Palmer (1990) e outros, que fazem parte do que foi exposto no Capítulo 2 deste trabalho.

5.3.2. Nível 2

A seguir faremos a análise do Nível 2, referentes às questões 11 a 20 do teste.

Quadro 11: Ementa da Disciplina Inglês 2

INGLÊS 2		
EMENTA	OBJETIVOS	ITENS
Consolidação da compreensão e produção oral e escrita com a utilização de funções sociais e estruturas simples da língua desenvolvidas na disciplina Inglês 1. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos sócio-culturais da língua inglesa.	O aluno deverá ser capaz de:	
	Comunicar-se utilizando frases simples em contextos pessoais e profissionais.	11,12
	Pedir e dar permissão.	X*
	Falar sobre o trabalho.	X
	Fazer comparações.	13
	Falar sobre eventos passados.	15, 16, 17
	Atender a uma ligação telefônica e anotar recado.	19
	Compreender gráficos e tabelas.	X
	Redigir correspondências rotineiras.	X
	Extrair informações de textos técnicos específicos da área.	14
	Entender diferenças de pronúncia.	19, 20

*X representa objetivo não avaliado pelos itens.

Ao se fazer uma análise sobre o nível 2(dois) do teste de classificação, podemos observar a partir do quadro 9 (nove), que abarca os conteúdos a serem desenvolvidos pelos docentes durante o 2º semestre de um Curso Superior de Tecnologia, que vários

conteúdos não estão explicitamente trabalhados no teste, como: pedir permissão, falar sobre o trabalho, compreender gráficos e tabelas, redigir correspondências rotineiras e extrair informações de textos técnicos específicos da área.

Uma amostra de todos os itens do teste pode ser vista no Anexo 1(um). A partir de uma análise de todos alguns itens do teste, podemos conferir que cada item apresenta um tópico de aprendizagem, entretanto, é importante que cada item do teste represente, pelo menos, um tópico de aprendizagem que está listando na ementa.

Assim, faremos comentários sobre os itens que estão contemplados na ementa e no teste de forma a verificar a pertinência deles em relação à validade de conteúdo.

Apesar da ementa não afirmar que estratégias de leitura o examinando deve utilizar para realizar o teste, podemos inferir pelo tipo de questões que o teste apresenta que as principais seriam:

- a) identificar a ideia principal do texto, compreendendo os propósitos implícitos e explícitos;
- b) identificar palavras/frases em um texto ou um diálogo que são inadequadas gramaticalmente em relação ao contexto;
- c) determinar respostas adequadas dependendo da situação apresentada.

Começaremos comentando os itens 15, 16 e 17, pois eles tratam do mesmo tópico gramatical: *simple past tense*.

Item 15

Sandro's dad is a director now. Ten years ago _____

- a. he was a manager.
- b. he will be a manager.
- c. he is going to be a manager.
- d. he is a manager.

Item 16

Leia o texto abaixo e responda qual afirmativa sobre a vida de Denise é verdadeira.

Denise was born in Sorocaba, moved to Indaiatuba when she was two years old. She studied Business Management at FATEC. She graduated in 2007 and then

worked for a multinational logistics company for 2 years. In 2009 she went to Canada to study English.

- a. She worked in a multinational after she traveled abroad.
- b. She studied French before she graduated.
- c. She has never traveled abroad.
- d. She traveled to Canada two years after she graduated.

Se considerarmos que a ementa do curso não trata em nenhum momento de tópicos gramaticais, constatamos que os itens simulam um diálogo, entretanto exige do examinando um conhecimento gramatical, e, na verdade está avaliando gramática e não a compreensão de leitura.

No item 15 temos o início da afirmação com o verbo no *simple present* e para completar a frase o examinando necessita saber o significado para palavra *ago* para identificar a alternativa que se encontra no *simple past*, uma vez que as alternativas apresentam verbos no *simple future*, *going to* e presente.

No item 16, temos um texto, que deveria servir ao propósito de entendimento, entretanto temos novamente um item gramatical, pois o examinando precisaria distinguir entre *simple past* e *present perfect tense*.

Item 17

Assinale a pergunta para a resposta abaixo.

_____?

Denise studied Business Management at FATEC from 2005 to 2007.

- a. What does Denise study?
- b. Where does Denise study?
- c. Where did Denise study?
- d. Why did Denise study Business Management?

O item 17 também verifica o conhecimento gramatical do examinando, uma vez que para ele fazer uma pergunta, a partir de uma resposta dada, ele deve distinguir o verbo auxiliar de presente (*does*) e o verbo auxiliar de passado (*did*), além do reconhecimento de pronomes interrogativos como: *what*, *where* e *why*.

Voltando nossa atenção para os itens 13 e 14, temos a apresentação de dois textos, sendo que o primeiro tem como título: *An Indian Culture Overview*, e o segundo: *A Chinese Culture Overview*.

An Indian Culture Overview	A Chinese Culture Overview
Population – 1,189,172,906 (July 2011 est.)	Population – 1,336,718,015 (July 2011 est.)
Area - 3,287,263 sq km	Area - 9,596,961 sq km
Official Languages – English, Hindi and 14 other official languages (2001 census)	Official Languages – 7 main Chinese languages including standard Chinese or Mandarin, Cantonese, and Wu.
Life expectancy at birth - 66.8 years	Life expectancy at birth - 74.68 years
Internet users - 61.338 million (2009)	Internet users - 389 million (2009)
Exports – commodities: petroleum products, precious stones, machinery, iron and steel, chemicals, vehicles, apparel	Exports – commodities: electrical and other machinery, including data processing equipment, apparel, textiles, iron and steel, optical and medical equipment
Agricultural products - rice, wheat, oilseed, cotton, jute, tea, sugarcane, lentils, onions, potatoes; dairy products, sheep, goats, poultry; fish	Agricultural products - rice, wheat, potatoes, corn, peanuts, tea, millet, barley, apples, cotton, oilseed; pork; fish

(Source: The World Factbook)

Item 13

Leia as informações abaixo e assinale a alternativa correta.

- Life expectancy in China is higher than in India.
- China is less populous than India.
- India is bigger than China.
- China is smaller than India.

Item 14

Das afirmações abaixo, assinale a única INCORRETA em relação às informações dos países acima.

- Both countries export steel and iron.
- India doesn't export any beef.
- There are more users of the Internet in China than in India.
- China exports fruit.

Esses textos trazem dados estatísticos relacionados à população da Índia e da China, áreas dos países, línguas oficiais, expectativa de vida, uso da internet, produtos agrícolas. Entretanto os textos são utilizados para uma verificação gramatical envolvendo o uso de comparativos, utilizando os adjetivos: *higher than*, *populous than*, *bigger than*, *smaller than*.

Podemos constatar dois aspectos na “inadequação” desse tipo de item, apesar da sua pertinência como “elemento motivador para a leitura”. Chama a atenção o título, que traz a expressão *culture overview* que, a princípio pode ser vista como um *distractor*, uma vez que os dados apresentados na tabela são prioritariamente dados geopolíticos dos países. Em segundo lugar, mais uma vez, a compreensão do texto é medida a partir da exigência do conhecimento do comparativo pelo examinando, fundamentalmente. Nesse sentido, o texto é mal explorado e parece ter sido utilizado apenas como pretexto para uma questão gramatical.

Em relação ao item 14 uma primeira observação diz respeito à cobrança do que está incorreto, exigindo do examinando “pensar a língua ao avesso”. No entanto, constatamos que apesar de o propósito desse item estar na avaliação da habilidade de entendimento do texto, ele pode ser respondido corretamente apenas por meio de comparações entre as palavras que aparecem nas alternativas e as que estão no texto. Observando o texto, a resposta correta para o item seria a alternativa **d**, uma vez que a China não exporta nenhum tipo de fruta.

Outro fato que podemos assinalar é que o examinando conta com a maioria das palavras que são similares às do português. A avaliação de leitura está no nível da decodificação das palavras, não necessitando que o examinando pense em um sentido pleno de leitura como construção de sentidos.

A seguir os itens 11, 12 e 18.

Item 11

Que frase completaria o e-mail abaixo?
Glenn, I called Rodrigo about the meeting.
He can't make it on Thursday, but Friday
morning is fine for him.

Regards,
Simone

- a. The meeting is canceled.
- b. Can I confirm it?
- c. Rodrigo can't meet on Friday.
- d. The meeting is on Thursday.

Item 12

Como você diria a seu professor, de uma maneira educada, que quer saber suas notas?

- a. Give me my grades.
- b. I'd like to check my grades.
- c. I need my grades.
- d. You have to give me my grades.

Item 18

Situação: Sofia, a nova estagiária da FATEC, conversa em casa sobre as suas funções

rotineiras no trabalho. Leia o texto abaixo e escolha a alternativa que continue o diálogo.

- Mom, my new job is great.

- Really! What do you have to do there?

- _____

- And how do you like it?

- It's fun.

- a. I park and wash the cars.
- b. I help the teachers and coordinators.
- c. I teach the students and prepare lessons.
- d. I plan lessons and correct tests.

Os itens 11, 12 e 18 são itens que envolvem diálogos incompletos, com diferentes expressões e tópicos, porém não apresentam uma correlação com o que é solicitado pela ementa.

De acordo com a ementa esses itens deveriam avaliar a habilidade do examinando em entender e saber utilizar expressões do dia a dia, como: pedir permissão, falar sobre o trabalho, compreender gráficos e tabelas, redigir correspondências rotineiras. Entretanto, esses conteúdos não estão contemplados nos itens. Além disso, o item 18 apresenta problemas ao pedir que o aluno apresente uma resposta para a continuação de um diálogo. A questão é mal formulada, pois mais de

uma resposta pode ser aceita. A exigência, no caso, está condicionada ao conhecimento bastante aprofundado das funções de um estagiário. Dentre as várias possibilidades de resposta, o examinando fica bastante confuso, o que prejudica o seu desempenho no teste.

A seguir apresentamos os itens 19 e 20 do nível 2 referentes à avaliação de compreensão auditiva.

As questões de 19 a 20 referem-se a uma gravação.

Item 19

Qual é o assunto da conversa telefônica?

- a. Uma negociação de compra e venda.
- b. Uma anotação de recado.
- c. Um pedido para deixar o endereço eletrônico.
- d. Uma gravação que foi deixada na secretária eletrônica.

Item 20

Qual é o número de telefone de Andrea Hann?

- a. 619-555-2437
- b. 609 –555–2427
- c. 609– 555-3437
- d. 609 –555–2477

Audio script

A: Carston Company. This is Michelle.

B: Hello, I'm looking for Patrick James, please. The last name is J-A-M-E-S.

A: He's stepped out. Would you like to leave a message?

B: Yes, please. Tell him that Andrea Hann called. That is A-N-D-R-E-A-H-A-N-N. I'd like him to call me back. It will be at 619-5552437, until 3 o'clock. He can call me at that number.

A: I'll give him the message.

B: Thank you.

Os dois itens que compõem o nível 2 do teste de classificação são adequados e coerentes com o que a ementa prioriza. As informações que estão sendo solicitadas nos itens, inclusive, podem ser facilmente identificadas e entendidas. O examinando deve saber identificar números, o que pode ser constatado no diálogo que apresenta um número de telefone para ser anotado para que um dos interlocutores retorne a ligação.

Além disso, por tratar-se de situação de rotina, o evento comunicativo foi bem escolhido, além de concordante com o que se revê na ementa.

Em relação ao vocabulário, as palavras são facilmente reconhecíveis e presumidamente do domínio de alunos que frequentam cursos superiores. Para que o examinando possa responder ao item 19, esse precisa reconhecer a palavra “*message*”, um cognato com a palavra mensagem do português. Um dos objetivos do item é o de analisar se o examinando tem a capacidade de distinguir informações mais importantes no texto das informações de menor importância para a sua compreensão. Avalia-se, também, o conhecimento mínimo do alfabeto em Inglês e a soletração das letras, conteúdo bastante trabalhado nos níveis mais básicos de ensino da língua.

Observamos, portanto aspectos bastante positivos em relação a esses dois itens de compreensão oral do nível 2: não oferecem dificuldade ao aluno e são representativos de uma linguagem realista e comunicativa de uma situação de interação em um ambiente de trabalho.

5.3.3. Nível 3

A seguir apresentamos a ementa do nível 3 de Língua Inglesa, representados pelos itens 21 a 30 e a análise dos itens do teste que apresentam concordância e discordância com a ementa.

Quadro 12: Ementa da Disciplina Inglês 3

INGLÊS 3		
EMENTA	OBJETIVOS	ITENS
Expansão da compreensão e produção oral e escrita com a utilização de funções sociais e estruturas básicas da língua. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e	O aluno deverá ser capaz de:	
	- participar de discussões em contextos sociais e empresariais usando linguagem apropriada de polidez e formalidade.	21, 23
	- Expressar opiniões e necessidade.	22, 25

abordando aspectos socioculturais da língua inglesa.	- Fazer solicitações.	25
	- Descrever habilidades, responsabilidades e experiências profissionais.	X*
	- Usar números para descrever preços, dados e gráficos.	X
	- Compreender informações de manuais, relatórios e textos técnicos específicos da área.	24, 26, 27
	- Aperfeiçoar a entoação e o uso dos diferentes fonemas da língua.	X

*X representa objetivo não avaliado pelos itens.

A partir de uma análise dos itens que compõem o teste, observamos que os conteúdos programáticos, que compõem a ementa referentes a descrever habilidades, responsabilidades e experiências profissionais; usar números para descrever preços, dados e gráficos e aperfeiçoar a entoação e o uso dos diferentes fonemas da língua não estão contemplados no teste.

Podemos observar que não são todos os itens do teste estão inseridos no conteúdo programático da ementa, entretanto, é importante que cada item do teste represente, pelo menos, um tópico de aprendizagem que está listando na ementa.

A partir do nível 3 as instruções para os itens não estão mais escritas em língua materna, porém em língua estrangeira.

Apresentamos a seguir os itens 21, 22, 23 e 25.

Item 21

You are in a meeting with a client and you need a repetition about an explanation. What would

- you say to the client?
- Repeat the explanation.
 - Could you repeat the explanation?
 - Do you repeat the explanation, please?
 - I want you to repeat the explanation

Item 22

After 2 hours studying English with a friend, you are exhausted. How would you *suggest* that you need a break?

- We must have a break now.
- I can have a break now.
- We should have a break now.
- Stop now.

Item 23

Choose the best option to fill the gap:

Hi, Pedro.

Just a short note to let you know that my printer is not working.
_____ by email.

Could you please print it for me?

Thanks, Rosa.

- I am sending the report
- I would send you the report
- I had sent the report
- I was sending the report

Item 25

Choose the best option to complete the instructions about a tour to the museum.

The bus will leave at 9:00 and get to the museum by 11:00. We plan to return at 4:00 in the afternoon and be back here at 6:00. _____ or enough money to get something to eat.

- You mustn't bring a snack
- You might bring a snack
- You can't bring a snack
- You should bring a snack

Os itens 21, 22 e 25 estão contemplados na ementa e pedem que o avaliando aponte de que forma se expressaria diante da situação sugerida pelo texto. Para isso, o examinando deveria saber diferenciar o uso de verbos modais para pedir algo de forma polida, utilizando o verbo “could”, deveria saber diferenciar o uso de must, can't, should, dentro do contexto sugerido pelo item.

Item 24

Read the text below and choose the best alternative.

All about Alice

Alice is an innovative 3D programming environment that makes it easy to create an animation for telling a story, playing an interactive game, or a video to share on the web.

Alice is a teaching tool for introductory computing. It uses 3D graphics and a drag-and-drop interface to facilitate a more engaging, less frustrating first programming experience.

(from <http://www.alice.org/>)

Alice is a 3D programming environment created to

- a. teach programmers how to make 3D graphics.
- b. teach people how to use computers in an easy way.
- c. teach people how to program experiences.
- d. teach advanced students how to work with 3D graphics.

Os itens 24, 26 e 27 são aqueles que têm a intenção de avaliar a habilidade de interpretação de textos em Inglês e que, de acordo com a ementa, avaliam a capacidade do aluno para compreender informações de manuais, relatórios e textos técnicos específicos da área. Não podemos dizer que os textos são específicos da área, uma vez que alunos de vários cursos participam da prova. Entretanto, apresentam um contexto possível para o perfil dos vários cursos de tecnologia e configuram uma tentativa adequada para avaliar a interpretação de textos curtos, prevista na ementa.

Item 26

Read the paragraph below and answer questions 26 and 27.

Choose the best alternative to complete the text.

When conducting a study on computers, whether you are working on the data for a class or for market research, you may wish to find out the number of personal computers per household. As computers have become more affordable, their use _____ more

widespread. Determining the number of PCs per household can help you understand the current state of the consumer market.
 (from http://www.ehow.com/how_7602206_determine-number-pcs-per-household.html)

- a. had become
- b. has becoming
- c. became
- d. has become

Um problema assinalado no item 26 é que o texto serviu como pretexto para uma questão gramatical, uma vez que o examinando tem de ter o conhecimento dos tempos verbais oferecidos como opção para completar a lacuna. Neste caso, o examinando precisa, antes de tudo, saber a gramática da língua, tratando-se, portanto, de uma avaliação centrada na língua e não na leitura. Nota-se que as alternativas tentam reproduzir pequenas partes do texto, de ordem puramente linguística, focando elementos isolados. Temos, então, uma avaliação muito mais pautada no reconhecimento de estruturas do que na compreensão.

Item 27

"Determining the number of PCs per household can help you understand the current state of the consumer market" (lines 4, 5 and 6) means

- a. The current state of the consumer market can be determined by the number of PCs per household.
- b. Families buy PCs to help the consumer market.
- c. You can determine how many families use PCs in their houses with the help of the current consumer market.
- d. The number of PCs per household is determined by the help of the current state of the market consumer.

O item 27 apresenta uma oração do texto e pede para que o aluno o seu significado, distinguindo qual seria a opção mais próxima da frase original. O examinando, no caso, somente precisa “pinçar” a melhor resposta por meio de uma comparação entre as palavras empregadas na frase original e as utilizadas nas alternativas. Esta questão limitou-se, portanto, em avaliar a sintaxe e o léxico,

desconsiderando seu funcionamento no texto como um todo. Neste caso, a leitura não é vista como uma processo ativo, mas como uma tarefa de decodificação.

A seguir os itens 28, 29 e 30.

Item 28

As questões de 28 a 30 referem-se a uma gravação.

Áudio: audio3.wma

Choose the correct alternative.

- a. Mr. James wants to go to Taipei tomorrow.
- b. Mr. James wants to leave a message.
- c. Mr. James is calling for a new internet connection.
- d. Mr. James wants to rent a car for 5 days.

Item 29

When will James leave the hotel?

- a. June 30
- b. June 28
- c. June 24
- d. June 22

Item 30

The dialogue is probably between:

- a. A receptionist and a guest.
- b. A husband and a wife.
- c. A boss and a secretary.
- d. A saleswoman and a client.

Audio script

A: Carston Company. This is Michelle.

B: Hi, Michelle. It's Patrick James.

A: Hello, Mr. James.

B: I need you to do some things for me today.

A: All right.

B: I'm travelling to Taipei next week. I need a hotel reservation for 5 nights starting on Sunday June 23.

A: OK.

B: I want a hotel with fitness center and room service. Make sure the room has an internet connection and I want a non-smoking room.

A: Of course.

B: I'll need a rental car too. I want a compact car, nothing big. Try to get an automatic transmission.

A: OK. What days will you need that for?

B: The pick up day will be the same, Sunday June 23. The drop off day will be Friday June 28.

A: Anything else? B: Yes, I need a hair cut and a massage. Please see if I can get an appointment this afternoon. A: OK. B: Are there any messages? A: Yes, one from Andrea Hann.

Ao se analisar as questões de recepção oral que constam no nível 3 do teste, observamos que os itens 28, 29 e 30, poderiam estar abarcados na ementa ao se pedir que o avaliando seja capaz de participar de discussões em contextos sociais e empresariais usando linguagem apropriada de polidez e formalidade e compreender informações de manuais, relatórios e textos técnicos específicos da área.

O objetivo das questões de recepção oral neste nível é avaliar se o examinando apresenta a capacidade de entender a ideia central do diálogo e as informações adicionais que apoiam a informação central. Entretanto percebemos que a avaliação da recepção oral está voltada para o processamento da memória e não necessariamente na compreensão do que está sendo tratado como objeto da avaliação.

O diálogo entre um executivo e sua secretária que é a base para as questões 28 a 30 apresentam muitos detalhes, não pedidos nos testes. O examinando pode perder o fluxo das informações ao prestar atenção aos detalhes e deixar de compreender o restante do texto. A habilidade de compreender o que está sendo falado em uma apresentação oral é uma tarefa que difere da compreensão de leitura de um texto escrito ou do reconhecimento informações linguísticas contidas no texto. Compreender o que se ouve numa língua estrangeira é um exercício que requer relacionar pistas e hipóteses, fazer inferências *do* e *sobre* o texto ouvido, no momento da interação.

No que se refere ao processamento de informações, a compreensão oral está sujeita a limitações da capacidade da memória humana. Na compreensão da linguagem, a memória desempenha duas funções: armazenamento de informação para posterior recuperação e processamento.

Quando as exigências das tarefas são elevadas, como em um teste de compreensão oral em que muitos detalhes se fazem presentes, muitas vezes devido à necessidade de armazenamento e processamento das informações, o processo de

lembrar as informações torna-se mais lento e alguns resultados parciais do processo da memória acabam por ser esquecidos.

Não é uma tarefa simples pedir para alguém lembrar-se de todos os detalhes presentes em uma mensagem oral: o que se retém é o que é considerado como importante para as necessidades na hora da escuta do texto oral. No caso do texto-base utilizado para as questões 28 a 30, os detalhes mencionados como: o tipo de quarto a ser reservado (*non-smoking room*), as opções de serviços desejados (*fitness center, room service, internet connection*), o tipo de carro (*compact car, nothing big, automatic transmission*), não são relevantes para responder aos itens. No entanto, fazem parte das informações perfeitamente previsíveis de um evento comunicativo em um contexto semelhante ao apresentado. De qualquer maneira, dado o excesso de informações que chamam a atenção do ouvinte, a compreensão pode ser prejudicada.

Provavelmente, responder ao item 29 também pode ser uma tarefa difícil, pois há nas alternativas novamente um detalhe relacionado a datas. Além desse aspecto, há também a necessidade de o examinando entender o que “*drop off*” quer dizer para responder ao item. A compreensão direta de um texto oral abrange também o reconhecimento de vocabulário, o que implica ter acesso às informações armazenadas na memória, o que, às vezes, não é uma tarefa das mais simples.

Se o excerto apresentado no teste fosse um diálogo utilizado na vida real, o interlocutor não ouviria todos os detalhes e certamente não se deteria em detalhes mínimos e supérfluos como datas. Esse tipo de texto acaba funcionando como uma forma de “retenção de informações textuais”, normalmente irrelevantes, e não como parte de um processo de aplicação de estratégias para auxiliar na compreensão.

5.3.4. Nível 4

A seguir apresentamos a ementa referente ao conteúdo programático referente a Inglês 4 e, em seguida, os itens de n^{os} 31 a 40 que compõem o teste de classificação referentes ao nível 4.

Quadro 13: Ementa da Disciplina Inglês 4

INGLÊS 4		
EMENTA	OBJETIVOS	ITENS
Consolidação da compreensão e produção oral e escrita com a utilização de funções sociais e estruturas básicas da língua desenvolvidas na disciplina Inglês 3. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa.	O aluno deverá ser capaz de:	
	- participar de discussões e negociações em contextos sociais e empresariais, destacando vantagens, desvantagens e necessidades.	35
	- preparar-se para participar de entrevistas de emprego presenciais e por telefone.	X*
	- compreender informações de manuais, relatórios e textos técnicos específicos da área.	34, 36, 37
	- redigir cartas e e-mails comerciais, relatórios e currículos.	X
	- aperfeiçoar a entonação e o uso de diferentes fonemas da língua.	X

* X representa objetivo não avaliado pelos itens.

** Os itens 31 e 32 não se encaixam em nenhum dos objetivos apresentados pela ementa.

A seguir os itens 31, 32, 35 e 36.

Item 31

Your boss is travelling on business and is on the phone now.
She needs some information that is in the report of the last meeting.

- What does she say to you?
- Could you send me the report?
 - Must you send me the report?
 - Would you like me to send the report?
 - Should you send me the report?

Item 32

- Choose the best alternative to complete the sentence.
I think that smoking cigarettes _____ illegal because it can hurt people around you.
- isn't allowed to be
 - should be
 - should to be
 - have to be

Item 35

- Complete the statement of a negotiation below with a suitable ending.
This deal is very important for the company _____
- then we shouldn't consider it.
 - so let's analyze it carefully.
 - then we shouldn't worry about it.
 - so let's postpone it.

Item 36

- Which sentence best completes the text?
I don't know if I like the idea of online shopping.
_____ **Can I return the item and get my money back?**
- Why don't we shop from home?
 - What if I like the product?
 - Why don't we buy the item on the internet?
 - What if the item I receive is damaged?

Podemos perceber que as questões que compõem o teste no nível 4, principalmente as de números 31, 32, 35 e 36 são elaboradas com a intenção de verificar se o aluno domina determinadas estruturas gramaticais que envolvem verbos modais como *should*, *could*, *would*, *to be allowed to*, *have to*.

Necessariamente o avaliando precisará dominar esse item gramatical para responder às questões, exigindo-se dele saber diferenciar o uso de cada um desses verbos nas situações apresentadas nos itens. Nesse sentido, avalia-se a habilidade de o aluno saber utilizar uma forma gramaticalmente correta em um dado contexto. Avalia-se, portanto, a competência linguística do examinando.

Item 33

Choose the best option to complete the sentence.

_____ they were extremelly tired, they went to school.

- a. On the other hand
- b. However
- c. Although
- d. But

O item 33 não está contemplado na ementa e trata do reconhecimento de vocabulário pelo uso de uma palavra de ligação. Não é um item de fácil resolução pelo aluno, pois normalmente as palavras de ligação não são muito vistas nos ensinios fundamental e médio como tal. Muitas delas são vistas como advérbios e, como não são palavras cognatas ao português, sua identificação pode ser muito difícil.

Item 34

Read the policy below on the use of calculators in university examinations and then choose the only sentence which is true:

- i) No programmes or information stored in any medium may be brought into the examination room.
 - ii) All calculators should be reset prior to entering the examination room.
 - iii) No calculator may be used which would disturb another candidate.
- a. Students shouldn't use calculators in university examinations.
 - b. Students can't use calculators in university examinations.
 - c. Students must use programmes stored in the calculators during their examinations.
 - d. Students can use calculators in the classroom if they follow certain rules.

Item 37

What's the idea of the text above?

- a. The safety of shopping online.
- b. Doubts about shopping online.
- c. The price for shopping online.
- d. Advantages of shopping online.

Os itens 34 e 37 avaliam a compreensão de textos. No item 34 um texto é apresentado e pede-se para que o examinando escolha a afirmação correta sobre o mesmo. O item 37 está relacionado ao item 36 e avalia a compreensão da ideia e do contexto central do texto. Os dois itens estão indiretamente contemplados na ementa do curso quando é mencionado que o aluno deverá ser capaz de compreender informações de manuais, relatórios e textos técnicos específicos da área.

Apresentaremos agora os itens 38, 39 e 40 que fazem parte do nível 4 do teste de classificação. Em seguida visualizamos os itens do teste.

As questões de 38 a 40 referem-se a uma gravação.

Item 38

What is the problem mentioned by Rick?

- a. He has financial problems.
- b. The computer is not working.
- c. Problems with the printer.
- d. Internet problems.

Item 39

Louise has offered some suggestions related to:

- a. Printer models
- b. Computer programs
- c. Computer store addresses
- d. Prices

Item 40

The dialogue is probably between:

- a. Salesman and buyer
- b. Mother and son
- c. Two friends
- d. Two salesmen

Audio script:

A: Hi Rick, what are you doing here?

B: Hei, Louise, I'm looking for a printer. I have a new computer and my old printer isn't working anymore. Do you have any suggestions?

A: I have an HD922. It's terrific.

B: Really, is it expensive?

A: Yes, it is, but It's worth the money.

B: I don't have much money. What else do you suggest?

A: Well, a friend of mine has the HD9102. It makes good prints. Maybe the clerk can help you.

B: Thanks, Louise

A: Good luck, see you in class.

O texto-base para os itens 38 a 40 inclui um tema que é bastante pertinente ao aluno de um curso de tecnologia, em especial: a escolha de um modelo de impressora. O vocabulário apresentado no diálogo provavelmente é conhecido por parte desse aluno, não trazendo, portanto, grandes dificuldades na compreensão. O que pode ser um empecilho para o entendimento da mensagem, nesse caso, é o adjetivo “*terrific*” - que para muitos examinandos pode ser traduzido como “terrível” em vez de “maravilhoso, excelente”. Nesse sentido, a escolha desse falso cognato atua no teste como um distrator, apesar da palavra não interferir na interpretação dos itens.

Os itens que compõem um teste de múltipla escolha são exemplos de uma tarefa “fechada”, com uma dificuldade fixa, em que a informação é pré-analisada e as opções linguísticas para expressar os pontos identificados são estabelecidos previamente (Rost, 1990).

Percebemos, portanto, que o propósito na elaboração dos itens do nível 4 é o de identificação de um tópico e sua ideia principal, sem se deter a detalhes, incluindo fatos que ajudem na identificação dessa ideia central. Dessa forma, os itens são coerentes com essa perspectiva, atendendo inclusive ao que está previsto na ementa “a consolidação da compreensão e produção oral com a utilização de funções sociais e estruturas básicas da língua e a ênfase na oralidade”

5.3.5. Nível 5

A seguir, a ementa referente a Inglês 5, seguida dos itens n°s 41 a 50, que fazem parte do nível 5 do teste de classificação.

Quadro 14: Ementa da Disciplina Inglês 5

INGLÊS 5		
EMENTA	OBJETIVOS	ITENS
Aprofundamento da compreensão e produção oral e escrita com a utilização de funções sociais e estruturas mais complexas da língua. Ênfase na escrita e na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa.	O aluno deverá ser capaz de:	
	- fazer uso das habilidades linguístico-comunicativas com maior espontaneidade e confiança.	X*
	- fazer uso de estratégias argumentativas.	X
	- acompanhar reuniões e apresentações orais simples e tomar nota de informações.	X
	-redigir correspondência comercial em geral.	47
	- compreender informações em artigos acadêmicos e textos técnicos específicos da área.	45
	- aperfeiçoar a entonação e o uso de diferentes fonemas da língua, de forma a garantir a inteligibilidade nos contatos em ambiente profissional, tanto pessoalmente quanto ao	X

	telefone.	
--	-----------	--

*X representa objetivo não avaliado pelos itens.

Nesta etapa do teste podemos identificar que vários aspectos propostos na ementa não são contemplados. Alguns aspectos da ementa como: fazer uso das habilidades linguístico-comunicativas com maior espontaneidade e confiança e aperfeiçoar a entonação e o uso de diferentes fonemas da língua, de forma a garantir a inteligibilidade nos contatos em ambiente profissional, tanto pessoalmente quanto ao telefone, não é possível ser verificado em um texto escrito, uma vez que são características relacionadas ao desempenho oral do examinando, podendo, portanto, ser verificadas durante a entrevista.

A seguir apresentaremos os itens 41, 42, 43 e 46.

Item 41

Choose the best alternative to complete the sentence:

**In response to growing demand for energy in recent years, Brazil
_____ and exporting ethanol.**

www.gov.uz

- a. stopped producing
- b. is decreasing production
- c. is planning to produce
- d. started producing

Item 42

You are discussing a new business opportunity and your boss says:

“This negotiation is too important to lose,
_____”

Choose the best alternative to complete the sentence.

- a. because we shouldn't make a mistake.
- b. as long as we don't make a mistake.
- c. however, we shouldn't make a mistake.
- d. so we shouldn't make a mistake.

Item 43

Choose the best alternative to complete the sentence.

Trade between Africa and Brazil _____ by 415% since 2002 when President Lula _____ the Presidency of Brazil.

<http://europafrica.net/2009/06/19>

- a. has increased – assumed
- b. increased – has assumed
- c. is increasing – assumes
- d. will increase – has assumed

Item 46

Choose the best alternative to complete the sentence.

It was almost eleven p.m. and I _____, but I couldn't go home because I had to finish the report.

- a. have been tired
- b. am tired
- c. am going to be tired
- d. was getting tired

Em relação aos itens 41, 42, 43 e 46 observamos que são itens que prioritariamente avaliam um conhecimento gramatical, mas não apresentam uma correlação direta com o que está previsto pela ementa. Por esse instrumento revela-se um conceito de proficiência cuja característica principal é dominar regras gramaticais, sem se levar em conta o uso da linguagem.

No item 41, para completar a lacuna, o examinando deve diferenciar o uso do tempo verbal *simple past* e o do *present continuous*. No item 42, também para completar a lacuna, é necessário, além do conhecimento do verbo modal *should*, o conhecimento do uso do verbo no *simple present*. Entretanto, talvez o que traz mais dificuldade para o aluno são as palavras de ligação *as long as* e *however* (esta última um pouco menos do que a primeira) uma vez que *because* e *so* costumam ser trabalhadas nas escolas de ensinos fundamental e médio. O mesmo se aplica ao item 45, que utiliza a palavra de ligação *although*. Além disso, nos itens 43 e 46, há a cobrança do uso do tempo verbal *present perfect* que normalmente traz dificuldades para alunos brasileiros.

A seguir os itens 44, 45 e 47.

Item 44

_____ the presidents of Brazil and the US came from different sides of the political spectrum, they had a good relationship.

Choose the best alternative to complete the sentence:

- a. As long as
- b. Although
- c. As
- d. Because

Item 45

Choose the best alternative to complete the text.

Nano Technology has tremendous potential and a lot of young people take a keen interest in this field. In India, _____ the economic slowdown, science & technology has done well and has shown steady growth.

adapted from www.nanotech-now.com/news

- a. however
- b. in spite of
- c. according to
- d. as a result of

Item 47

You work in the human resources department of a large corporation.

You are writing a report explaining the decision for choosing one particular candidate. Which sentence **would not** be appropriate to start the document:

- a. Because the position required computer skills we decided to choose Mr. Silva.
- b. How the position required computer skills we decided to choose Mr. Silva .
- c. As the position required computer skills, we decided to choose Mr. Silva.
- d. As long as the position required computer skills we decided to choose Mr. Silva .

Os itens 44, 45 e 47 são itens interessantes, pois são questões que estabelecem relações, pelo uso de elementos de ligação para garantir a coesão da oração, e que requerem que o examinando aprenda a relação existente entre os elementos do texto.

O item 47 é o único item nos quatro níveis do teste até agora que utiliza as instruções para como redigir uma correspondência comercial. Um ponto que podemos argumentar como não adequado para esse item é pedir para que o examinando aponte a forma errada e não a forma correta. O ideal seria indicar qual seria a alternativa correta e as demais incorretas, assim como oferecer elementos que permitam compreender o acerto na resolução do item, de acordo com observações apontadas sobre a elaboração de testes, no capítulo 3.

O nível 5 é composto de 3 itens que avaliam a recepção oral dos examinandos. O texto-base para a composição dos itens de avaliação versa sobre uma pesquisa de mercado e quais produtos de alimentação podem ser agregados à empresa para estender seu leque de opções de venda.

A seguir, os itens 48, 49 e 50 que compõem a compreensão oral do nível 5.

Listen to the extract of a meeting in a company that produces frozen food and answer questions 48 - 50:

Item 48

The aim of the meeting is to...

- a. discuss the results of a recent market research.
- b. talk about healthy food and its trends.
- c. decide how to improve the line of products.
- d. discuss what kind of new products the company should develop.

Item 49

According to Donna's market research:

- a. Consumers don't have much interest in healthy desserts.
- b. Consumers have a great interest in buying healthy food.
- c. Less than 50% of consumers buy healthy food some or all of the time.
- d. Consumers buy more vegetarian food than healthy food with meat or fish.

Item 50

What are the consumer's options in order of preference, according to Donna's research?

- a. fat free meat and fish, sugar free desserts, fruit.
- b. low fat meat and fish, fruit and low sugar desserts, vegetarian dishes.
- c. healthy food with little salt, diet desserts and fruit, vegetarian dishes.
- d. healthy food without salt, fruit, vegetables.

Audio script:

A: OK, let's start. At the last meeting, we agreed that we need to extend our product range. The purpose of this meeting is to decide what kind of new products we should develop. Donna, you've looked at the market research. What have you found out?

B: Well, the trend is definitely towards healthy foods. Market research shows that 60 per cent of people buy healthy foods some or all of the time. And 35 per cent often buy vegetarian food. If we go with the trend, we've got three options. The first is healthy dishes made with meat or fish and very little salt or fat. The second is healthy desserts, made with fruit and not much sugar. And the third is vegetarian dishes.

Para avaliar se o examinando tem capacidade de processar e compreender a linguagem falada é preciso fornecer-lhe textos que sejam representativos da linguagem falada em contextos reais. Trazer textos que sejam representativos da linguagem falada no dia a dia traz autenticidade ao teste, e pode contribuir para a construção da validade desse teste.

Ao reconhecer uma determinada situação de fala como autêntica e possível, o examinando pode levar em consideração seu conhecimento prévio, auxiliando, portanto, na compreensão. Ao se trazer um texto contendo um tópico familiar, acrescido de alguma informação um pouco além da já conhecida pelo examinando, faz com que este consiga usar o conhecimento prévio para inferir determinadas informações desconhecidas, não somente no nível linguístico, mas também no da conceituação e entendimento de fatos. Ao apresentar-se um tópico completamente desconhecido desse examinando, o reconhecimento linguístico acaba se tornando inacessível, trazendo confusão e pavor. Para Rost (1990), compreender a língua falada é um processo essencialmente inferencial, baseado em pistas e ativação de conhecimento prévio, em vez de meras associações entre o que é ouvido e os aspectos do contexto.

O assunto do texto não é dos mais conhecidos dos examinandos, por tratar de produtos de alimentação, diferenciado em categorias. Algumas palavras oferecem uma

dificuldade maior de compreensão, tais como: “*healthy*”, “*dessert*”, “*dishes*”, “*research*”, algumas delas estranhas ao aluno. Além disso, algumas opções de tipos de comida, não tão fáceis de serem identificadas, são apresentadas. O conhecimento de expressões como *low fat*, *sugar free*, por exemplo, nem sempre auxiliam na compreensão do texto, o que é uma pena. Para chegar a compreender a alternativa é preciso “transformar os substantivos *salt* ou *fat* e o quantitativo *little* na locução *low fat meat and fish*”. O mesmo acontece com a frase: *The second is healthy desserts, made with fruit and not much sugar*, que teria de ser entendida e transformada em: *fruit and low sugar desserts*.

No entanto, apesar desses itens do teste avaliarem o conhecimento do léxico e, a entonação e o uso de diferentes fonemas da língua, em palavras como *healthy*, *desserts*, *vegetarian* e *vegetables*, por exemplo, (cujas compreensão oral garante a inteligibilidade do evento comunicativo), eles não contemplam o que está estabelecido na ementa como tópicos de aprendizagem. O que continua a ser avaliado é o conhecimento da estrutura e do sistema da língua.

Considera-se, à medida que se chega aos níveis mais altos, pertinente avaliar dificuldades especialmente relacionadas ao léxico e às estruturas da língua e à configuração dos itens, com um número maior de detalhes e informações, muitas vezes supérfluas.

5.3.6. Análise do Nível 6

O 6º e último nível do teste é representado pelos testes de nºs 51 a 60 e a ementa, cujo conteúdo deve seguir como subsídio para o teste segue abaixo.

Quadro 15: Ementa da Disciplina Inglês 6

INGLÊS 6		
EMENTA	OBJETIVOS	ITENS
Aprimoramento da	O aluno deverá ser capaz de:	

compreensão e produção oral e escrita com a utilização de funções sociais e estruturas mais complexas da língua desenvolvidas na disciplina Inglês 5. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa.	- fazer uso das habilidades linguístico-comunicativas com mais autonomia, eficiência e postura crítico-reflexiva.	X*
	- Interagir em contextos de socialização e entretenimento.	X
	- Redigir textos técnicos e acadêmicos.	X
	- Compreender informações em artigos acadêmicos e textos técnicos específicos da área.	54, 57
	- Aperfeiçoar a entonação e o uso de diferentes fonemas da língua, de forma a garantir a a inteligibilidade nos contatos em ambiente profissional, tanto pessoalmente quanto ao telefone.	X

*X representa objetivo não avaliado pelos itens.

A seguir os itens 54 e 57.

Item 54

According to the text below:

About a third of the fuel Brazilians use in their vehicles is ethanol, known in Brazil as 'alcohol.'

That compares with three percent in the United States. All gasoline sold in Brazil contains at

least 26% ethanol, but motorists driving flexible-fuel cars have the option of filling up with pure ethanol, or E100, which currently is selling for about half the price of the blend.

The trend in Brazil has far-reaching implications for environmental policy, trade and economic development in poor countries that may have a bright future producing crops that can be easily turned into fuels.

<http://www.researchandmarkets.com/report>

- a. Poor countries have environmental problems with ethanol.
- b. In Brazil it is possible to choose from pure gasoline and pure ethanol.
- c. Brazil produces 26% more ethanol than gasoline.
- d. Many more cars use ethanol in Brazil than in the USA.

Item 57

According to the text below, if we work from home _____

Advantages and disadvantages of working from home

Working from home can have many advantages. You may find it helps you to:

- arrange work around family commitments;
- get help from family, e.g. they may help you with filing or general administration tasks;
- save on time and money spent traveling to work;
- avoid noise and distractions of the workplace;

However, there are disadvantages from home working too. You may find you have trouble with:

- separating home and work life;
- the initial expense of setting up;
- domestic distractions and interruptions;
- isolation/loneliness.

<http://www.businesslink.gov.uk>

- a. we can make it easier to maintain team work.
- b. we can avoid distractions and interruptions.
- c. it's impossible to keep home and work life apart.
- d. we can spend less money and time travelling to work.

Os únicos itens do nível 6 que contemplam o conteúdo da ementa são os itens 54 e 57, uma vez que apresentam textos para interpretação. Sabemos que em um tipo de teste como o teste de classificação, um teste de múltipla escolha acaba por ser a forma mais rápida para avaliar os itens, especialmente a competência de leitura do examinando. Entretanto, devemos observar que, ao escolhermos tal tipo de teste, o examinando não constrói o sentido do texto, ou seja, a leitura em si não é processada pelo leitor e sim,

pelo autor do item, uma vez que o sentido já está determinado nas alternativas, só cabendo ao examinando identificar o sentido proposto. Nesse sentido, a multiplicidade de sentidos no processamento de leitura não se faz presente no construto do teste, conforme pode ser observado nos itens apresentados a seguir.

A seguir os itens 51 e 53.

Item 51

Which statement *is false* according to the table?

Table 1. Brazilian Ethanol Production

Year*	Million Gallons
2003/04	3,910
2004/05	4,068
2005/06	4,174
2006/07	4,719
2007/08	5,916
2008/09**	7,054

* marketing year = May – April

** forecast

Source: GAIN report BR8013, USDA Foreign Agricultural Service, 2008.

- a. Ethanol production is expected to grow.
- b. Ethanol production has increased significantly in the last six years.
- c. Ethanol production has expanded in recent years.
- d. Ethanol production dropped in 2007.

Item 53

Choose the best alternative:

- a. Brazilian exports fell sharply between 2003 and 2006.
- b. Brazilian exports doubled between 2002 and 2004.
- c. Brazilian exports have increased continuously since 2002.
- d. Brazilian exports were steady in the period of 2002 and 2006.

Em relação aos itens 51 e 53, indicados adiante, verificamos que o item 51 também se trata de uma questão gramatical, na qual o examinando, a partir da leitura e análise de um gráfico, precisa chegar a uma resposta. Entretanto, como as alternativas

giram em torno de tempos verbais, o examinando deve escolher entre o uso da voz passiva (alternativa a: is expected); *present perfect* (alternativa b: has increased e alternativa c: has expanded) e *simple past* (alternativa d: passed).

O item 53 é um item com foco gramatical combinado com o conhecimento lexical, pois utiliza palavras antagônicas como: a) *fell sharply*; b) *doubled*; c) *has increased*; d) *were steady*.

Desse modo, o examinando tem de escolher a melhor alternativa, tendo conhecimento dessas expressões. Além disso, há no item 53 um problema de clareza no enunciado, uma vez não é informado ao aluno de que o item 53 também se refere à tabela apresentada no item 51. Espera-se, portanto, que o aluno consiga inferir essa informação. Por mais que possa ser possível tal inferência pelos alunos, essa falta de clareza no enunciado prejudica o teste em tempo gasto para a resposta, indecisão e confusão no aluno, entre outros problemas que porventura venham a ocorrer. Como essa informação não faz parte do enunciado da questão, o examinando pode escolher qualquer possibilidade de resposta e anular o item, uma vez que não há uma alternativa correta (ou todas elas podem ser corretas) se não for indicado, de algum modo no enunciado, que o teste 53 está atrelado à tabela do teste 51.

Os itens 52, 55 e 56 também são itens que parecem avaliar, prioritariamente, o conhecimento gramatical dos examinandos.

Item 52

Choose the best sentence to complete the gap.

Effective meetings that produce results begin with meeting planning. First, identify if other employees are needed to help you plan the meeting, then decide _____ .

Source: humanresources.about.com

- a. if you are going to organize it or not
- b. what the actual results will be
- c. if you can make all the decisions by yourself
- d. what you hope to accomplish by holding the meeting

Item 55

Choose the best alternative for a suitable answer:
 You are the manager of a supermarket. You are angry with your supplier because they didn't deliver the products you bought on time. The supplier tries to calm you down.
 You - We would like to cancel the order.
 Supplier- What if we offer you a special discount?
 You - If you offered me a 50% discount _____.

- I want to reconsider my decision.
- I reconsider my decision.
- I would reconsider my decision.
- I will reconsider my decision.

Item 56

Choose the best alternative to report Obama's opinion about Brazil:
Obama: "Brazil is an economic power and a key player in the international scenario".

<http://www.rio2016.org.br/en/Noticias>

- Obama asked if Brazil was an economic power.
- Obama spoke that Brazil was an economic power.
- Obama said that Brazil was an economic power .
- Obama told Brazil is an economic power.

Como já observamos anteriormente, tópicos gramaticais não são mencionados na ementa e estão embasados na concepção behaviorista da linguagem, pela qual a estrutura da língua pode ser "dividida" em pequenos elementos ou partes, que mesmo divididas, não alteram o significado. No item 52 temos um pequeno trecho incompleto que deve ser completado por uma oração, cujo objetivo é o de diferenciar o uso da conjunção "if" e "what". O vocabulário desse item não é fácil, mas pode ser inferido em algumas das alternativas. No entanto, como no enunciado há a escolha de if, sobra para a alternativa correta a indicação do uso de What. Nesse sentido, há um problema com os distratores, permitindo que a alternativa correta possa ser inferida. O item 55 reproduz uma situação comunicativa. Entretanto, as alternativas giram em torno de tempos verbais, como: *simple present: want, reconsider; conditional: would reconsider; e future: will reconsider*. Novamente, o que é priorizado é o conhecimento do sistema e a correção na língua. O item 56 refere-se ao que chamamos de discurso indireto ou *reported speech*, utilizando os verbos: *asked, spoke, said e told*.

A seguir os itens 58, 59 e 60.

Two colleagues are talking about moving to a new office building. Listen to the dialogue and answer questions 58 - 60:

Item 58

Choose the best alternative:

- a. The second speaker does not want to move to the new building.
- b. The second speaker is not convinced of the benefits of the new office.
- c. The first speaker thinks the old office building was perfect.
- d. Both the first and second speaker believe the new office will be more comfortable.

Item 59

Which advantage of the *new building* is mentioned in the conversation?

- a. It has a lot of space for all the staff.
- b. It has a good bus service.
- c. It is near restaurants and shops.
- d. It is near downtown.

Item 60

What does the *second speaker* suggest at the end of the conversation?

- a. The employees should hire a company bus service.
- b. They should talk to the manager about their problems.
- c. They should go to work by car.
- d. The manager should look for a new location.

Audio script:

A: What do you think about moving to the new offices, then?

B: Personally, I'm not keen on the idea. I like being here in the centre of town. I've been able to go out to the shops on the lunch hour. I really miss going into our favorite café for lunch. You know, we'll have to eat in the office canteen. Because there will be nowhere else to go.

A: Well yes, it won't be as nice being outside town. But think how much more practical the new building will be. We'll have more space, more light, more modern conveniences. We know how small that old building is. And the company is expanding. In fact you are always complaining about the size of your office.

B: Yes, that's true I suppose. But I will have a problem getting to work.

I'll have to go by car, because the bus service out there is terrible.
A: Yes, I agree, and not everyone has a car. Maybe we should talk to the management about that. They should organize a company bus service for all the staff.

Os itens 58, 59 e 60, a seguir, avaliam a recepção aural dos examinandos. Entretanto, os itens descritos na ementa como: “fazer uso das habilidades linguístico-comunicativas com mais autonomia, eficiência e postura crítico-reflexiva e aperfeiçoar a entonação e o uso de diferentes fonemas da língua, de forma a garantir a inteligibilidade nos contatos em ambiente profissional, tanto pessoalmente quanto ao telefone” não parecem diretamente contemplados nos itens.

Contudo, o que é pedido nos itens do teste pode abarcar o item previsto na ementa que objetiva que: “o avaliando seja capaz de interagir em contextos de socialização e entretenimento”, uma vez que, apesar do contexto profissional escolhido para compor os itens, o que é avaliado nos itens é a capacidade de interagir em contextos de socialização. Nesse caso, podemos perfeitamente admitir que em contextos de trabalho, situações variadas de socialização e de interação entre as pessoas, podem ocorrer com frequência.

Mesmo assim, como podemos observar no item 60, a ênfase apenas na competência linguística (no sistema, na estrutura e nos componentes linguísticos), em casos que exploram o uso de verbos como *should*, continua a fazer parte do repertório das alternativas e, por conseguinte, da avaliação.

Observamos também que o texto-base para os itens 58 a 60 utiliza uma língua falada que sofre assimilações e outras modificações da forma padrão e que implicam no não reconhecimento de unidades linguísticas. A linguagem natural traz outros recursos que facilitam a compreensão, tais como as expressões “*well*”, “*Yes*”, “*Personally*”, que acabam diluindo a concentração de informações, de forma a dar tempo para o examinando processar a língua ouvida.

Mesmo assim, esse texto apresenta vários detalhes que podem deixar o examinando bastante confuso na hora de respondê-los. Se o texto-base for ouvido somente uma vez, algumas informações poderão ser deixadas de lado, uma vez que o

ouvinte, em uma interação real, não observa todos os detalhes presentes em uma conversa, anotando mentalmente apenas os pontos mais relevantes.

Ao compararmos a ementa do curso e as 10 questões do nível 6 observamos que basicamente temos questões gramaticais, dando ênfase ao conhecimento da estrutura da língua. A ementa enfatiza a compreensão e produção oral e escrita com a utilização de funções sociais, ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos socioculturais da língua inglesa, entretanto não apresenta itens que contemplem essas atividades.

Segundo Brown (1994:235), “A recepção oral não se resume a uma mera transmissão e recepção de sons, mas num processo complexo de ordem psicomotora que vai desde a percepção pelo ouvido de um estímulo sonoro, até a recepção e processamento da mensagem no cérebro, tudo em questão de micro segundos.” Compreender linguagem é um processo de re(construir) sentidos, sejam ele a partir de um texto falado ou escrito.

Comparando-se a recepção de língua materna e língua estrangeira, o processo é bem mais difícil com a última, uma vez que o ouvinte precisa prestar atenção para compreender o significado do que está sendo falado, isto é, a mensagem, que é emitida por uma determinada pessoa, que pode possuir sotaque, tonalidade de voz, articulações diferenciadas, além de se deter, muitas vezes, a um vocabulário desconhecido, que para ser entendido, o ouvinte precisa se valer do contexto, para sua compreensão.

O discurso dentro de um mundo real contém muitas ambiguidades, pausas e falsos começos. Se alguns desses problemas afeta a compreensão do ouvinte, este pode esclarecer por meio de perguntas. Numa situação de exame, esse processo é impossível, de modo que os textos e diálogos que são utilizados para testes de audição precisam ser inequívocos. Ao se ter contato com um material gravado, e este não apresentar nenhum apoio para compreensão, como aspectos visuais, movimentos faciais e labiais, gestos extralinguísticos, acaba por prejudicar a compreensão do ouvinte.

Além disso temos também a limitação de tempo e a necessidade de mediar referências culturais específicas, e até mesmo um conhecimento prévio do assunto

tratado. Por todas essas razões, um texto ou diálogo utilizado no mundo real vai precisar de um grau de adequação para utilização em uma avaliação de recepção oral.

A recepção/compreensão oral tem sido reconhecida como uma parte essencial da comunicação, e uma forma importante para a aquisição de linguagem. Devido à globalização e ao comércio entre países (cujo idioma de contato é preferencialmente o inglês), bem como à educação (sem fronteiras), ao uso da internet para a comunicação (via Skype) e a algumas formas de entretenimento (como a TV a cabo), a necessidade de compreender o inglês oral está aumentando. Em particular, no domínio da língua para efeitos acadêmicos, a habilidade de compreensão de textos orais é um componente crítico na competência comunicativa dentro de um cenário universitário, necessitando, portanto, de melhor desenvolvimento e atenção.

Testes de compreensão auditiva para fins acadêmicos procuram avaliar o entendimento de mini conferências ou diálogos para que, em seguida, o examinando responda algumas tarefas cuja intenção é verificar a quanto o examinando consegue compreender desses textos. Assim, é inevitável entender primeiramente qual é o construto que embasa essa avaliação de recepção/compreensão oral, para em seguida entendermos qual a opção de escolha dos textos e dos itens que compõem a avaliação.

Como seria esperado em um teste de recepção/compreensão oral, de um modo geral, o processo de compreensão oral propriamente dito é difícil de ser percebido e avaliado, o que não ocorre, por exemplo, durante uma entrevista. Podemos, no primeiro caso, apenas fazer inferências do resultado do processamento na forma das respostas dos examinandos. Devido a esse fato, é particularmente importante que os itens que compõem o teste estejam de acordo com o construto da prova.

O teste que é alvo de nossa pesquisa apresenta como construto, segundo as ementas que compõem os 6 (seis) níveis de aprendizagem de língua Inglesa, “habilidades de compreensão e produção oral e escrita por meio de funções sociais e estruturas simples da língua. Ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos sócio-culturais da língua inglesa”.

Podemos observar, então, que o objetivo maior desse teste de classificação reside na ênfase na oralidade. No entanto, em função da quantidade de itens que avaliam essa habilidade ser muito pequena, uma boa avaliação da compreensão auditiva do

examinando não é feita e, por consequência, a avaliação não está de acordo com os objetivos previstos no curso. Os textos que compõem a prova de recepção/compreensão oral são em sua maioria representativos de uma linguagem real utilizada no dia a dia em um ambiente de trabalho, com exceção do texto-base do nível 5, que por apresentar um diálogo falando sobre alimentação, foge parcialmente dos assuntos que se espera nos testes.

A avaliação da compreensão da língua oral pode ser estabelecida em três níveis de processamento para entendimento de textos. No *nível 1* ocorre a compreensão direta, que corresponde ao reconhecimento ou adivinhação do vocabulário, a lembrança de fatos e detalhes do texto oral e a identificação das ideias principais nele contidas. No *nível 2* ocorre a interpretação com inferência, que procura identificar o propósito e atitudes daquele que se comunica através do texto, assim como as implicações do que está dito, as interligações das ideias e alguns sentidos específicos que são apreensíveis através da fala, e a avaliação e extensão, por meio da qual o ouvinte utiliza recursos lógico-dedutivos, procurando contradições, estabelecendo juízos de valor, através de verdades e/ou falácias nos argumentos. Finalmente, no *nível 3* ocorre a avaliação do resultado na construção de um sentido interno ao texto e dos possíveis efeitos externos ao mesmo (Almeida Filho & El Dash, 2002).

No teste de classificação analisado observa-se que em todos os níveis o examinando é praticamente condicionado a utilizar para a compreensão oral a “apreensão de informações”, a partir da identificação de um tópico e a ideia principal, que não é muito geral ou restrita.

Alguns aspectos devem ser salientados em relação ao teste de classificação mediado pelo computador: a rapidez na avaliação e a possibilidade de utilizar fones de ouvido, o que melhora muito a recepção do som, sem a interferência de ruídos que possam prejudicar o entendimento do texto. Nesse sentido, o teste apresenta um aspecto bastante positivo, que lhe garante, no mínimo, uma maior validade de conteúdo. Outro tipo de validade que pode ser facilitada pelo uso do fone de ouvido é a validade de face, por permitir uma audição do que é dito com clareza e perfeição, próximas das situações concretas de interação.

5.4. Análise da Produção Oral e a Avaliação nas Entrevistas

Nesta seção apresentaremos aspectos relacionados à entrevista que é feita a partir do nível 3 aos alunos que têm um mínimo de 60% de resultado satisfatório no teste, conforme já salientado anteriormente. No caso específico deste trabalho o foco não foi na entrevista pela qual os examinandos são submetidos ao chegarem ou ultrapassarem o nível 3 do teste. Entretanto, alguns comentários e análise se fazem necessários, mesmo que sejam meras leituras impressionistas sobre esse processo. Nossa intenção aqui é a de auxiliar na melhoria do processo ao apontarmos os pontos merecedores de uma atenção maior por parte dos profissionais que atuam na instituição

A avaliação da produção oral é feita por meio de uma entrevista presencial, com duração de 5 a 15 minutos, envolvendo o examinador, que no caso é o professor da disciplina e o examinando. A entrevista não tem um formato fechado, e as questões e interações são baseadas nas ementas e objetivos das disciplinas de inglês. Como já foi por vezes observado, somente fazem a avaliação oral os alunos aprovados para o nível três, ou acima, no exame escrito. O objetivo do exame oral é confirmar o nível obtido pelo aluno no exame escrito.

Os elaboradores do teste sugerem que a interação se inicie com o cumprimento e apresentação do examinador, para que se tenha uma atmosfera mais receptiva e agradável. O examinador deve fazer as perguntas e oferecer comentários como forma de dar continuidade à conversa. O examinador tem também a liberdade de criar suas próprias perguntas ou situações para cada nível, de acordo com o que está estipulado na ementa, ou seguir a sugestão de um roteiro previamente elaborado e entregue ao professor antes de dar início à entrevista.

Abaixo, descrevemos os objetivos expostos na ementa para cada nível e alguns exemplos de perguntas que constam como sugestões para os examinadores.

Quadro 16: Sugestões de perguntas – Nível 3

Nível 3: *Objetivos descritos: descrever experiências profissionais e habilidades, expressar opiniões e necessidades e fazer solicitações.*

Sugestão de perguntas/interação:

Do you work? What do you do in your job? Talk about your responsibilities.

Have you worked before? Talk about your previous job.

In your opinion what is more important, money or quality of life?

In your opinion, what is the most prestigious job at the moment?

(The examiner could talk about the place he/she lives and for how long to make the students more comfortable with the interaction).

Quadro 17: Sugestões de perguntas – Nível 4

Nível 4: Objetivos descritos: participar de entrevistas de emprego presenciais e por telefone; destacar vantagens, desvantagens e necessidades; usar linguagem apropriada de polidez e formalidade.

Sugestão de perguntas/interação:

(Ask questions using modal verbs to express opinion/suggestion and/or obligation. Let the students express their opinion in a free and relaxed way).

Make comments on the following excerpt:

“EVERYONE SEES LIFE THROUGH THEIR JOB. TO THE DOCTOR, THE WORLD IS A HOSPITAL. TO TRUCK DRIVERS, IT’S A ROAD SYSTEM.”

Quadro 18: Sugestões de perguntas – Nível 5

Nível 5: Objetivos descritos: aperfeiçoar as estratégias argumentativas; acompanhar reuniões e apresentações orais simples.

Sugestão de perguntas/interação:

Do you work? How long and where? Do you use English in your everyday activities?

What is your opinion about a placement test? Do you think it can really help teachers and students? Justify your answer.

ST: Suppose you regularly have meetings to launch new products. You will call a meeting tomorrow to talk with your team about the launch of a new chocolate bar. You

have already decided on the name of the product, price and places to be advertised and want your team to accept your ideas. How would you convince them?

Quadro 19: Sugestões de perguntas – Nível 6

Nível 6: Objetivos descritos: discutir planejamento; lidar com conflitos em negociações; lidar com valores e dados numéricos.

Sugestão de perguntas/interação:

What is your opinion about a placement test? Do you think it can really help teachers and students? Justify your answer.

ST: You have to negotiate the price, quantity and payment terms of a certain product. You could start by asking the unit price and then negotiate a discount for larger orders. You could also negotiate the payment terms in order to get more discount. Try to get the best possible deal. Interact with the examiner.

Em exames como os de classificação e proficiência, que a partir do desempenho demonstrado classificam os examinandos em faixas de proficiência, os critérios de avaliação e classificação devem estar muito claros, e a grade de correção e posicionamento do examinando deve ser feita de forma criteriosa. É preciso prever e abarcar as possíveis diferenças demonstradas pelos examinandos e saber classificá-las de tal forma que o resultado seja o mais adequado possível e pertinente ao desempenho esperado.

Entretanto, principalmente por falta de tempo, a competência oral acaba sendo negligenciada em muitos testes de classificação, reduzindo-se por vezes a uma avaliação holística, concretizada em uma apreciação global e vaga da participação oral do examinando.

Um fato muito importante a se considerar é que a grade de posicionamento dos alunos é muito superficial, não apresentando descritores que descrevam o que os examinandos são capazes de fazer, mas também como devem fazê-lo, como podemos observar nas orientações a seguir.

Quadro 20: Grade de Desempenho – Teste oral

O aluno é capaz de atingir os objetivos abaixo, conforme as ementas, com que qualidade?

Boa – o aluno continua o exame, podendo subir até um nível. Por exemplo, o aluno alcançou o Nível 3 no exame escrito, mas pode se matricular no Nível 4, devido ao resultado do exame oral.

Satisfatória – nível confirmado do exame escrito.

Insatisfatória – o aluno é matriculado num nível abaixo do resultado do exame escrito.

Esses descritores poderiam ser apresentados como: *uma escala*, com uma combinação de descritores de categorias diferentes; *uma lista de verificação*, agrupando itens de acordo com sua relevância; *uma grade holística* que aborde o desempenho global do examinando; *uma grade analítica*, que descreva em detalhes as habilidades que o examinando é capaz de realizar e que habilite o examinando a usar a língua estrangeira em vários universos.

A proficiência em língua pode ser avaliada de forma global, subjetiva e impressionista ou de forma analítica, por meio de instrumentos de análise do objeto avaliado. O que temos na escala exemplificada utilizada no teste é, na verdade, uma apreciação geral e subjetiva do examinador. Nesse sentido, trata-se a análise e avaliação da proficiência como algo possível e dependente da apreciação do examinador. Temos assim uma avaliação pouco aconselhável, uma vez que depende das impressões do avaliador que não se pauta em critérios para garantir uma homogeneidade ao sistema de avaliação.

Como sugestão, uma possibilidade bastante simples, até que uma análise mais aprofundada seja feita dessa etapa tão importante para a classificação desses alunos, é seguir a grade de avaliação oral do Quadro Comum Europeu de Referência para Línguas, que estabelece um esquema descritivo e os níveis comuns de referência. Esta poderia ser uma linha de ação a ser explorada pela instituição para contemplar momentaneamente a exigência de se posicionar os alunos em determinados níveis, a partir de critérios já testados e observáveis.

Assim, segundo o Conselho da Europa, organização que elaborou o Quadro Comum Europeu de Referência para Línguas, os descritores para os níveis de proficiência devem ser relevantes e apropriados para um contexto em que serão utilizados e as categorias utilizadas para descrever o que os examinandos são capazes de fazer devem relacionar-se com os contextos da língua-alvo, baseando-se em teorias da competência linguística. Além disso, esses descritores devem ser determinados objetivamente, na medida em que se baseiem em uma teoria de medição.

O Conselho da Europa estabeleceu seis níveis gerais que abrangem integralmente o espaço da aprendizagem pertinente para os alunos europeus de línguas (figura 7). Como já dissemos anteriormente, o quadro seria uma referência para posteriormente a instituição realizar a elaboração de um conjunto de pontos de referência para descrever com mais propriedade aquilo que se deseja dos alunos de cursos superiores de tecnologia. Desse modo, uma descrição global e mesmo simplificada do Quadro Comum Europeu de Referência das línguas pode oferecer algumas orientações importantes aos professores que fazem o papel de entrevistador nesse teste. A figura e os quadros a seguir ilustram essa orientação para o teste.

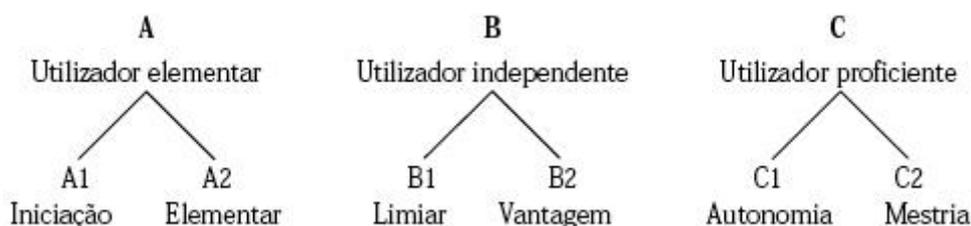


Figura 7: Níveis Comuns de Referência – Conselho da Europa

Fonte: QUADRO EUROPEU COMUM DE REFERÊNCIA PARA AS LÍNGUAS, 2001:48

	Produção Oral Geral
C2	É capaz de produzir um discurso claro, bem estruturado, fluente e com uma estrutura lógica eficaz que ajuda o receptor a aperceber-se e a lembrar-se das questões mais significativas.
C1	É capaz de fazer descrições claras e pormenorizadas e de expor assuntos complexos, que integrem subtemas, desenvolvendo questões específicas e terminando com uma conclusão adequada.
B2	É capaz de fazer, metodicamente, uma descrição ou uma exposição,

	<p>sublinhando as questões mais importantes e justificando as suas ideias com pormenores pertinentes.</p> <p>É capaz de fazer uma descrição ou uma exposição pormenorizada acerca de uma vasta gama de assuntos relacionados com os seus interesses, desenvolvendo ou defendendo ideias, fazendo notar questões secundárias e dando exemplos relevantes.</p>
B1	É capaz de manter razoavelmente bem e com fluência uma descrição direta de um dos muitos assuntos do seu interesse, apresentando-a como uma sucessão linear de questões.
A2	É capaz de fazer uma descrição simples ou uma apresentação de uma pessoa, das condições de vida ou de trabalho, das atividades quotidianas, daquilo de que gosta ou não, etc., numa série curta de expressões e de frases ligadas como numa lista.
A1	É capaz de produzir expressões simples e isoladas sobre pessoas e lugares.

Fonte: QUADRO EUROPEU COMUM DE REFERÊNCIA PARA AS LÍNGUAS, 2001:91.

Quadro 21: Níveis Comuns de Referência – produção oral geral – escala global

	MONÓLOGO EM SEQUÊNCIA: descrever uma experiência
C2	É capaz de fazer descrições elaboradas e, por vezes, memoráveis, de forma clara e corrente.
C1	<p>É capaz de fazer descrições de assuntos complexos, de forma clara e pormenorizada.</p> <p>É capaz de fazer descrições e narrativas elaboradas, integrar subtemas, desenvolvendo questões específicas e concluindo de forma apropriada.</p>
B2	É capaz de fazer descrições acerca de uma vasto leque de assuntos relacionados com a sua área de interesse, de forma clara e pormenorizada.
B1	<p>É capaz de fazer descrições simples e diretas acerca de uma variedade de assuntos que lhe são familiares, da sua área de interesse.</p> <p>É capaz de relatar com fluência uma narrativa ou de fazer uma descrição simples como uma sequência linear de tópicos.</p> <p>É capaz de dar conta de experiências, descrevendo sentimentos e reações de forma pormenorizada.</p> <p>É capaz de relatar pormenores de ocorrências imprevisíveis, p. ex.: um acidente.</p> <p>É capaz de contar a intriga de um livro ou de um filme e de descrever as suas próprias reações.</p> <p>É capaz de descrever sonhos, esperanças e ambições.</p> <p>É capaz de descrever acontecimentos, reais ou imaginários.</p> <p>É capaz de contar uma história.</p>
A2	<p>É capaz de contar uma história ou de descrever algo, fazendo uma lista de tópicos.</p> <p>É capaz de descrever aspectos do seu dia-a-dia, p. ex.: as pessoas, os lugares, uma</p>

	<p>experiência de trabalho ou de estudo.</p> <p>É capaz de fazer descrições elementares e curtas de acontecimentos e de atividades.</p> <p>É capaz de descrever planos e preparativos, hábitos e rotinas, atividades passadas e experiências pessoais.</p> <p>É capaz de usar uma linguagem descritiva simples para fazer declarações breves acerca de objetos e do que lhe pertence, bem como fazer comparações.</p> <p>É capaz de explicar do que gosta ou do que não gosta em qualquer coisa.</p> <p>É capaz de descrever a sua família, as condições de vida, a sua formação, o seu trabalho atual ou mais recente.</p> <p>É capaz de descrever pessoas, lugares e coisas em termos simples.</p>
A1	É capaz de se descrever a si próprio, descrever o que faz e onde mora.

Quadro 22: Níveis Comuns de Referência – monólogo em sequência –descrever uma experiência

Fonte: QUADRO EUROPEU COMUM DE REFERÊNCIA PARA AS LÍNGUAS, 2001: 92.

O Quadro Europeu Comum de Referência de Línguas contempla 6 níveis de proficiência, o que se encaixaria momentaneamente muito bem ao teste de classificação que é dividido em 6 níveis; e a grade de avaliação seria uma ótima linha mestre para o professor/avaliador utilizar como guia para avaliar a produção oral do examinando, mesmo sabendo que, como é afirmado no documento que traz as diretrizes para o exame oral, a entrevista somente confirma a nota que o examinando obteve no teste escrito.

Como a ementa de Língua Inglesa enfatiza a oralidade como prioridade para o ensino nos cursos superiores analisados, a principal forma de avaliação para o aluno ingressante nesses cursos seria a entrevista oral.

5.5. Análise dos itens dos questionários: tentativas de caracterização da validade de face

De acordo com Brown (2001) a validade de face corresponde à opinião do examinando que está sendo avaliado em relação ao teste que o avalia, de forma que um teste terá validade de face se atender às expectativas dos examinandos com relação ao tipo de tarefa, tópicos, etc.

Neste trabalho, conforme já salientado anteriormente, a intenção é verificar se o teste de inglês aplicado aos alunos ingressantes de três cursos superiores em tecnologia

é válido para o seu propósito: avaliar diferentes níveis de desempenho de acordo com a ementa unificada dos cursos para que o aluno ingressante possa ser posicionado em um nível de conhecimento e uso de língua de acordo com a sua capacidade e competência. Para tanto, apresentaremos nesta seção os resultados referentes ao questionário II – Percepção dos itens do teste de classificação, aplicado a 106 alunos ingressantes em agosto de 2012; e o questionário III - *Experts Post-Test Questionnaire.*, respondido por duas professoras de inglês da instituição, como já foi apontado em outras seções deste trabalho. Discutiremos brevemente a opinião dos professores sobre o teste, buscando ampliar o universo do processo avaliativo para a caracterização da validade de face, de modo a complementar ou contrapor as interpretações sobre o teste, levando em conta dois polos extremos: alunos e professores.

5.5.1 Percepção dos examinandos: Apresentação dos Resultados do Questionário II - Percepção da prova – Placement Test – Agosto 2012

1) Qual o grau de dificuldade deste teste na parte de leitura?

Tabela 16: Grau de dificuldade na parte de leitura

	Nº Alunos	%
a) muito fácil	2	2%
b) fácil	7	7%
c) médio	64	60%
d) difícil	27	25%
e) muito difícil	6	6%

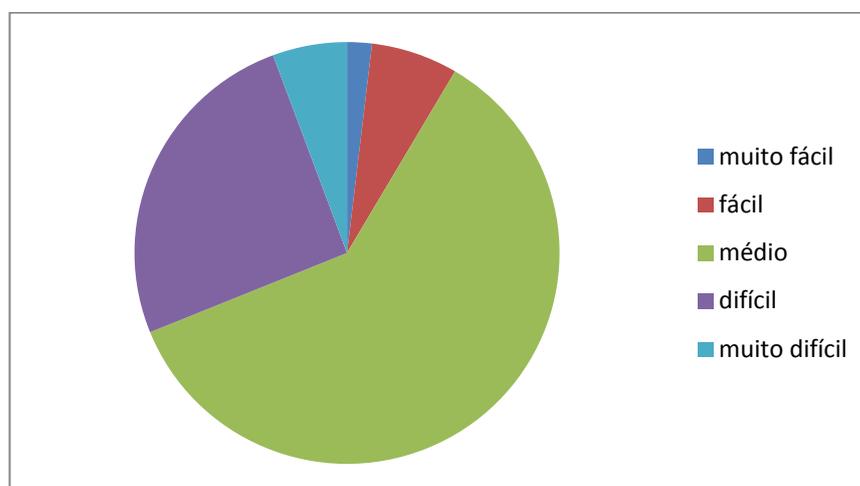


Gráfico 3: Grau de dificuldade na parte de leitura

Por meio do questionário para percepção da prova aplicado aos alunos ingressantes em agosto de 2012, podemos constatar que dos 106 alunos que responderam ao questionário, 60% (64 alunos) acharam que o teste apresentou um nível médio em relação às questões de leitura; 25% (27 alunos) acharam que o nível era difícil; 7% (7 alunos) que o nível era fácil; 6% (6 alunos) que o nível era muito difícil; e somente 2% (2 alunos) acharam que era muito fácil. Podemos perceber que os alunos apresentaram certo nível de dificuldade ao fazer o teste, o que remete para o conhecimento dos alunos e o grau de dificuldade dos itens.

2) Qual o grau de dificuldade deste teste na parte de “listening”?

Tabela 17: Grau de dificuldade na parte de “listening”

	Nº Alunos	%
a) muito fácil	5	5%
b) fácil	14	13%
c) médio	50	47%
d) difícil	29	27%
e) muito difícil	9	8%

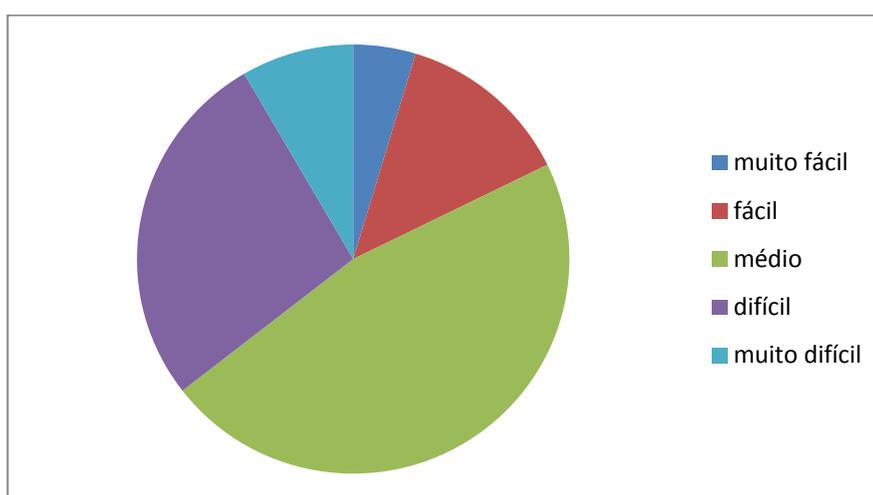


Gráfico 4: Grau de dificuldade na parte de “listening”

Em relação ao nível de dificuldade das atividades de recepção oral, as diferenças não foram tão gritantes em relação à habilidade de leitura, com dados muito próximos. Dos 106 alunos que responderam ao questionário, 47% (50 alunos) declararam que

acharam o teste de nível médio de dificuldade; 27% (29 alunos) acharam o teste difícil; 13% (14 alunos) acharam a prova fácil; 8% (9 alunos) responderam que o teste estava muito difícil; 5% (5 alunos) que estava muito fácil. Normalmente os alunos apresentam dificuldade nas questões de recepção oral, por ser uma das habilidades menos trabalhadas nas escolas de ensino médio. O resultado remete para um provável nível de dificuldade dos itens do teste de recepção escrita maior do que os de recepção oral

3) Considerando a extensão do teste, em relação ao tempo total, você considera que o teste foi

Tabela 18: Tempo para realização do teste

	Nº Alunos	%
a) muito longa	6	5%
b) longa	17	16%
c) adequada	68	66%
d) curta	10	9%
e) muito curta	5	4%

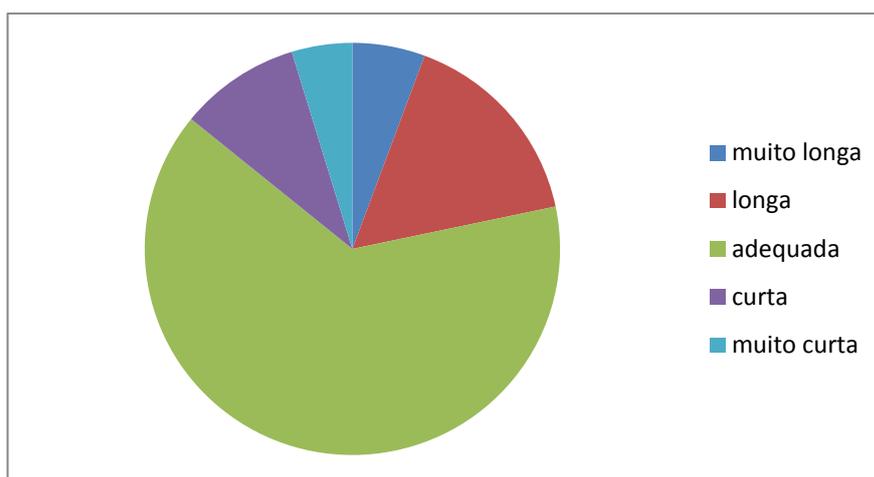


Gráfico 5: Tempo para realização do teste

Dos 106 alunos questionados, 66% (68 alunos) acharam a prova adequada em relação ao tempo para conclusão da mesma; 16% (17 alunos) responderam que a prova é muito longa; 9% (10 alunos) que a prova é curta; 5% (6 alunos) que a prova é muito

longa; 4% (5 alunos) que a prova é muito curta. O teste que é feito no computador, somando as partes de leitura e recepção oral tem a duração de 75 minutos. A parte de produção oral realizada em outro momento e somente para os alunos que foram aprovados nos níveis 1,2 e 3 dura em média 10 a 15 minutos.

4) Os enunciados dos itens do teste estavam claros e objetivos?

Tabela 19: Qualidade do enunciado dos itens

	Nº Alunos	%
a) Sim, todos	19	18%
b) Sim, a maioria	54	51%
c) Apenas cerca da metade	14	13%
d) Poucos	14	13%
e) Não, nenhum	6	5%

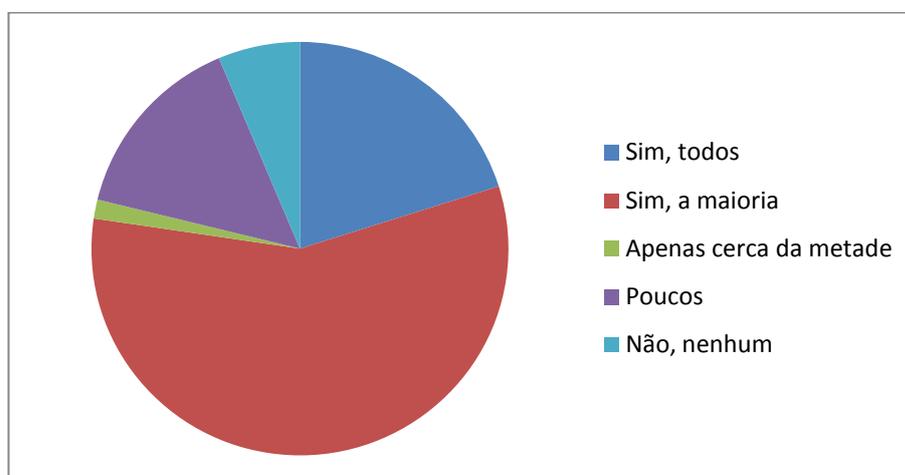


Gráfico 6: Qualidade do enunciado dos itens

Em relação a se os enunciados das questões estavam claros e objetivos 51% (54 alunos) responderam que a maioria deles estava; 18% (19 alunos) comentaram que todos os enunciados estavam claros; empatados com 13% (14 alunos) os dois grupos de alunos que responderam “apenas cerca da metade” e “poucos”; e finalmente 5% (6 alunos) disseram que não estavam claros. Pelos resultados parece-nos que a grande maioria dos alunos teve poucos problemas para entender o desenvolvimento do teste.

5) As informações/instruções fornecidas para a resolução dos itens foram suficientes para resolvê-los?

Tabela 20: Instruções do teste

	Nº Alunos	%
a) Sim, até excessivas.	5	5%
b) Sim, em todas elas.	14	13%
c) Sim, na maioria delas.	50	47%
d) Sim, somente em alguma	29	27%
e) Não, em nenhuma delas.	9	8%

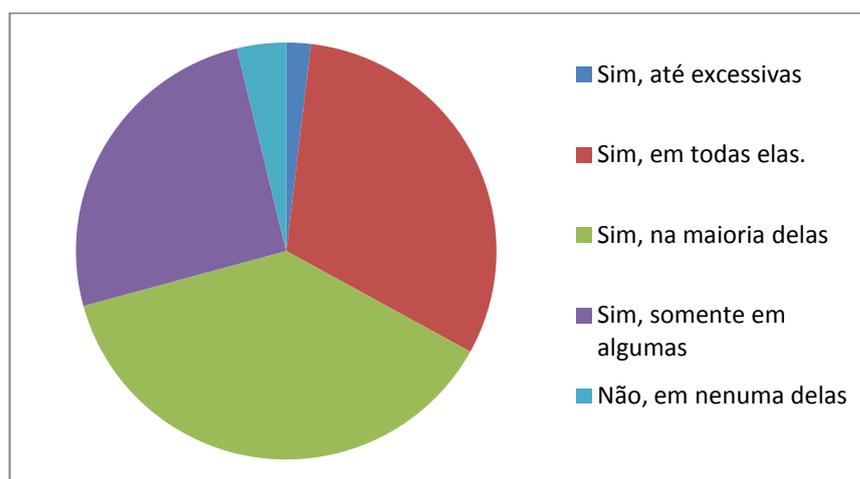


Gráfico 7: Instruções do teste

Fazendo-se um paralelo com a questão anterior, percebe-se que os alunos não tiveram muitos problemas para entender as instruções fornecidas para o desenvolvimento das questões, uma vez que 38% (40 alunos) afirmaram que as instruções foram fornecidas na maioria delas; 31% (33 alunos) comentaram que em todas elas; 25% (27 alunos) disseram que somente em algumas, e somente 4% (4 alunos) em nenhuma delas; e 2% (2 alunos) que as instruções eram excessivas.

6) Você se deparou com alguma dificuldade ao responder o teste Qual?

Tabela 21: Dificuldade ao responder o teste

	Nº Alunos	%
a) Desconhecimento do conteúdo.	26	25%
b) Falta de motivação para fazer a prova.	10	9%
c) Desconhecimento do vocabulário.	24	23%
d) Desconhecimento gramatical/estruturas.	38	36%
e) Não tive qualquer tipo de dificuldade para responder à prova.	8	7%

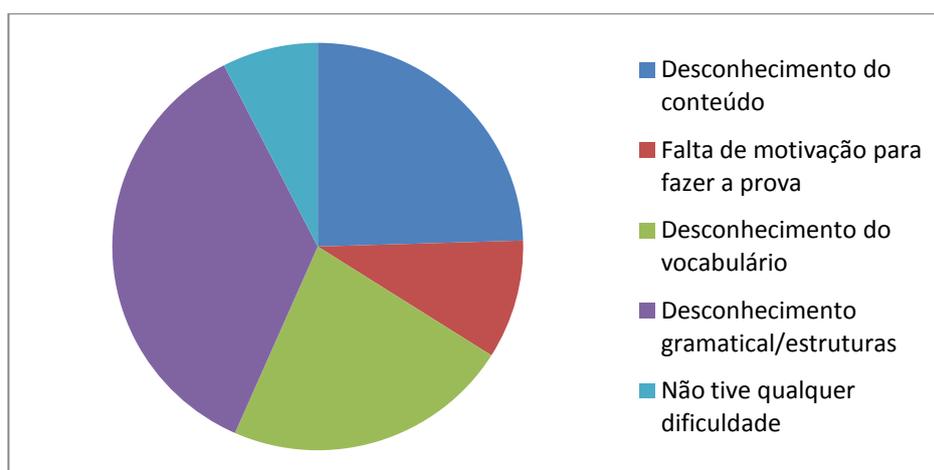


Gráfico 8: Dificuldade ao responder o teste

Voltando-nos agora para questões mais pontuais sobre o conteúdo do teste, podemos observar que algumas dificuldades que os alunos tiveram para responder às questões do teste foram sobre o desconhecimento de estruturas da língua e problemas gramaticais (36%, equivalente a 38 alunos); 25% (26 alunos) desconheciam o conteúdo da prova; 23% (24 alunos) tiveram problemas com o vocabulário; 9% (10 alunos) apresentaram falta de motivação para fazer a prova, e somente 7% (8 alunos) não tiveram problemas para realizar o teste. Podemos perceber pelos resultados que o grande problema dos alunos ingressantes está relacionado ao conteúdo do teste, total ou parcial, incluindo-se aí a gramática, o vocabulário e as estruturas da língua. A falta de motivação para fazer o teste pode tanto resultar de problemas externos como do conhecimento que o aluno possa ter da língua (pouco, com dificuldade ou razoável, sem dificuldade).

7) Considerando os itens do teste, você percebeu que

Tabela 22: Dificuldade no conteúdo do teste

	Nº Alunos	%
a) Não estudou ainda a maioria do conteúdo da prova.	33	31%
b) Estudou alguns desses conteúdos, mas não os aprendeu.	29	27%
c) Estudou a maioria desses conteúdos, mas não os aprendeu.	15	14%
d) Estudou e aprendeu muitos desses conteúdos.	23	22%
e) Estudou e aprendeu todos esses conteúdos.	6	6%

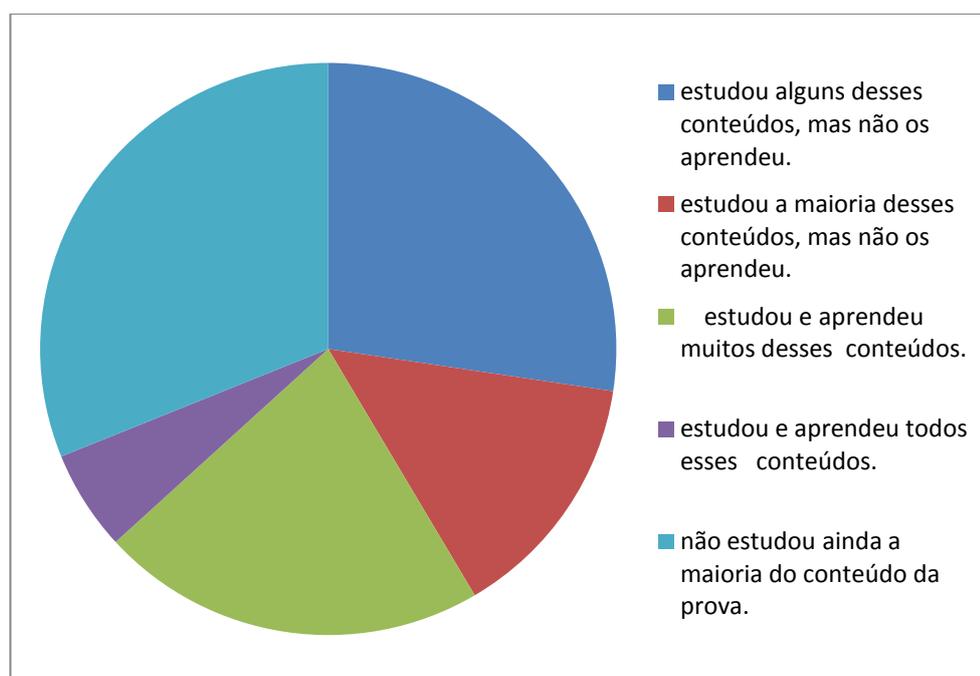


Gráfico 9: Dificuldade no conteúdo do teste

Ratificando as respostas do item anterior, constata-se novamente que o problema dos alunos ingressantes no ensino superior é em relação ao que eles aprenderam de língua inglesa nos anteriores, pois 31% (33 alunos) afirmaram que ainda não estudaram a maioria do conteúdo do teste; 27% (29 alunos) apontaram que já estudaram alguns itens do conteúdo, mas não os aprendeu; e 14% (15 alunos) que estudaram a maioria dos conteúdos, mas não os aprendeu. No entanto, felizmente, 22% (23 alunos) disseram

que já tinham estudado o conteúdo e aprendido muitos deles e 6% (6 alunos) que estudaram e aprenderam todo o conteúdo do teste. Novamente, percebemos que o aluno que chega ao ensino superior apresenta um déficit no conhecimento da língua estrangeira.

8) Em que nível você chegou ao fazer a prova?

a) 1 b) 2 c) 3 d) 4 e) 5 f) 6

Tabela 23: Nível atingido no teste

	Nº Alunos	%
a) 1	58	56%
b) 2	12	11%
c) 3	16	15%
d) 4	9	8%
e) 5	9	8%
f) 6	2	2%

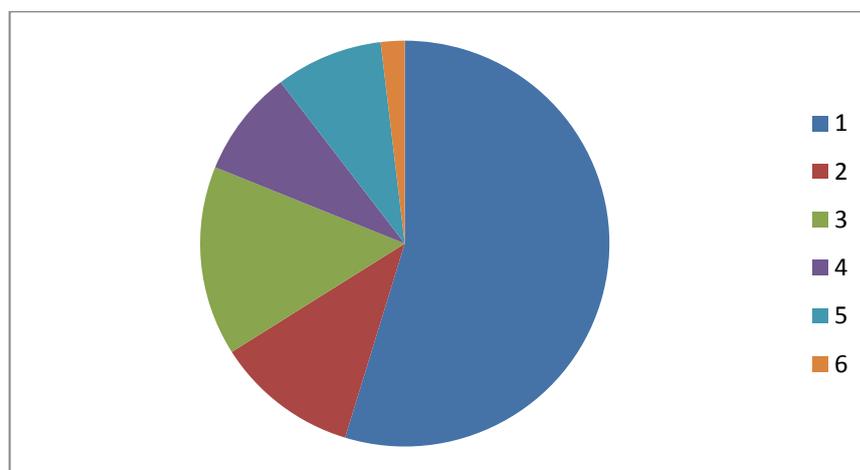


Gráfico 10: Nível atingido no teste

Novamente, ratificando os resultados das questões 6 e 7 podemos verificar que 56% (58 alunos) se mantém no nível 1 do teste, provavelmente por problemas de conteúdo, 15% (16 alunos) conseguiram chegar ao nível 3; 11% (12 alunos) no nível 2. Em relação aos níveis 4 e 5, apenas 8%, o equivalente a 18 alunos, 9 alunos (4%) para cada nível e no nível 6 somente 2% (2 alunos).

8) Eu me senti confortável em fazer o teste no computador?

Tabela 24: Aceitabilidade da realização do teste no computador

	Nº Alunos	%
a) Concordo totalmente.	34	32%
b) Concordo.	41	38%
c) Não concordo nem discordo – neutro.	23	22%
d) Discordo.	5	5%
e) Discordo Totalmente.	3	3%

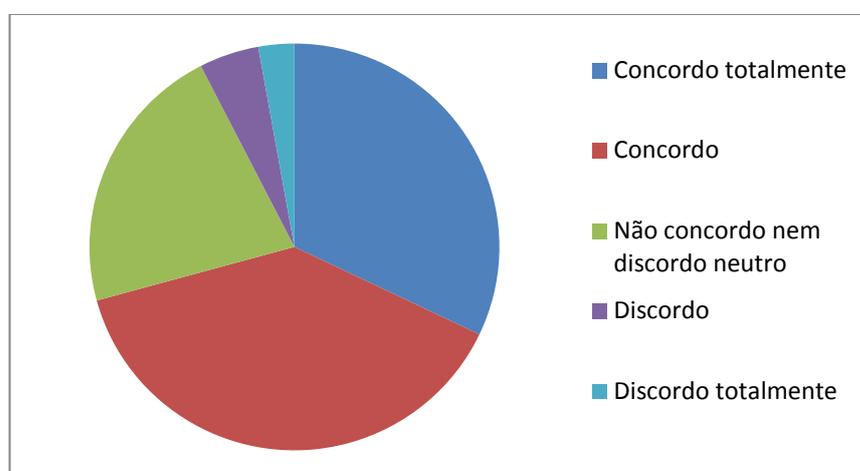


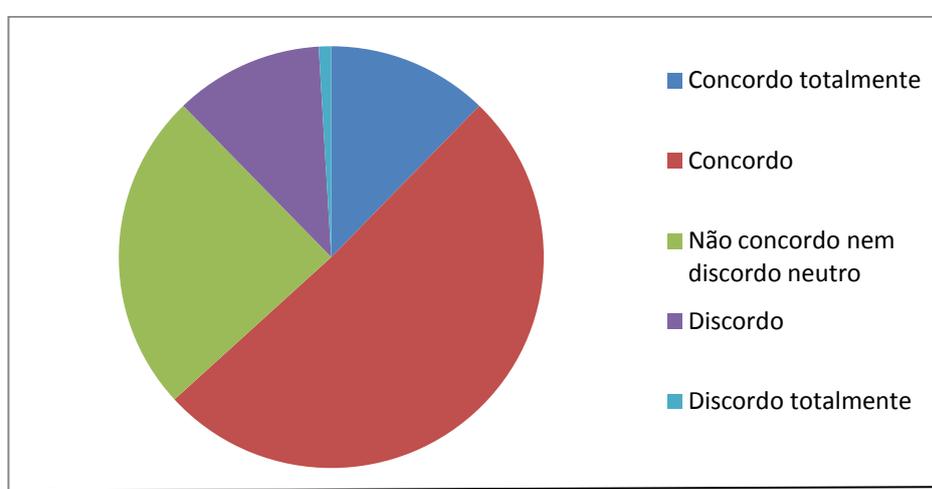
Gráfico 11: Aceitabilidade da realização do teste no computador

Uma das grandes dúvidas sobre a aplicação do teste mediado pelo computador seria a dificuldade que os examinandos teriam para utilizar esse meio; 38% (41 alunos) disseram que se sentiram confortáveis em fazer o teste por esse meio; 32% (34 alunos) inclusive enfatizaram a satisfação concordando totalmente, talvez por estarem ingressando em uma faculdade de tecnologia e já possuírem algum conhecimento no uso dessa ferramenta; 22% (23 alunos) se mantiveram neutros, sem concordar ou discordar e 5% (5 alunos) demonstraram seu desconforto, sendo que 3% (3 alunos) enfatizaram esse desconforto.

10) Há uma boa variedade de tipos de exercícios, deixando a prova agradável para se fazer.

Tabela 25: Variedade de itens no teste

	Nº Alunos	%
a) Concordo totalmente.	13	12%
b) Concordo.	54	52%
c) Não concordo nem discordo – neutro.	26	24%
d) Discordo.	12	11%
e) Discordo Totalmente.	1	1%

**Gráfico 12:** Variedade de itens no teste

Em relação ao tipo de exercícios que o teste oferece 52% (54 alunos) opinaram que o teste apresenta exercícios variados; 24% (26 alunos) se mantiveram neutros; 12% (13 alunos) concordaram com a afirmação de que os exercícios são variados, deixando a prova agradável para ser feita; 11% (12 alunos) não gostaram do teste; e somente 1% (1 aluno) apontou que discordava totalmente que o teste era agradável e com variedades de exercícios.

5.5.2. A percepção dos professores: Apresentação dos Resultados do Questionário III – Experts Post-Test Questionnaire

Os professores também opinaram sobre os testes. Duas professoras de inglês, conforme já assinalado anteriormente responderam às perguntas do Questionário III,

baseado em Papadyma (2005:208), porém com algumas modificações, cujo objetivo é auxiliar na validação do processo avaliativo por meio dos testes de classificação e garantir e melhorar a qualidade dos cursos da instituição.

A seguir, apresentamos o levantamento e a análise dos dados do Questionário III.

This questionnaire is part of the validation process of the Placement Test (henceforth – PT) designed for the Curso Superior em Tecnologia – (henceforth CST). The aim of PT is to improve the academic quality and efficiency of the CST in order to establish its content validity. The questionnaire is entirely anonymous.

1 – Strongly disagree

2 – Disagree

3- Neutral

4 – Agree

5 – Strongly agree

	Prof 1	Prof 2
1. PT meets the requirements of an English placement test for the needs of CST students.	4	4
2. PT accurately represents the content of CST English Language Programme at the various levels.	2	3
3. The grammatical points represent the structure content of the respective levels.	4	3
4. The lexical points represent the vocabulary content of the respective levels.	4	3
5. The reading comprehension items represent the level of understanding of reading texts of the respective levels.	4	3
6. The topics are appropriate for the respective levels.	4	3
7. The topics are interesting.	4	4
8. The instructions are clear.	5	2
9. The test reflects settings test-takers would come across in their studies.	4	3
10. The test reflects settings and contexts in English speaking settings.	5	3
11. The test reflects the setting where test-takers will be studying.	2	3
12. There is sufficient variety in test types, appropriate for the respective levels.	2	3
13. PT items generally target the appropriate English level.	4	3

14. The kinds of activities are appropriate.	5	3
15. The activity types cover variety of learning styles.	3	3
16. The time given to the test-takers to take the test is appropriate.	5	2
17. There is time flexibility to cater for test-takers' difference in time needs.	5	4
18. Item selection criteria are appropriate.	5	4
19. The scoring system is satisfactory.	5	4
20. The test results are appropriately reported to the test-takers.	5	4
21. The instructions help test-takers familiarize themselves with the test, the test format, the test layout and the test items.	3	2
22. The computer screen is clear.	4	2
23. The activity layout is clear.	4	2
24. The visuals are appropriate.	4	2
25. The students are given opportunities to control the test (Review items, going backwards and towards, etc.).	1	2

Questão 26: Tell what you did not like about the PT.

Prof. 1- In fact that it focuses too much on business English, but teaching business English is not a generally accepted position in the CST.

Prof. 2 – I guess there should be a kind of “mock test” to get the students aware of the exercises, system, time. By doing so many questions that pop up during the exam would be solved beforehand.

Questão 27: Tell what you liked most about PT (if anything).

Prof. 1 – The grammatical structure of the different levels is appropriate.

Prof. 2 –It's incredible to think that students can be evaluated through an online test that teachers can manage to open and close from time to time. It is an exam similar to the international ones (TELP, for example).

Questão 28 – Tell how PT can be improved.

Prof. 1 – By including more general English and by making all students take the test (another edition) again after one year and a half.

Prof. 2 – I guess the test should not depend on the internet. Students should be able to go back and forth in their questions as well.

Questão 29 – Do you have problems with students that aren't placed correctly?

Prof.1 – A few students claim that they didn't take the test seriously and were placed in the wrong level. This is not a general complaint.

Prof. 2 – As I have been working at the institution for less than a year, I have little experience with this, but, as far as I'm concerned, the exams I have held showed that the speaking abilities of the students do not match their writing skills at many times.

Questão 30 - Do you have problems with students that don't have to attend classes for one or several terms and they have to go back later? Which are the problems?

Prof. 1 – I've had problems with students who have been able to skip one or two semesters at most. When they come back, they sometimes disturb the "flowing" of the class, because they are not used to it or because they have problems in their oral performance. With students who have been dismissed for longer periods, I have never had problems.

Prof. 2 – Because of my lack of experience there are only a few points I'd like to share. Students that come back to classes after being dismissed generally have their listening skills compromised. Another point I could notice is that the students feels like an "outsider" in the classroom for a while.

As afirmações das duas professoras que lecionam nos cursos representados neste estudo indicam que o teste tem validade de face, pois elas forneceram algumas razões e evidências que apoiam suas afirmações. Entretanto, algumas críticas foram indicadas, o que é revelador dos problemas que os protagonistas enfrentam ao serem submetidos ao teste.

As duas professoras concordam que o teste contempla as necessidades dos alunos de cursos superiores em tecnologia (1), os tópicos são interessantes (7) e as atividades do teste abordam estilos diferentes de aprendizagem (15).

A opinião das duas professoras que responderam ao questionário é discordante em relação aos itens: (8) as instruções são claras, (10) o teste reflete contextos de uso da língua oral, (14) os tipos de atividades são adequados; (16) o tempo dado para os examinandos para fazer o teste é apropriado, (22), a tela do computador é clara; (23) o layout das atividades é claro; e os visuais são apropriados (24). Enquanto uma delas discorda dessas afirmações propostas no questionário, a outra concorda.

Em relação aos outros itens do questionário, as respostas das duas professoras foram muito próximas. Entretanto, podemos verificar que as entrevistadas apresentam posições negativas em relação ao que está afirmado nos itens (2) O teste de classificação representa exatamente nos seus vários níveis o conteúdo do programa de Inglês para os Cursos Superiores em Tecnologia; (11) o teste representa o cenário no qual os examinandos estudarão; (12) Há uma variedade suficiente em tipos de teste apropriados para os diferentes níveis; (21) As instruções ajudam os examinandos a se familiarizarem com o teste, seu formato, layout e seus itens; (25) os examinandos têm a possibilidade de controle do teste (revisar itens, voltar e ir para frente).

Uma das críticas apresentadas é a de que o teste está centrado em Inglês para Negócios. No entanto, as ementas de cada nível do curso não afirmam que esse seja o foco dos cursos analisados na unidade de ensino. Outro problema grave apontado é que alguns alunos não levam o teste a sério e acham que é um teste sem valor para o seu desempenho acadêmico. Assim, acabam fazendo o teste sem a atenção necessária e o resultado é negativo, posicionando o aluno em um nível aquém de sua competência. Provavelmente está faltando um nível de conscientização maior para esse aluno, alertando-o sobre suas decisões, que terão um peso muito grande na sua vida acadêmica futura. A percepção que o examinando tem sobre o teste, e o modo como deve se preparar para fazê-lo, não no sentido de estudar um determinado conteúdo, mas de se envolver com a prova, sua motivação ao fazer o teste, as estratégias que utilizará para resolver os itens apresentados são fatores decisivos para um bom resultado do examinando e do teste, com benefícios para ele e a instituição.

O problema de dispensa o aluno por um ou mais semestres e a ausência do aluno pelo período de seis meses a um ano também foi notado como um ponto negativo, pois alguns alunos retornam com grandes dificuldades, não só no conteúdo, mas no

entrosamento com o grupo na classe (alguns alunos estão juntos desde o primeiro semestre do curso e seu desempenho, em alguns casos, melhora ao manter-se no grupo). Muitos alunos que estiveram ausentes do curso por um período têm nitidamente a percepção de que seu desempenho piorou, segundo as docentes.

Vale aqui ressaltar, que em essência, a validade de face é importante para a aceitação que esse teste tem socialmente, e que as percepções que os protagonistas têm sobre a validade desse teste influenciam no crédito que é dado a ele.

Assim, finalizamos a exposição dos resultados nesta seção e partimos para as discussões dos mesmos no próximo capítulo, revendo conceitos e retomando as perguntas de pesquisa que nortearam o trabalho.

CAPÍTULO 6 - O PROCESSO DE VALIDAÇÃO E A CARACTERIZAÇÃO DAS VALIDADES DE CONSTRUTO, DE CONTEÚDO E DE FACE: DISCUSSÃO DOS RESULTADOS

A análise desenvolvida neste trabalho sobre o instrumento de avaliação escolhido como referência para o tratamento e caracterização das validades de construto, de conteúdo e de face apoiou-se nos dados obtidos nos questionários respondidos pelos alunos e professores já apresentados no capítulo anterior.

Neste trabalho não tivemos como foco a análise de pessoas, mas a análise de um instrumento de avaliação, sendo que essa análise foi apoiada pela ajuda da análise de outras pessoas que estavam envolvidas como o teste de classificação, foco desta pesquisa. Contamos com a ajuda dos próprios examinandos, que nos deram suas opiniões sobre o teste e de professores envolvidos na aplicação do mesmo.

Com o objetivo de sistematizar os resultados deste estudo, primeiramente nos deteremos em realizar uma visão geral do teste, retomando em seguida as perguntas de pesquisa que nortearam este trabalho.

6.1. Visão geral do teste

Comparativamente com outros tipos de testes, os testes de classificação (*placement*) têm recebido pouca atenção por pesquisadores da área de línguas. Alguns autores como Hughes (1989); Bachman (1990); Alderson et al.(1995); Bachman and Palmer (1996); Brown (1996) apontam dois papéis fundamentais aos testes de classificação:

- a) auxiliar na criação de grupos homogêneos de estudantes em termos de habilidade linguística; e
- b) averiguar o domínio dos alunos em relação ao conteúdo do curso.

Vários autores têm discutido o potencial dos resultados dos testes de classificação para informar como está o ensino e contribuir para a avaliação de programas educacionais. Wall et AL. (1994) descreve o uso de testes de classificação para EAP (*English for Specific Purposes*) – Inglês para Fins Específicos como forma de diagnóstico para a deficiência de determinadas áreas que podem ser trabalhadas por

ESL (*English as a Second Language*) – Inglês como Segunda Língua. Brown (1989, p. 79) também afirma que há implicações curriculares no desenvolvimento de testes de classificação.

O teste que é objeto de estudo neste trabalho serve, na verdade, para dois objetivos: é um teste de classificação, pois é aplicado aos alunos ingressantes dos cursos superiores em tecnologia para o posicionamento desses junto a colegas que tenham o mesmo nível linguístico, a fim de que se tenham turmas homogêneas, entretanto, também é um teste de proficiência, uma vez que esse aluno faz o teste ao ingressar em um curso, sem se levar em conta qual é o seu conhecimento anterior sobre a língua, somente seu uso futuro. Para justificar essa hipótese, deveríamos estabelecer primeiramente que as tarefas que aparecem no teste poderiam ser posicionadas em termos de dificuldade e que esse posicionamento seria aplicado a todos os alunos no programa, sem se olhar seu background. Em outras palavras, temos que estar confiantes que tanto a dificuldade dos itens do teste quanto à habilidade de cada examinando possa ser posicionada em uma escala unidimensional, que nos permita classificar os alunos de forma consistente, no que diz respeito a gama de habilidades a serem testadas (Henning, 1987). Em segundo lugar, as tarefas do teste devem permitir inferências válidas sobre o domínio do conhecimento, habilidades ou competências ensinadas durante o curso ou previamente.

Além disso, baseado no seu desempenho, e de acordo com a média que esse aluno obtiver, que tem que ser 6,0 ou superior a isso, o aluno poderá ser dispensado da frequência das aulas de Inglês em todos os semestres, efetivando desta forma seu conhecimento sobre a língua, naquilo que é esperado que esse aluno consiga desempenhar.

De acordo com as ementas de língua inglesa que são unificadas para a grande maioria dos cursos, (exceção de Comércio Exterior e Secretariado, pois os alunos têm o dobro de aulas de Inglês por semana), podemos perceber que o que subjaz como abordagem de ensino é a abordagem comunicativa.

Ao se trabalhar a abordagem comunicativa, diferentes habilidades de uma língua são enfocadas, como: compreensão aural, leitura, escrita e fala. Seguindo essa linha, as avaliações devem testar as mesmas habilidades, de forma que os alunos deveriam ser

testados em sua compreensão aural, textual, expressão oral, escrita, de estruturas gramaticais e vocabulário.

Entretanto, ao se falar em avaliação, temos que pensar que ela é um elemento fundamental em todo o processo de ensino e aprendizagem, de forma que deve ter objetivos coerentes com a visão de linguagem que orienta o ensino e com a abordagem inferida através da ementa apresentada para a disciplina.

Hughes (1989) e Bailey (1998) afirmam que esse tipo de avaliação deve espelhar em suas tarefas o conteúdo do currículo dos níveis a serem avaliados, refletindo em suas tarefas o construto elaborado pela instituição. Além disso, a construção de um bom teste de classificação/proficiência significa a tomada de decisões coerentes em relação ao nível que o aluno deve ser posicionado.

O conteúdo do teste de classificação também deveria abordar as mesmas habilidades expostas na ementa: as habilidades de falar, compreender, ler, escrever, de forma integrada, sendo que os aspectos linguísticos, gramatical, vocabulário e pronúncia, deveriam estar apresentados de forma contextualizada. O que se observa é que não há uma relação coerente entre o conteúdo proposto na ementa e a forma como as habilidades são avaliadas no teste, uma vez que o teste somente contempla, nos três primeiros níveis, questões de compreensão oral e compreensão escrita; e somente nos três últimos níveis, além das questões de compreensão escrita e de recepção oral, há uma entrevista em Inglês, utilizando o nivelamento da parte escrita para confirmar o desempenho da parte oral. Temos, neste caso, uma avaliação holística, correspondendo a uma nota atribuída pelo avaliador a partir do desempenho do examinando, que no caso do teste em questão, é o próprio professor da disciplina. Entretanto, o avaliador não possui descritores de desempenho, que possam auxiliá-lo na interpretação da produção do examinando.

O que acontece nesse caso é o que é denominado de *efeito halo* (Santos, 2007), que define que o bom desempenho de um examinando em uma habilidade tem influência no desempenho das demais habilidades. No caso deste estudo, por exemplo, um examinando que se saia muito bem na compreensão escrita, acertando uma quantidade suficiente para ser considerada como um bom desempenho deva ser exatamente igual à parte de recepção e produção oral.

Como estamos falando de um teste para cursos superiores em tecnologia, muitos alunos que ingressam nesses cursos, principalmente os alunos que procuraram os cursos de Gestão da Tecnologia da Informação e Análise e Desenvolvimento de Sistemas apresentam um excelente desempenho na habilidade de leitura, pois estão acostumados a utilizar videogames e outros programas para computador, normalmente em inglês, entretanto, não conseguem nem mesmo reproduzir oralmente o que está escrito, e quando precisam falar determinada palavra, que aparecem nos jogos ou programas, preferem soletrar as letras, para que haja um entendimento do vocábulo. Inclusive não sabem expressões que são utilizadas no dia a dia.

Entendemos que, como é um teste de classificação, e este precisa trazer a possibilidade de avaliar rapidamente e com precisão, por isso é que foi escolhida a modalidade eletrônica e o formato de múltipla escolha, um total de 60 questões com 4 alternativas para cada questão, para que os alunos pudessem fazê-lo, em, no máximo 75 minutos. No entanto, o teste não abarca vários itens que estão discriminados nas ementas de cada nível de inglês, além disso, se o propósito do curso de Inglês é dar ênfase à oralidade, somente duas questões de recepção oral nos três primeiros níveis e em seguida três questões de recepção oral nos três últimos níveis é insuficiente para avaliar um dos aspectos da oralidade, mesmo a quantidade de questões para cada nível não é suficiente para a avaliação do candidato.

Vários especialistas consideram os testes de múltipla escolha particularmente adequados para avaliar atributos mentais e apontam que a amostragem do conteúdo é geralmente superior quando comparado com outros formatos. Essa flexibilidade para acomodar uma diversidade de conteúdo significa que o formato de múltipla escolha pode ser usado em uma variedade de instrumentos de avaliação educacional e psicológica.

Um aspecto da flexibilidade do formato de testes de múltipla escolha é que, quando as questões são bem formuladas, o formato permite a avaliação de complexas e sofisticadas habilidades mentais, que envolvem habilidades cognitivas.

Outra vantagem do formato de questões de múltipla escolha é que permitem uma interpretação precisa, que pode levar a uma evidência importante da validade de conteúdo de uma prova.

Entretanto, uma questão ainda controversa é referente à quantidade de opções de resposta para cada teste. Há autores que recomendam a opção de 5 opções, (uma correta e quatro distratores), cujo argumento se baseia na diminuição da probabilidade de “chutes”, sendo a proporção de 20% de chance. Outros propõem 4 opções, cujo argumento está pautado na visão de que não adianta incluir distratores que não sejam plausíveis, isto é, distratores que não serão selecionados por menos de 20% dos examinandos. Um bom distrator, de acordo com Haladyna (1994), é aquele que é escolhido pelos examinandos menos preparados e capazes e ignorado pelos mais capazes.

Shohamy (1995: 188) aponta que houve uma busca no desenvolvimento e no uso de testes que ressaltassem mais características do uso real da língua e que cobrem, no desempenho dos alunos, uma língua mais autêntica, direta e comunicativa. Acreditava-se que tais testes refletiam melhor o que quer dizer “saber uma língua”. Entretanto o teste não mensura o desempenho do aluno para compreender o uso real da língua diante de uma situação, apesar da questão apresentar uma situação real de sala de aula, na qual o professor está se apresentando aos alunos. Poderíamos dizer que representa um teste autêntico, pois apresenta uma situação que reproduz o uso real da língua, porém, o teste utiliza como meio para a orientação da tarefa a língua materna, para em seguida, utilizar a língua inglesa como fim para a realização da mesma.

O examinando necessariamente não precisaria saber se comunicar na língua alvo para responder ao item. Ele poderia estar atento somente à estrutura gramatical da língua, e poderia respondê-la utilizando o recurso de comparação entre a língua alvo e a língua materna.

Seguem abaixo tabelas e gráficos que demonstram nos quatro cursos envolvidos nesta pesquisa, Eventos, Logística (manhã), Logística (noite) e Gestão da Tecnologia da Informática, como os alunos ingressantes nesses cursos foram posicionados em cada nível de acordo com seu desempenho no teste. Em seguida, elaboramos uma tabela final, com porcentagem para visualizarmos possíveis problemas no conteúdo do teste, de acordo com a quantidade de alunos que foram posicionados nos níveis.

Tabela 26: Posicionamento dos alunos nos níveis - Eventos

Nível	2º sem. 2010	1º sem. 2011	2º sem. 2011	1º sem. 2012	2º sem 2012
Nível 1	7	36	29	17	31
Nível 2	4	1	2	11	2
Nível 3	10	14	5	4	0
Nível 4	5	5	1	3	0
Nível 5	2	10	5	3	0
Nível 6	1	3	2	0	0
Dispensado	0	1	0	0	0

Tabela 27: Posicionamento dos alunos nos níveis - Logística (manhã)

Nível	2º sem. 2010	1º sem. 2011	2º sem. 2011	1º sem. 2012	2º sem 2012
Nível 1	27	27	28	28	22
Nível 2	5	1	2	4	2
Nível 3	5	3	6	0	2
Nível 4	2	1	0	3	2
Nível 5	1	1	1	1	0
Nível 6	0	1	2	0	1
Dispensado	1	1	0	0	0

Tabela 28: Posicionamento dos alunos nos níveis - Logística (noite)

Nível	2º sem. 2010	1º sem. 2011	2º sem. 2011	1º sem. 2012	2º sem 2012
Nível 1	17	16	25	12	26
Nível 2	6	1	1	10	5
Nível 3	9	10	4	6	2
Nível 4	5	3	1	3	0
Nível 5	4	4	1	4	0
Nível 6	1	2	0	3	3
Dispensado	17	1	2	12	26

Tabela 29: Posicionamento dos alunos nos níveis - Gestão da Tecnologia da Informação

Nível	2º sem. 2010	1º sem. 2011	2º sem. 2011	1º sem. 2012	2º sem 2012
Nível 1	3	5	13	6	20
Nível 2	0	6	4	12	18
Nível 3	1	12	7	6	4
Nível 4	0	1	1	9	2
Nível 5	1	5	6	5	0
Nível 6	0	3	0	1	3
Dispensado	0	3	3	6	0

As tabelas a seguir representam os resultados do teste de classificação dividido em faixas de nivelamento. Em cada quadro temos o período de pesquisa, a quantidade

de examinandos que fizeram o teste, o número de examinandos e a porcentagem referente a cada faixa de nivelamento. Representamos em seguida a quantidade de alunos em cada faixa nas 5 edições do teste e a última tabela aponta a porcentagem dos alunos em cada faixa de nivelamento no período que compreende do 2º semestre de 2010 ao 2º semestre de 2012.

Tabela 30: Número/porcentagem de examinandos por nível no 2º semestre de 2010

Nível (2010)	1	2	3	4	5	6	Dispensado
Eventos 29 alunos	7 25%	4 14,2%	10 35,7%	5 17,8%	2 7,1%	1 3,5%	0 0%
LOG M 41 alunos	27 65,8%	5 12,1%	5 12,1%	2 4,8%	1 2,4%	1 2,4%	0 0%
LOG N 42 alunos	17 40,4%	6 14,2%	9 21,4%	5 11,9%	4 9,5%	1 2,3%	0 0%
GTI 5 alunos	3 60%	0 0%	1 20%	0 0%	1 20%	0 0%	0 60%

Tabela 31 : Número/porcentagem de examinandos por nível no 1º semestre de 2011

Nível (2011-1)	1	2	3	4	5	6	Dispensado
Eventos 49 alunos	18 34,6%	1 1,9%	14 26,9%	5 9,6%	10 19,2%	3 5,7%	1 1,9%
LOG M 35 alunos	27 65,8%	5 12,1%	5 12,1%	2 4,8%	1 2,4%	1 2,4%	0 0%
LOG N 37 alunos	17 40,4%	6 14,2%	9 21,4%	5 11,9%	4 9,5%	1 2,3%	0 0%
GTI 35 alunos	3 60%	0 0%	1 20%	0 0%	1 20%	0 0%	0 60%

Tabela 32 : Número/porcentagem de examinandos por nível no 2º semestre de 2011

Nível (2011-2)	1	2	3	4	5	6	Dispensado
Eventos 44 alunos	29 65%	2 4,5%	5 11,3%	1 2,2%	5 11,3%	2 4,5%	0 0%
LOG M 41 alunos	28 71,7%	2 5,1%	6 15,3%	0 0%	1 2,5%	2 5,1%	0 0%
LOG N 42 alunos	25 73,5%	1 2,9%	4 11,7%	1 2,9%	1 2,9%	0 0%	2 5,8%
GTI 34 alunos	13 38,2%	4 11,7%	7 20,5%	1 2,9%	6 17,6%	3 8,7%	0 0%

Tabela 33 : Número/porcentagem de examinandos por nível no 1º semestre de 2012

Nível (2012-1)	1	2	3	4	5	6	Dispensado
Eventos 38 alunos	17 44,7%	11 28,9%	4 10,5%	3 7,8%	3 13%	0 0%	0 0%
LOG M 36 alunos	28 77%	4 11,1%	0 0%	3 8,3%	1 2,7%	0 0%	0 0%
LOG N 38 alunos	12 31,5%	10 26,3%	6 15,7%	3 7,8%	4 10,5%	3 7,8%	0 5,8%
GTI 39 alunos	6 15,3%	12 30,6%	6 15,3%	9 23%	5 12,8%	1 2,5%	0 0%

Tabela 34: Número/porcentagem de examinandos por nível no 2º semestre de 2012

Nível (2012-2)	1	2	3	4	5	6	Dispensado
Eventos 33 alunos	31 93%	2 7%	0 0%	0 0%	0 0%	0 0%	0 0%
LOG M 32 alunos	25 78,1%	2 6,2%	2 6,2%	2 6,2%	0 0%	1 3,1%	0 0%
LOG N 36 alunos	26 72,2%	5 13,8%	2 5,5%	0 0%	0 0%	3 8,3%	0 5,8%
GTI 37 alunos	20 54%	8 21,6%	4 10,8%	2 5,4%	0 0%	3 8,1%	0 0%

Tabela 35: Número de examinandos por nível

Nível	1	2	3	4	5	6	D
Eventos	102	20	33	14	20	1	1
LOG M	135	16	16	8	4	5	1
LOG N	96	23	31	12	13	9	3
GTI	47	30	30	13	17	10	3

Tabela 36: Porcentagem de examinandos por nível

Nível	1	2	3	4	5	6	Dispensado
Eventos 196 alunos	52%	10,2%	16,8%	7,1%	10,2%	3,1%	0,5%
LOG M 185 alunos	72,9%	8,6%	8,6%	4,3%	2,1%	2,7%	0,5%
LOG N 187 alunos	48,7%	12,2%	16,5%	6,4%	6,9%	4,8%	1,6%
GTI 150 alunos	31,3%	20%	20%	8,6%	11,3%	6,6%	2%

Ao se fazer uma análise quantitativa do desempenho dos examinandos no teste de acordo com o nível atingido, observamos que massivamente os alunos são

posicionados no nível 1 do teste , começando pelos alunos de Logística (manhã) 72,9%, seguido de Eventos (52%), Logística (noturno) 48,7% e Gestão da Tecnologia da Informação 31,3%.

Podemos observar que os examinandos apresentam uma maior dificuldade nos níveis 2 e 4 do teste, uma vez que a porcentagem dos alunos que são posicionados nesses níveis apresenta um decréscimo em relação ao nível 3 e, em alguns casos, em relação ao nível 4. No curso de Logística (manhã) 8,6% dos examinandos foram posicionados no nível 2, a mesma porcentagem no nível 3, e 4,3% no nível 4, reduzindo a porcentagem nos níveis subsequentes e somente 0,5% dos examinandos, ou apenas 1 aluno conseguiu ser dispensado de cursar a disciplina Inglês, por ter um nível de proficiência compatível com o que era esperado pelo curso.

No curso de Eventos 10,2% dos examinandos foram posicionados no nível 2, sendo que no nível 3 temos 16,8%, e no nível 4 apenas 7,1% dos examinandos foram posicionados neste nível, sendo ainda que apenas 0,5% ou somente 1 examinando conseguiu ser dispensado da disciplina nos 6 semestres. Em Logística (noturno) temos números muito parecidos, sendo que 10,5% foram posicionados no nível 2; 16,3% no nível 3 e novamente o nível 4 apresenta um decréscimo, com 10,5% dos alunos posicionados nesse nível.

No curso de Logística (noturno) 12,2% foram posicionados no nível 2; 16,5% no nível 3; e 6,4% no nível 4. Assim como nos outros cursos, a maior discrepância pode ser observada no nível 4, pois a quantidade de alunos posicionados neste nível é inferior às do nível 2 e 3.

O curso que apresenta certa estabilidade é o curso de Gestão da Tecnologia da Informação, uma vez que os alunos são posicionados de uma forma equilibrada, sendo 31,3% no nível 1, 20% nos níveis 2 e 3, 8,6% no nível 4; 11,3% no nível 5; 6,6% no nível 6 e 3 alunos dispensados de cursar a disciplina Inglês nos 6 semestres. O fato que nos chama a atenção é o posicionamento dos alunos em relação ao nível 4, pois os alunos estavam sendo posicionados em uma curva descendente, uma vez que os itens que compõem o teste exigem mais habilidade desse examinando de acordo com o nível, sendo que nos últimos níveis deveríamos ter menos alunos; entretanto, o que se pode constatar é que o nível 4 parece ser aquele que apresenta dificuldade maior.

Ao se analisar o conteúdo do teste referente ao nível 4, constatamos que os itens que compõem esse nível são contemplados com itens gramaticais que apresentam estruturas que envolvem o uso de verbos modais como *should, could, would, to be allowed to, have to*. A ementa não é clara quanto ao uso desse tópico gramatical, apenas deixa implícito que o aluno deve ser capaz de participar de discussões e negociações em contextos sociais e empresariais, *destacando vantagens, desvantagens e necessidades*. Podemos inferir que no quesito vantagens, desvantagens e necessidades, o uso de verbos modais se fazem presente. Além disso, esse mesmo nível apresenta itens que o examinando precisa utilizar palavras de ligação. Provavelmente os examinandos estão apresentando uma dificuldade maior neste nível, pois tanto verbos modais quanto palavras de ligação não são assuntos geralmente abordados com profundidade em algumas escolas de nível médio.

Outro nível que gostaríamos de destacar é o nível 2, que apresenta discrepância em relação aos outros níveis do teste. Observamos que quatro assuntos da ementa não são contemplados em itens no teste, além de que os itens que estão relacionados à habilidade de leitura mostram-se inadequados para o propósito, uma vez que estão contemplando conhecimento gramatical, e não a habilidade de interpretar informações do texto.

Outro dado que podemos verificar através dos gráficos e tabelas apresentadas nesta seção é que no 2º semestre de 2012 os alunos ingressantes tiveram um desempenho abaixo do normal em relação aos outros semestres apresentados neste trabalho, e mesmo em relação aos outros cursos, que se mantiveram um desempenho semelhantes aos outros semestres verificados.

6.2. Retomada das perguntas de pesquisa

Ao abordar o tema avaliação de proficiência linguística em língua inglesa, e fazer uma análise de um teste de classificação procuramos demonstrar alguns encontros e desencontros em relação ao teste, voltamos-nos agora para a discussão das questões que nortearam este trabalho, isto é, sobre a validação do teste em relação à validade de construto, conteúdo e validade de face.

Em relação aos questionamentos que direcionaram este trabalho temos as seguintes respostas:

1) O teste de classificação proposto é válido para o posicionamento dos alunos dos cursos superiores em tecnologia analisados neste trabalho? (Cursos de Eventos, Logística e Gestão da Tecnologia da Informação).

A validade de um teste é um dos procedimentos mais importantes utilizados em processos avaliativos. A validade seria a investigação que se faz a partir de indagações se um teste mede aquilo para o qual foi desenvolvido para medir. Entretanto, a validade não é um critério, porém interpretações que se faz de seu uso (Messick, 1989), assim, amparadas pelas análises quantitativas que foram expostas neste trabalho, pela discussão dos itens do teste de classificação, e a partir da análise das ementas unificadas para a disciplina língua inglesa propostas para os cursos superiores em tecnologia de Eventos, Logística e Gestão da Tecnologia da Informação, podemos dizer que o teste de classificação não está adequado para o posicionamento dos alunos nos 6 níveis de língua inglesa dos cursos.

Para ratificar a afirmação da inadequação do teste, vamos utilizar alguns teóricos que fazem comentários sobre avaliação e validação. O primeiro quesito que devemos pensar para responder à questão é se o teste é válido para o posicionamento dos alunos é pensar no construto deste teste. Segundo Tumolo e Tomich, 2007: 68 apud Scaramucci, 2009: 35),

O primeiro estágio, planejamento, envolve a definição do que se quer medir, as especificações norteadoras do desenvolvimento de um teste. Faz-se, assim, necessária a definição do construto, entendido como o conjunto de habilidades e/ou conhecimentos que pode ser plausivelmente argumentado e/ou teoricamente justificado como esperado. (TUMOLO e TOMICH, 2007:68)

E que de acordo com Ebel e Frisbie (1991, p. 108), que comentam que construto “refere-se [...] a uma conceitualização teórica sobre um aspecto do comportamento humano que não pode ser medida ou observada diretamente”.

E ainda, segundo Scaramucci (2009:35) “Construto, portanto, não é universalmente compartilhado, mas definido localmente, situado, uma construção, uma

teorização que, como os próprios autores sugerem, deve ser “argumentado e teoricamente justificado”.

A autora continua sua argumentação, observando que:

O entendimento do que é um construto está na base do ato de avaliar. Vale lembrar que avaliações em geral são baseadas em inferências sobre um determinado critério, visto como o conjunto de comportamentos que se deseja avaliar. Esses comportamentos são subsequentes a um teste e, portanto, não observáveis. A única maneira de torná-los observáveis é caracterizá-los para que possam ser simulados ou representados, sempre de forma amostral, na elaboração do instrumento. Os dados de desempenho observados a partir da aplicação do teste serão usados para fazermos inferências sobre o critério, permitindo observar o que antes não era observável (McNamara, 2000). É importante, portanto, uma distinção clara entre o critério, ou seja, o comportamento comunicativo na situação alvo que se quer avaliar (leitura, por exemplo) e o teste ou instrumento para avaliá-lo. (Scaramucci, 2009:35)

Assim, a ementa dos cursos preconiza que as quatro habilidades linguísticas serão trabalhadas durante todo o curso, entretanto a ênfase será maior no componente oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos sócio-culturais da língua inglesa. Deste modo temos aí a definição de linguagem que permeará o ensino de língua inglesa do curso, que compreende a linguagem como uma ação social, usada como uma forma de ação conjunta, agindo em coordenação uma com outra, e que focaliza o ensino numa perspectiva sociointeracionista (Clark, 2007). Podemos dizer, portanto, que essa é a base para o construto do teste.

A partir desse construto temos também uma definição do que é ser proficiente em uma língua estrangeira, e portanto, a base para a construção do teste, que requer que o falante para se mostrar proficiente, além de conhecer o sistema abstrato da língua, precisa saber utilizá-lo adequadamente em determinadas situações, e como é registrado pelas ementas do curso, que saiba se comunicar utilizando as especificidades acadêmico-profissionais da área de atuação, abordando aspectos sociais.

Para podermos responder com precisão à pergunta de pesquisa acima delineada, dividiremos o teste em partes, sendo essa divisão composto de três partes: a parte escrita (que engloba leitura, com a finalidade de identificar informações gerais e específicas em um texto; o conhecimento de vocabulário e estruturas sociais da língua); a recepção oral (reconhecimento de informações aurais); e produção oral, que tem por objetivo compreender as perguntas feitas pelo examinador e responder a elas de forma coerente).

Começando pela parte de escrita, constatamos que vários itens do teste usam texto como pretexto para para testar vocabulário e estruturas gramaticais. De acordo com o construto do teste, que seguiu as ementas do curso, deveríamos ter itens que englobassem situações de comunicação que utilizem recursos linguísticos para apresentar-se, fornecendo informações pessoais e cotidianas, descrever locais e pessoas, fazer comparações, pedir e dar permissão, etc. Essas funções que estão apontadas nas ementas demandam o uso de estruturas gramaticais corretas, léxico pertinente a cada situação e podem ser testadas em vários tipos de situação como leitura, recepção e produção oral.

A parte do teste referente à recepção oral é pouco representativa, pois até o nível 3 do teste somente dois itens são utilizados para avaliação e a partir do nível 4, três itens. Se temos como afirmação nas ementas que as quatro habilidades serão trabalhadas durante o curso, porém a ênfase será na oralidade, o teste deixa a desejar uma vez que avalie superficialmente a recepção oral dos examinandos.

Observaremos que o teste não contempla a habilidade de produção escrita, apesar desse conteúdo constar nas ementas, solicitando que o aluno seja capaz de dar e anotar recados; redigir correspondências rotineiras; descrever habilidades, responsabilidades e experiências profissionais, (que pode ser feito tanto oralmente, quanto por escrito); redigir cartas e e-mails comerciais, relatórios e currículos, redigir correspondência comercial em geral e redigir textos técnicos e acadêmicos.

Em relação à produção oral, percebemos inadequações mais acentuadas, uma vez que a prova oral, representada por uma entrevista, só é efetuada a partir do momento que o examinando ultrapassar o nível 3, e em alguns casos, até mesmo o nível 2 do teste. Novamente, citando as ementas unificadas dos cursos, se a intenção é enfatizar a oralidade, devemos cobrar oralidade no teste. Segundo Bachman (1990) o uso e as interpretações que fazemos de um determinado teste podem não ser válidos para todas as habilidades e em todos contextos. Um teste que apresenta tarefas de compreensão de leitura, grande número de itens gramaticais, com uma pequena amostragem de questões de recepção oral, não pode ser considerado válido para definir habilidade de fluência oral. Da mesma forma que ao se afirmar que o curso

Muitos testes de classificação e exames de proficiência reconhecidos internacionalmente utilizam até hoje questões de múltipla escolha, que têm o propósito de testar elementos gramaticais e o uso lexical. Outras habilidades como de recepção/produção oral e produção escrita são avaliadas de forma isolada. Uma das razões para a utilização desse tipo de verificação é a praticidade e a confiabilidade da correção, principalmente com um retorno rápido da pontuação para o examinando, sem a possibilidade de possíveis reclamações por parte do examinando, uma vez que as questões escritas somente contemplam uma resposta. Entretanto, ao se elaborar um teste, outros itens se fazem presentes além da praticidade e confiabilidade. Precisamos também pensar na autenticidade e a validade desse teste.

Portanto, se nos orientarmos pelas ementas, que contribuíram para o construto do teste, podemos dizer que para que o examinando fosse posicionado de forma justa nos 6 níveis do teste, precisaríamos ter mais questões de recepção oral, produção oral a partir do primeiro nível e incluir a parte de produção escrita. Podemos inferir que para que um aluno acompanhasse o curso em um nível mais elevado, não seria suficiente este saber itens gramaticais, pois o foco das ementas é a prática da oralidade em diferentes situações de uso da língua inglesa.

2) O conteúdo do teste é condizente com a ementa e o conteúdo programático do curso?

Voltando-nos para Hughes (1994:22), um teste tem validade de conteúdo se seu conteúdo constitui uma amostra significativa das habilidades, sub-habilidades e estruturas que se pretende avaliar.

A partir da definição do construto do teste, que se baseou nas seis ementas de língua inglesa, que abordam o uso das quatro habilidades, ler, ouvir, falar e escrever, entretanto com ênfase na oralidade, que seria falar e ouvir, podemos dividir os itens que compõem o teste, e que são representações do conteúdo abordado no mesmo em defensáveis e não defensáveis.

Segundo Scaramucci (2009:35), defensáveis são aqueles “que permitem coletar evidências para a interpretação de habilidade com maior grau de validade de construto”. As ementas dos 6 semestres são compostas no total por 40 tópicos, que deveriam ser

abordados em 3 anos de curso. De acordo com os quadros apresentados no Capítulo V, detalhando as ementas e a análise dos itens do teste, temos 31 itens que foram abordados no teste e 22 que não foram.

Os itens mais representativos do teste, pois são evidências do construto do mesmo, portanto, são defensáveis estão relacionados à recepção oral, sendo os de nºs 9, 10, 19, 20, 28, 29, 30, 38, 39, 40, 48, 49, 50, 58, 59 e 60.

Os níveis 4, 5 e 6 do teste são aqueles que possuem uma amostragem reduzida do conteúdo das ementas, porém possuem itens que correspondem a tópicos abordados na ementa, só que em menor número. Como já foi observado anteriormente, os resultados do posicionamento dos examinandos nesses três níveis foi relativamente baixo em comparação com os outros níveis. Observamos que itens que fazem referência a elementos de ligação e ao uso dos verbos modais são os que mais trazem dificuldades aos examinandos.

Podemos dizer que o teste apresenta validade de conteúdo, entretanto, gostaríamos de sinalizar que alguns testes ainda seguem um padrão estruturalista de linguagem, abordando itens gramaticais. Alguns autores como Alderson (2002), Bachman (1998), após pesquisas feitas com exames de proficiência internacionais, afirmam que 70% dos itens de leitura do TOEFL estão relacionados à gramática e a conteúdos acadêmicos. Essa noção de conteúdo, que presumidamente estaria relacionada à leitura pode levar a uma noção equivocada de que há uma equivalência entre o conhecimento de estruturas gramaticais e interpretação de textos.

Uma revisão da grande maioria dos itens do teste de classificação é recomendável para que os mesmos se adaptem ao conteúdo solicitado pelas ementas, inclusive à definição de linguagem que está implícita na ementa.

3) Os examinandos identificam que o resultado do teste os está classificando adequadamente?

No que diz respeito à opinião dos alunos e professores (validade de face do teste) o resultado é favorável, uma vez que a maioria dos examinandos (52%) consideraram as tarefas adequadas e o tempo para a realização dos itens também (66%). Os examinandos consideraram os enunciados dos exercícios claros e objetivos (51%), sendo que as informações e instruções fornecidas para a realização do teste foram

suficientes em sua grande maioria (38%). A grande dificuldade encontrada pelos examinandos foi o desconhecimento de pontos gramaticais e estruturais da língua inglesa (36%), juntamente com o desconhecimento do vocabulário (23%) e do conteúdo do teste em geral (25%), o que pode ser confirmado em relação ao nível 4 do teste, cujas questões versavam sobre verbos modais e palavras de ligação, fato confirmado pela resposta ao dos mesmos, que ao realizarem o teste perceberam que ainda não tinham estudado a maioria dos conteúdos da prova (31%), ou que já estudaram, porém não aprenderam (27%). Além disso, 52% dos examinandos responderam que os tipos de exercícios eram variados deixando a prova agradável.

Uma das grandes inovações desse teste foi utilizar o computador como meio para a realização do teste, ao invés de se fazer pelo tradicional método de papel e lápis, e segundo os examinandos é um fator muito positivo, pois 38% dos alunos que responderam o questionário se sentiram confortável em fazer a prova mediada por um meio eletrônico e 32% se sentiram muito confortáveis. Somente 3% dos alunos que fizeram a prova, de um universo de 106 questionários respondidos se sentiram desconfortáveis. Essa aceitação em fazer um teste mediado pelo computador é um fator muito positivo, uma vez que os examinandos que fizeram o teste são todos ingressantes em cursos superiores de tecnologia, que têm como princípio básico o uso, a difusão, a compreensão de formas de tecnologia que possam ajudar uma sociedade em geral.

Além dos examinandos, duas professoras do curso também foram questionadas a respeito do teste e responderam que elas concordam que o teste contempla as necessidades dos alunos de cursos superiores em tecnologia, que os tópicos expostos no teste são interessantes, e que as atividades do teste abordam estilos diferentes de aprendizagem. Entretanto, alguns pontos acabam por divergir em relação às respostas dos examinandos, sendo um deles que as instruções para o teste não são claras e objetivas, que o tempo para fazer o teste não é adequado.

Sobre o conteúdo do teste, as duas professoras apresentam críticas que o conteúdo, em sua maioria, não é aquele apresentado nas ementas e que o teste está centrado em Inglês para Negócios, entretanto as ementas de cada nível do curso não afirmam que esse seja o foco dos cursos, propondo assuntos que abarquem às

especificidades acadêmico-profissionais da área, abordando aspectos sócio-culturais da língua inglesa.

4) Em que medida tem sido um norteador para o planejamento do ensino e atividades nos cursos superiores em tecnologia?

Podemos afirmar com segurança que sim, pois o teste é bem aceito pela comunidade ingressante nos cursos superiores em tecnologia, entretanto alguns alunos não se dão conta na hora de fazer o teste da importância que esse terá durante seu tempo de estudos na faculdade. Segundo McNamara (2000) toda avaliação provoca consequências diversas nos contextos educacionais em que é praticada. Sabemos que essas práticas provocam impacto e efeito retroativo. Em conversas informais com alunos que fizeram o teste em outros semestres e que agora já estão finalizando os cursos, pudemos perceber o quanto alguns alunos lastimam não terem dado a devida atenção quando foram fazer o teste ao ingressarem na faculdade, pois o que definia se eles continuariam no nível I, no nível II, ou se seriam dispensados de alguns níveis é o acerto de um único item, que muitas vezes foi feito sem a devida atenção.

Da mesma forma que os ingressantes fizeram um teste de classificação para posicioná-los em classes homogêneas, para ajudar na aprendizagem da língua, incentivamos os alunos concluintes a prestarem um exame de proficiência reconhecido internacionalmente, como o TOEIC, que tem por objetivo avaliar a proficiência linguística do candidato que está voltado ao mercado de trabalho; ou TOEFL, IELTS, que está voltado para a área acadêmica. Percebemos assim que muitas atividades utilizadas em sala de aula, e mesmo quando pensamos em reformulações do plano pedagógico do curso e conseqüentemente das ementas, do material didático a ser utilizado, os tipos de avaliação, sendo o teste de classificação é um parâmetro para mudanças.

Ao se preparar o aluno para as tarefas que constam nas ementas dos cursos, algumas tarefas que estão presentes também no teste, estamos preparando-os para usar a língua em situações reais de uso da língua, de forma que podemos concluir de uma forma bastante primário, pois algumas pesquisas mais detalhadas precisariam ser feitas, que o teste está tendo um impacto no ensino e aprendizagem desses alunos, e que talvez

já esteja tendo efeitos positivos, pois os professores que ministram as aulas de Inglês estão preocupados em trazer inovações para suas aulas.

Com isso, finalizamos este capítulo e passamos às considerações finais.

Considerações Finais

O objetivo deste trabalho foi a análise de um teste de classificação, que está sendo aplicado em cursos superiores de tecnologia, durante cinco semestres consecutivos – do 2º semestre de 2010 ao 2º semestre de 2012, que tem como característica avaliar o nível de competência linguística do aluno de forma que possa ser colocado em um estágio de um programa de ensino apropriado as suas habilidades com o propósito de validação desse teste.

Para tanto fizemos uma análise dos itens que compõem cada nível desse teste de classificação, fazendo-se uma comparação com o conteúdo programático mínimo da disciplina Inglês, que consta na ementa dessa faculdade de tecnologia, além da validação de seu construto, que se destina a conhecimentos sobre a língua inglesa e formas de como usar essa língua em situações diversas, principalmente, na área de negócios, de forma que analisamos por meio deste trabalho a validade de construto, a validade de conteúdo e a validade de face desse teste.

Pela análise das ementas unificadas dos cursos envolvidos neste estudo que são Eventos, Logística e Gestão da Tecnologia da Informação, observamos que a definição de linguagem que permeia as ementas, podemos concluir definitivamente que está baseada em Clark (2007) que compreende a linguagem como uma ação social conjunta, usada como uma forma de ação conjunta, agindo em coordenação uma com outra, e que pressupõe muito mais do que simplesmente dominar regras gramaticais ou funções comunicativas; é uma forma de interagir com o mundo.

Seguindo agora os princípios norteadores que as ementas pressupõem, temos que o foco dos estudos de Inglês nos 6 semestres do curso é que o aluno tenha as habilidades de compreensão e produção oral e escrita por meio de funções sociais e estruturas da língua, com ênfase na oralidade, atendendo às especificidades acadêmico-profissionais da área e abordando aspectos sócio-culturais da língua inglesa, sendo esses princípios considerados o construto que fundamenta o teste.

Além disso, fizemos uma análise de cada item com o conteúdo especificado nas ementas, sendo dessa forma verificada a relevância do conteúdo que está sendo

utilizado no teste. E para completar a aplicamos questionários para os alunos ingressantes do 2º semestre de 2012 como forma de verificação da atitude desses em relação ao teste, a visão do teste de duas professoras que administram a disciplina Inglês na unidade de ensino.

A validação de um teste nem sempre fornece uma resposta positiva (sim) ou negativa (não) como resposta (Chapelle, 2004). É um processo dinâmico que nunca termina, mas que vai se desenvolvendo de acordo com o desenvolvimento da ciência que trabalha com a área de avaliação.

Os resultados deste estudo oferecem evidências de que o teste apresenta aceitação pelos alunos ingressantes, apresentando baixos índices de rejeição, sendo que o maior problema que os examinandos observam no teste é o não conhecimento do conteúdo, ou se já viram o conteúdo, não aprenderam, apesar de que todos eles já tiveram, no mínimo, 7 anos de contato com a língua inglesa. Segundo os dados do questionário sociocultural que os examinandos respondem antes de iniciar o teste, a grande maioria desses alunos é oriunda de instituições públicas e não frequentam institutos de idiomas para aprimorarem seus conhecimentos em língua estrangeira.

Conforme observamos na análise dos itens que compõem o teste, o construto prevê o uso das quatro habilidades, ler, ouvir, falar e escrever, com ênfase na oralidade, porém o teste avalia, nos três primeiros níveis somente duas delas, que são ler e ouvir para em seguida, nos três últimos níveis avaliar o ler, ouvir e falar, através de uma entrevista, feita pelo próprio professor da disciplina como forma de confirmação do nível escrito. O teste não contempla nenhuma parte escrita, apesar de tarefas escritas serem solicitadas na ementa.

Apesar do foco deste trabalho não ser a validação da parte oral, percebemos que embora a ementa enfatize o uso da oralidade, essa parte acaba sendo negligenciada, uma vez que a avaliação oral só ocorrerá a partir do 3º nível do teste, os itens do teste correspondentes à recepção oral são insuficientes para poder avaliar a proficiência oral do examinando.

Em relação ainda ao teste oral, verificamos que não há uma grade de avaliação holística, nem analítica, que possam dar suporte ao examinador, e para que ele tenha parâmetros de avaliação.

Embora estejamos falando de um teste de classificação, e sabemos que esse deva avaliar com rapidez e com precisão, para que o examinando tenha seu resultado o mais rápido possível e seja posicionado numa sala, cujo ensino esteja em um nível muito próximo de seu conhecimento, e para isso foi escolhida a modalidade eletrônica e o formato de múltipla escolha, um total de 60 questões com 4 alternativas para cada questão, para que os alunos pudessem fazê-lo, em, no máximo 75 minutos, constatamos que o teste não abarca vários tópicos que estão discriminados nas ementas de cada nível de inglês, apesar de que a maior parte do conteúdo estar presente nos itens.

A análise das amostras dos itens do teste demonstra que ainda temos uma prova centrada em itens gramaticais, vocabulário, estruturas, testados através de pontos discretos, ou a utilização de textos para leitura, como pretexto para verificação de gramática.

Apesar de sabermos que no sistema educacional os movimentos são lentos, e que temos um longo caminho pela frente, nossa intenção ao analisarmos esse teste de classificação foi trazer subsídios para que pudéssemos entender melhor o processo de elaboração, desenvolvimento e validação de um instrumento avaliativo, e que os comentários feitos neste trabalho ajudem a implementar mudanças tanto no teste em si, quanto em nossas práticas avaliativas.

Gostaríamos de ressaltar a importância positiva que esse teste de classificação está trazendo para o ensino de língua inglesa nos cursos superiores em tecnologia. Destacamos a iniciativa desses professores que se preocuparam em fazer com que seus alunos pudessem aproveitar melhor as aulas, com uma menor quantidade de alunos em sala de aula, para que o ensino fosse mais homogêneo.

Ainda, sabemos que todo teste provoca consequências diversas nos contextos educacionais, e, como afirma Bachman²⁷ (1990:279) apud Scaramucci (2009:44) “testes não são desenvolvidos e usados em um tubo de ensaio psicométrico, despido de valores; eles são sempre elaborados para servir às necessidades de um sistema educacional ou da sociedade como um todo”.

Em 2010 quando iniciamos esta pesquisa, até o ponto que estamos hoje, percorremos caminhos longos e tortuosos, que implicou idas e vindas, esperanças e desesperos, soluções e muitas dúvidas, o ultrapassar obstáculos, lampejos de desistência

²⁷ “Tests are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or society at large” (Bachman, 1990, p. 279).

e recomeços inibidos. Sabemos que este trabalho é o ponto de partida para outras pesquisas que por aí virão, que se preocupem em contribuir para o ensino de línguas, que sejam dinâmicas como a linguagem o é, procurando uma interação com o mundo.

Esperamos que este estudo possa contribuir na tentativa de se preencherem algumas lacunas sobre a validação de exames, e acreditamos que este trabalho pode ser relevante não só para professores da área de língua estrangeira, mas para todos que estejam direta ou indiretamente ligados ao ensino/aprendizagem de línguas, entretanto sabemos que várias lacunas ficaram para serem preenchidas com novas pesquisas e que pesquisadores que se dedicam a estudos relacionados ao ensino/aprendizagem de línguas poderão contribuir com novos questionamentos sobre o assunto.

BIBLIOGRAFIA

- AERA/APA/NCME. S. Standards for Educacional and Psychological Testing. Washington DC: American Psychological Association, 1999.
- AFONSO, A. J. *Avaliação Educacional: Regulação e emancipação*. São Paulo: Cortez, 2000.
- ALDERSON, J.C., CHAPHAM, C. & WALL, D. *Language test construction and evaluation*. Cambridge; Cambridge University Press, 1995.
- ALMEIDA FILHO, J.C.P. & EL DASH, L.G. Compreensão de Linguagem Oral no Ensino de Língua Estrangeira. Revista Horizontes de Linguística Aplicada, LET/UNB, vol. 01, nº 01(p.19-37), Brasília, 2002.
- ANASTASI, A. Psychological testing. New York: MacMillan Company, 1961.
- ANASTASI, A. Psychological testing.. 5th ed. MacMillan Company,1986.
- ANDERSON, A., & T. LYNCH. *Listening*. Oxford: Oxford University Press, 1988.
- ARAÚJO, K. S. *A perspectiva do examinando sobre a autenticidade de avaliações em leitura em língua estrangeira*. Dissertação de Mestrado. IEL-UNICAMP, 2007.
- BACHMAN, L.F. *Fundamental Considerations in Language Testing*. Oxford: Oxford, 1990.
- BACHMAN, L. F. A habilidade comunicativa de linguagem. In: *Linguagem & Ensino*, Vol. 6, No. 1, 2003, p. 77-128.
- BACHMAN, L.F. & PALMER, A.S. *Language Testing in Practice*. Oxford: Oxford, 1996.
- BACHMAN, L.F & PALMER, A. S. *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press, 2010.
- BAILEY, K. M. *Learning about language assessment: Dilemmas, decisions, and directions*. New York: Heinle & Heinle, 1998.
- BAKER, D. *Language testing: a critical survey and practical guide*. London : Edward Arnold, 1989.
- BAKHTIN, M. *Marxismo e filosofia da linguagem*. São Paulo: Hucitec, 1992.

BENNETT, R.E., BRASWELL, J., ORANJE, A., SANDENE, B., KAPLAN, B., YAN, F. Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, 6(9), (2008). Disponível em: <<http://www.jtla.org>>. Acesso em: 12 abr. 2010.

BERNHARDT, E. B. & KAMIL, M. L. Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, 16, 2, 1995, p. 16-34.

BINET, A. New methods for the diagnosis of the intellectual level of subnormals. *L'Anné Psychologique*, 12, 1905, p. 191-244.

BINET, A. & HENRI, B. La psychologie individuelle. *Amiee Psychol.*, 2, 1899, p. 411-465.

BLAIS, J.G. & LAURIER, M.D. The dimensionality of a placement test from several and analytical perspectives. *Language Testing*, 12, 1, 1995, p.72-98.

BRADSHAW, J. Test taker's reaction to a placement test. *Language Testing* 7, 1, 1990, p. 13-30.

BROWN, H. D. *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education, 2004.

BROWN, J.D. Improving ESL placement tests using two perspectives. *TESOL Quarterly* 23, 1989, p. 65-83.

BROWN, J. D. *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents, 1996.

BRUMFIT, C. & K. JOHNSON (orgs.) 1979. *The communicative approach to language teaching*. London: Oxford University Press.

BUCK, G. *Assessing Listening*, Cambridge: Cambridge University Press, 2001.

CAMBRIDGE ESOL: RESEARCH NOTES: ISSUE 32 / MAY 2008.

CANALE, M e SWAIN, M. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, v.1(1), 1980, p. 1-47.

CANALE, M. "From Communicative Competence to Communicative Language Pedagogy". In: Richards, J. e Schmidt, R. (orgs.) *Language and Communication*. London: Longman, 1983.

CARSON, G. R. 'Education: Testing in the 21st Century'. *Focus on Spartanburg Business*. Volume 26, No. 1; Spartanburg Area Chamber of Commerce, 2003.

CELCE-MURCIA, M. The elaboration of sociolinguistic competence: implications for teacher education. *LANGUAGES AND LINGUISTICS*, 1995, Round Table. Washington: Georgetown University.

CHAPELLE, C. A. Validity in language assessment. *Annual Review of Applied Linguistics*. 1999, 19: 254-272.

CHAPELLE, C.A.; DOUGLAS, D. (2006). *Assessing language through computer technology*. Cambridge: Cambridge University Press, 2006.

CHASTAIN, K. The ACTFL proficiency guidelines: a selected sample of opinions. *ADFL Bulletin* 20, 1988, p. 47-51.

CHIZZOTTI, A. *Pesquisa qualitativa em ciências humanas e sociais*. Petrópolis: Vozes, 2006.

CHOMSKY, N. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press, 1965.

CHOI, S. W., & TINKLER, T. *Evaluating comparability of paper and computer based assessment in a K-12 setting*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA, 2002.

CLARIANA R., & WALLACE P. Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, 33 (5), 593-602, 2002.

CLARK, H.H. O uso da Linguagem. *Cadernos de Tradução do IL/UFRG*, Porto Alegre, 9, 2000, p.49-71.

CRESWELL, J. *Qualitative Inquiry and Research Design: Choosing among Five Traditions*. Thousand Oaks, CA: Sage Publications, 1998.

CRONBACH, L. J. & MEEHL, P. E. Construct Validity in Psychological tests. *Psychological Bulletin*, 52, 1955, p. 281-302.

CUMMING, J.J. & MAXWELL, G.S. Contextualising authentic assessment. In: *Assessment in education* vol. 6 (2) Abingdon, 1999, p. 177-197.

DAVIES, A. The logic of testing language for specific purposes. *Language Testing*, 18/2, 2001, p. 133-147.

DAVIES, A. & ELDER, C. (reds.) *The handbook of applied linguistics*. Victoria, Australia: Blackwell, 2004.

DOUGLAS, D. *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press, 2000.

DOUGLAS, D. *Technology and the construct of language for specific purposes*, paper presented at the 40th annual meeting of the British Association for Applied Linguistics, Edinburgh, 2007.

DUDLEY-EVANS & M. J. ST. JOHN. *Developments in English for Specific Purposes - A multidisciplinary approach*. Cambridge: Cambridge University Press, 1998.

EBEL R.L.; FRISBIE, D.A.. *Essentials of Educational Measurement*. 5^a ed., Englewood Cliffs, Prentice-Hall, 1991.

ELLIS, M. & C. JOHNSON. *Teaching Business English*. Oxford: Oxford University Press, 1994.

FERNANDES, J. R. 2011. *Educação profissional integrada à educação básica, na modalidade de educação de jovens e adultos: políticas públicas federais emergentes no Brasil e em Portugal (2005-2010)*., 2011. Disponível em:

<<http://www.anpae.org.br/simposio2011/cdrom2011/PDFs/trabalhosCompletos/comunicacoesRelatos/0247.pdf>>

Acesso em: 15 abr. 2012.

FINI, M. I. Erros e acertos na elaboração de itens para a prova do ENEM. In: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, *Exame Nacional do Ensino Médio (ENEM): fundamentação teórico-metodológica*. Brasília, DF: O Instituto, 2005, p. 101-106.

FULCHER, G. An English language placement test: Issues in reliability and validity. *Language Testing* 14, 2, 1997a, p. 113–138.

- FULCHER, G. Computerizing na English language placement test. *English Language Teaching Journal* 53, 4, 1999b, p. 289-299.
- GIL, A. C. *Como elaborar projetos de pesquisa*. São Paulo: Atlas, 1991.
- GIL, A. C. *Métodos e Técnicas de pesquisa social*. São Paulo: Atlas, 1995.
- GUILFORD, J.P. New Standards for test evaluation. *Educational and Psychological Measurement*, 6, 1946, p. 427-439.
- GRADDOL, D. *The future of English? A guide to forecasting the popularity of the English language in the 21st century*. London: British Council. Available for free from the website of the British Council, 1997.
- GRADDOL, D. *English Next*. The British Council, 2006.
- HALADYNA, T. M. *Developing and validating multiple-choice test items*. Lawrence Erlbaum Associates, Publishers, UK, 1994.
- HALADYNA, T. M., & DOWNING, S. M. A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1), 1989a, p. 37-50.
- HAMP-LYONS, L. Washback, impact, and validity: ethical concerns. *Language Testing* 14(3), 1997, p. 295-303.
- HENNING, G. *A guide to language testing*. Rowley, MA: Newbury House, 1987.
- HOFFMAN, J. *Avaliação mediadora: uma prática em construção – da pré-escola à universidade*. Porto Alegre, Educação e Realidade, 1993.
- HOFFMANN, J. M. L. *Contos e contrapontos: do pensar ao agir em avaliação*. Porto Alegre: Mediação, 1998.
- HOLTZMAN W. H. *Computer-assisted instruction testing and guidance*. Harper Row, 1970.
- HOWE, K. & EISENHART, M. Standards for qualitative (and quantitative) research: A prolegomenon. *Educational Researcher*, 19 (4), 1990, p. 2-9.
- HUGHES, A. *Testing for language teachers*. Cambridge: Cambridge University Press, 1989.
- HYMES, D. "On communicative competence." In: PRIDE, J. B.; HOLMES, J. (eds.) *Sociolinguistics: selected readings*. Harmondsworth: Penguin Books, 1972.

HYMES, D. H. "On communicative competence". In: WILKINS, D.A. *Notional Syllabuses*. OUP, 1976, p. 5-26.

JENKINS, J.G. Validity for what? *Journal of Consulting Psychology*, 10, 1946, p. 93-98.

KANE, M. 2006. Validation. In R.L. Brennan (Ed.) *Educational measurement*. 4th Edition. Washington, DC: American Council on Education/Praeger, 2006, p. 17-64.

KIM, D.-H., & HUYNH, H. (2007). Comparability of computer and paper-and-pencil versions of Algebra and Biology assessments. *Journal of Technology, Learning, and Assessment*, 6(4). Disponível em:

< <http://www.jtla.org>>. Acesso em: 10 abr. 2010.

LADO, R. *Language testing: the construction and use of foreign language tests*. New York: McGraw-Hill Book Company, 1961.

LEESON H. V. The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, 6 (1), 1-24, 2006.

LLURDA, E. 2000. On competence, proficiency and communicative language ability. In: *International Journal of Applied Linguistics*, vol 10, n^o 1, p.85-95.

LUCKESI, C.C. *Avaliação da aprendizagem escolar*. São Paulo, Cortez Editora, 1996.

MARCUSCHI, L. A. "Gêneros textuais: definição e funcionalidade". In: DIONÍSIO, A. P.; MACHADO, A. R.; BEZERRA, M. A. (Org.). *Gêneros textuais e ensino*. Rio de Janeiro: Lucerna, 2003.

McNAMARA, T. F. *Measuring Second Language Performance*. Longman: London, New York, 1996.

MARTINS, T. H. B. *Subsídios para a Elaboração de um Exame de proficiência para Professores de Inglês*. Dissertação de Mestrado. Campinas: UNICAMP, 2005.

MAZZEO, J., & HARVEY, A. L. *The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature*. (College Board Report 88-8). New York: College Entrance Examination Board, 1988.

MESSICK, S. The standard problem: meaning and values in measurement and evaluation. *American Psychologist*, 30, 1075, p. 955-966.

MESSICK, S. Validity and washback in language testing. *Language Testing*. 13(3), 1996, p. 241-256.

MESSICK. S. "Validity". In R. L. Linn (Ed.), *Educational measurement* (3th edition), New York: American Council on Education & Macmillan, 1989, p. 13-103.

MINISTÉRIO DA EDUCAÇÃO. Censo da Educação Superior 2010. Disponível em: <www.portal.mec.gov.br>. Acesso em: 12 jan. 2012.

MODER, C.L.; HALLECK, G.B. "Framing the language proficiency interview as a speech event: native and nonnative speakers' questions". In R. Young e A .W. He (orgs) *Talking and testing: discourse approaches to the assessment of oral proficiency*. Philadelphia: John Benjamins, 1998, p .117 -48.

MORROW, K.E. "Communicative language testing, revolution or evolution?" In: Brumfit, C. &K. Johnson (orgs.) *The communicative approach to language teaching*. London: Oxford University Press, 1979.

MIYATA, R. A Review of TOEIC: A Critical Analysis of the Structure, Methods, and Purpose of TOEIC, Including its Effect on Teaching and Learning English as a Second or Foreign Language". *Bulletin of Beppu University Junior College* 23, pp. 59-69, 2004.

NEVO, B. Face Validity Revisited. *Journal of Educational Measurement* 22 (4), 1985, p. 287-293.

PAEK, P. *Recent trends in comparability studies* (PEM Research Report 05-05), 2005.

Disponível em:

<http://www.pearsonedmeasurement.com/downloads/research/RR_05_05.pdf> . Acesso em: 12 jan. 2012.

PERKINS, K., GUPTA, L. & TAMMANA,R. Predicting item difficulty in a reading comprehension test with an artificial neural network. *Language Testing*, 12(1), 1995, p.34-53.

POMPLUN M., RITCHIE, T., & CUSTER M. Factors in paper-and-pencil and computer reading score differences at the primary grades. *Educational Assessment*, 11(2), 127-143, 2006.

POMPLUN, M., & CUSTER, M. The score comparability of computerized and paper-and-pencil formats for K-3 reading tests. *Journal of Educational Computing Research*, 32(2), 153-166, 2005.

POMPLUN, M., FREY, S., & BECKER, D.F. The score equivalence of paper and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, 62(2), 337-354, 2002.

RAYMUNDO, V. P. Construção e validação de instrumentos: um desafio para a psicolinguística. *Letras de Hoje*. Porto Alegre, v. 44, n.3, p. 86-93, jul./set. 2009.

RETORTA, M.S. Efeito retroativo do vestibular da Universidade Federal do Paraná no ensino de língua inglesa em nível médio no Paraná: uma investigação em escolas públicas, particulares e cursos pré-vestibulares. *Tese de Doutorado*. Campinas, UNICAMP, 2007.

ROSENBERG, M. 2004. “Never the twain shall meet”, *English Teaching Professional* 11(35), 2004, p. 36–37.

ROBINSON, P. *ESP today: a practitioner’s guide*. Hertfordshire: Prentice Hall International, 1991.

ROST, M. *Listening in language learning*. New York: Longman, 1990.

SALOMAO, A. C. B.. Fatores a serem levados em consideração para o desenvolvimento de testes de proficiência oral em contexto virtual. *Trabalhos em Linguística Aplicada*, Campinas, v. 49, n. 2, Dec. 2010.

SANTOS, L.G. Avaliação de desempenho para nivelamento de alunos de português como língua estrangeira. *Dissertação de Mestrado*. Porto Alegre, UFRGS, 2007.

SCARAMUCCI, M. V. R. Avaliação: mecanismo propulsor de mudanças no ensino/aprendizagem de língua estrangeira. In: *Contexturas*, vol.4, 1999/2000, p. 115-124.

SCARAMUCCI, M. V. R. “Proficiência em LE: considerações terminológicas e conceituais”. In: *Trabalhos em Linguística Aplicada*, Campinas, 36, 2000, p. 11 – 22.

SCARAMUCCI, M. V. R. “O professor avaliador: sobre a importância da avaliação na formação do professor de língua estrangeira”. In: ROTTAVA, L.; SANTOS, S. S. (Orgs.) *Ensino e aprendizagem de línguas: língua estrangeira*. Editora da UNIJUI, 2006, p. 49-64.

SCARAMUCCI, M.V.R. “Efeito retroativo da avaliação no ensino/aprendizagem de línguas: o estado da arte”. *Trabalhos em Linguística Aplicada*, Campinas, 43 (2),2004, p. 203-226.

SCHOFFEN, J. R.. *Avaliação de proficiência oral em língua estrangeira: descrição dos níveis de candidatos falantes de espanhol no exame Celpe-Bras*. Dissertação de mestrado. PPG-Letras, UFRGS, 2003.

SCHALATTER, M.; GARCEZ, P.de M.; SCARAMUCCI, M.V.R. O papel da interação na pesquisa sobre aquisição e uso de língua estrangeira: implicações para o ensino e para a avaliação. *Letras de Hoje* 39 (3), 2004, p. 345-378.

SHAW, S.D. Legibility and the rating of second language writing: the effect on examiners when assessing handwritten and word-processed scripts, *Research Notes* 11, 7-10, 2003.

SHEPARD, L.A. Evaluating test validity. In: Darling-Hammond (Ed.), *Review of research in education* Vol. 19, 1993, p.405-450. Washington, DC: American Educational Research Association.

SHERMAN, J. The effect of question preview in listening comprehension tests. *Language Testing*, 14(2), 1997, p. 185–213.

SHOHAMY, E. “A proposed framework for testing the oral language of second/foreign language learners”. *Studies in Second Language Acquisition*, 10. Cambridge University Press, 1988.

SHOHAMY, E. “Beyond Proficiency Testing: A Diagnostic Feedback Testing Model for Assessing Foreign Language Learning”. *Modern Language Journal*, 76, 1992, p. 513-21.

SHOHAMY, E. “Performance Assessment in Language Testing”. *Annual Review of Applied Linguistics*, 1995, p. 188-211.

SHOHAMY, E. *The Power of Tests: A critical perspective on the uses of language tests*.” London: Longman, 2001.

SOUZA, S.M.Z. “Possíveis impactos das políticas de avaliação no currículo escolar”. *Cadernos de Pesquisa* no.119, 1993, p.175-190.

SPEARMAN, C. General Intelligence: objectively determined and measured. *American Journal of Psychology*, 18, 1904, p. 161-169.

SPOLSKY, B. *Language testing: art or science?* Paper presented at the 4th AILA International Congress, Stuttgart, 1975.

SPOLSKY, B., "Language Testing: Art or Science", In Nickel, G. (ed.), *Language Testing*, Stuttgart: Hoschulverlag, 1978, p. 216-225.

STAKE, R. E. *The art of case study research*. Thousand Oaks, CA.: Sage, 1995.

STAKE, R. E. *Program evaluation, particularly responsive evaluation*. Champaign, IL: Center for Instructional Research and Curriculum Evaluation (CIRCE), 1973. Keynote presentation at a conference on "New Trends in Evaluation". Disponível em: <http://www.ed.uiuc.edu/circe/Publications/Responsive_Eval.pdf>. Acesso em: 18 ago. 2012.

STERN, H.H. Models of second language learning and the concept of proficiency. In: *Fundamental concepts of language teaching*. Oxford University Press, 1983.

STRAUB, D.W. Validating Instruments in MIS Research. *MIS Quarterly* (13:2), 1989, p. 146-169.

STREVENS, P. ESP after twenty years: a re-appraisal. In: *ESP: state of the art*. M. L. TICKOO (ed.) Anthology Series 21. SEAMEO Regional Language Center, 1988.

TAYLOR, D.S. The meaning and use of the term 'competence'. in *Linguistics and Applied Linguistics*. In: *Applied Linguistics*, 9 (2), 1988, p.148-168.

VASCONCELOS, C. S. *Avaliação: concepção dialética-libertadora do processo de avaliação escolar*. 11 ed. São Paulo: Libertad, 2000.

VIANNA, H. M. *Avaliação educacional: teoria, planejamento, modelos*. São Paulo: Ibrasa, 2000.

VYGOSTSKY, L.S. *Pensamento e Linguagem*. São Paulo: Martins Fontes, 1993.

WAINER, H. Epilogue. IN: Wainer, H. (Ed). *Computerized adaptive testing: A primer*. 2nd. Edition. Hillsdale, NJ: Lawrence Erlbaum, 2000.

WAY, W. D., DAVIS, L. L., & FITZPATRICK, S. *Score comparability of online and paper administrations of Texas assessment of knowledge and skills*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. San Francisco, CA, 2006.

WALL, D., CLAPHAM, C. AND ALDERSON, J.C. *Evaluating a placement test*. *Language Testing* 11, p. 321-344, 1994.

WEIR, C. J. *Communicative language testing*. New Jersey: Prentice Hall International, 1990.

WEIR, C. J. A survey of the history of the Certificate of Proficiency in English (CPE) in the twentieth century, in Weir, C J and Milanovic, M (Eds), *Continuity and Innovation: Revising the Cambridge Proficiency in English examination 1913–2002*, Studies in Language Testing volume 15, Cambridge: UCLES/Cambridge University Press, 2003, p. 1–56.

WEIR, C. J. *Language Testing and Validation: An evidence-based approach*. Palgrave-Macmillan, 2005.

WIDDOWSON, H.G. *Teaching language as communication*. Oxford: Oxford University Press, 1978.

WILKINS, D.A. *Notional Syllabuses*. OUP, 1976, p. 5-26.

YIN, Robert K. *Estudo de Caso, planejamento e métodos*. 4.ed. São Paulo: Bookman, 2010.

ZHU, W. On differences between General English Teaching and Business English Teaching. *English Language Teaching*, 2008.

APÊNDICE 1**Questionário de percepção da prova
Placement Test - Agosto 2012**

- 1) Qual o grau de dificuldade deste teste na parte de leitura?
a) Muito fácil b) Fácil c) Médio d) Difícil
e) Muito difícil
- 2) Qual o grau de dificuldade deste teste na parte de “listening”?
a) Muito fácil b) Fácil c) Médio d) Difícil
e) Muito difícil
- 3) Considerando a extensão do teste, em relação ao tempo total, você considera que o teste foi
a) Muito longa b) Longa c) Adequada d) Curta
e) Muito curta
- 4) Os enunciados dos itens do teste estavam claros e objetivos?
a) Sim, todos. b) Sim, a maioria. c) Apenas cerca da metade.
d) Poucos. e) Não, nenhum.
- 5) As informações/instruções fornecidas para a resolução dos itens foram suficientes para resolvê-los?
a) Sim, até excessivas.
b) Sim, em todas elas.
c) Sim, na maioria delas.
d) Sim, somente em algumas.
e) Não, em nenhuma delas.
- 6) Você se deparou com alguma dificuldade ao responder o teste. Qual?
a) Desconhecimento do conteúdo.
b) Falta de motivação para fazer a prova.
c) Desconhecimento do vocabulário.
d) Desconhecimento gramatical/estruturas.
e) Não tive qualquer tipo de dificuldade para responder à prova.
- 7) Considerando os itens do teste, você percebeu que
a) não estudou ainda a maioria do conteúdo da prova.
b) estudou alguns desses conteúdos, mas não os aprendeu.
c) estudou a maioria desses conteúdos, mas não os aprendeu.
d) estudou e aprendeu muitos desses conteúdos.
e) estudou e aprendeu todos esses conteúdos.
- 8) Em que nível você chegou ao fazer o teste?
a) 1 b) 2 c) 3 d) 4 e) 5 f) 6
- 9) Eu me senti confortável em fazer o teste no computador?

- a) Concordo totalmente
- b) Concordo
- c) Não concordo nem discordo – neutro
- d) Discordo
- e) Discordo totalmente

10) Há uma boa variedade de tipos de exercícios, deixando o teste agradável para se fazer.

- a) Concordo totalmente
- b) Concordo
- c) Não concordo nem discordo – neutro
- d) Discordo
- e) Discordo totalmente

APÊNDICE 2

Questionário III - Aplicado aos professores da unidade de ensino

O questionário II é constituído de 30 itens, combinando perguntas de respostas fechadas e abertas. Seu objetivo é coletar dados que permitam delinear percepções e atitudes com relação ao teste de inglês entre os profissionais envolvidos com a disciplina de Inglês nessa unidade de ensino, para se estabelecer uma validade de conteúdo.

Experts Post-Test Questionnaire

This questionnaire is part of the validation process of the Placement Test (henceforth – PT) designed for the Curso Superior em Tecnologia – (henceforth CST). The aim of PT is to improve the academic quality and efficiency of the CST in order to establish its content validity. The questionnaire is entirely anonymous. (Baseado em Papadyma, 2005:208).

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. PT meets the requirements of an English placement test for the needs of CST students.					
2. PT accurately represents the content of CST English Language Programme at the various levels.					
3. The grammatical points represent the structure content of the respective levels.					
4. The lexical points represent the vocabulary content of the respective levels.					
5. The reading comprehension items represent the level of understanding of reading texts of the respective levels.					
6. The topics are appropriate for the respective levels.					
7. The topics are interesting.					
8. The instructions are clear.					
9. The test reflects settings test-takers would come across in their studies.					
10. The test reflects settings and contexts in English speaking settings.					
11. The test reflects the setting where test-takers will be studying.					
12. There is sufficient variety in test types, appropriate for the respective levels.					

13. PT items generally target the appropriate English level.					
14. The kinds of activities are appropriate.					
15. The activity types cover variety of learning styles.					
16. The time given to the test-takers to take the test is appropriate.					
17. There is time flexibility to cater for test-takers' difference in time needs.					
18. Item selection criteria are appropriate.					
19. The scoring system is satisfactory.					
20. The test results are appropriately reported to the test-takers.					
21. The instructions help test-takers familiarize themselves with the test, the test format, the test layout and the test items.					
22. The computer screen is clear.					
23. The activity layout is clear.					
24. The visuals are appropriate.					
25. The students are given opportunities to control the test (Review items, going backwards and towards, etc.).					

26. Tell what you did not like about the PT?

27. Tell what you liked most about PT (if anything)?

28. Tell how PT can be improved?

29. Do you have problems with students that aren't placed correctly?

30. Do you have problems with students that don't have to attend classes for one or several terms and they have to go back later? Which are the problems?

ANEXO 1
TESTE DE CLASSIFICAÇÃO

Questões 1 a 60

Em cada questão há somente uma resposta certa.

1) Escolha a melhor alternativa para completar o diálogo entre um professor e seus alunos, no primeiro dia de aula.

Teacher - GOOD MORNING, CLASS. NICE TO MEET YOU. I'M YOUR ENGLISH TEACHER. AND YOU, WHAT DO YOU DO?

Student A – I AM A FULL-TIME STUDENT.

Student B - I _____ WITH INFORMATION TECHNOLOGY.

Teacher – WHAT ABOUT YOU? ...

- a. not work
- b. are work
- c. work not
- d. work

2) Assinale a melhor alternativa para completar a frase abaixo.

Currently _____ 46,000 students at the 49 FATECs in São Paulo State.

- a. there was
- b. are
- c. there is
- d. there are

3) Escolha a melhor alternativa para completar a frase abaixo.

Felipe and Mariana _____ classmates. They both study Computer Database.

- a. are
- b. is
- c. was
- d. am

4) Situação: O coordenador de uma FATEC está apresentando o professor de Inglês aos alunos.

Como os alunos o cumprimentariam?

Coordinator – This is Antonio, your new English teacher.

Students - _____.

- a. See you soon.
- b. Good to see you again.
- c. It's nice to meet you.
- d. Very good.

5) Escolha a resposta mais apropriada para a seguinte pergunta.

What do you do in your free time?

- a. I watched TV.
- b. I travel on business.
- c. Sometimes I read magazines.

d. I always work late.

6) Em uma entrevista de emprego, qual seria a melhor resposta para a pergunta abaixo.

- Do you have any experience in finance?

- a. Yes, I do. I can answer the phone.
- b. Unfortunately I don't.
- c. No, I am sorry but I can't.
- d. No, I'm not.

7) Qual seria a pergunta para a seguinte frase?

Eduardo works with International Relations.

- a. What does Eduardo do?
- b. When does Eduardo work with International Relations?
- c. Does Eduardo work overtime?
- d. Why does Eduardo work?

8) Como você diria que sabe falar inglês?

- a. I can speak English.
- b. I speaking English.
- c. I am speak English.
- d. I can talk English.

9) As questões seguintes referem-se a uma gravação.

Áudio: audio1.wma

THE GIRL IS FREE AFTER _____.

- a. 9:00
- b. 8:30
- c. 5:00
- d. 12:30

10) Ainda em relação ao áudio acima:

Áudio: audio1.wma

In the end, the boy and girl decide to _____.

- a. attend a concert.
- b. attend a show.
- c. watch a movie.
- d. stay in the park near school.

11) Que frase completaria o e-mail abaixo?

Glenn, I called Rodrigo about the meeting.

He can't make it on Thursday, but Friday morning is fine for him.

Regards,

Simone

- a. The meeting is canceled.
- b. Can I confirm it?
- c. Rodrigo can't meet on Friday.

d. The meeting is on Thursday.

12) Como você diria a seu professor, de uma maneira educada, que quer saber suas notas?

- a. Give me my grades.
- b. I'd like to check my grades.
- c. I need my grades.
- d. You have to give me my grades.

13) Leia as informações abaixo e assinale a alternativa correta.

An Indian Culture Overview	A Chinese Culture Overview
Population – 1,189,172,906 (July 2011 est.)	Population – 1,336,718,015 (July 2011 est.)
Area - 3,287,263 sq km	Area - 9,596,961 sq km
Official Languages – English, Hindi and 14 other official languages (2001 census)	Official Languages – 7 main Chinese languages including standard Chinese or Mandarin, Cantonese, and Wu.
Life expectancy at birth - 66.8 years	Life expectancy at birth - 74.68 years
Internet users - 61.338 million (2009)	Internet users - 389 million (2009)
Exports – commodities: petroleum products, precious stones, machinery, iron and steel, chemicals, vehicles, apparel	Exports – commodities: electrical and other machinery, including data processing equipment, apparel, textiles, iron and steel, optical and medical equipment
Agricultural products - rice, wheat, oilseed, cotton, jute, tea, sugarcane, lentils, onions, potatoes; dairy products, sheep, goats, poultry; fish	Agricultural products - rice, wheat, potatoes, corn, peanuts, tea, millet, barley, apples, cotton, oilseed; pork; fish

(Source: The World Factbook)

- a. Life expectancy in China is higher than in India.
- b. China is less populous than India.
- c. India is bigger than China.
- d. China is smaller than India.

14) Das afirmações abaixo, assinale a única INCORRETA em relação às informações dos países acima.

- a. Both countries export steel and iron.
- b. India doesn't export any beef.
- c. There are more users of the Internet in China than in India.
- d. China exports fruit.

15) Sandro's dad is a director now. Ten years ago _____

- a. he was a manager.
- b. he will be a manager.

- c. he is going to be a manager.
- d. he is a manager.

16) Leia o texto abaixo e responda qual afirmativa sobre a vida de Denise é verdadeira. Denise was born in Sorocaba, moved to Indaiatuba when she was two years old. She studied

Business Management at FATEC. She graduated in 2007 and then worked for a multinational logistics

company for 2 years. In 2009 she went to Canada to study English.

- a. She worked in a multinational after she traveled abroad.
- b. She studied French before she graduated.
- c. She has never traveled abroad.
- d. She traveled to Canada two years after she graduated.

17) Assinale a pergunta para a resposta abaixo.

_____?
Denise studied Business Management at FATEC from 2005 to 2007.

- a. What does Denise study?
- b. Where does Denise study?
- c. Where did Denise study?
- d. Why did Denise study Business Management?

18) Situação: Sofia, a nova estagiária da FATEC, conversa em casa sobre as suas funções

rotineiras no trabalho. Leia o texto abaixo e escolha a alternativa que continue o diálogo.

- Mom, my new job is great.
- Really! What do you have to do there?

- _____

- And how do you like it?
- It's fun.
- a. I park and wash the cars.
- b. I help the teachers and coordinators.
- c. I teach the students and prepare lessons.
- d. I plan lessons and correct tests.

19) As questões de 19 a 20 referem-se a uma gravação.

Áudio: audio2.wma

Qual é o assunto da conversa telefônica?

- a. Uma negociação de compra e venda.
- b. Uma anotação de recado.
- c. Um pedido para deixar o endereço eletrônico.
- d. Uma gravação que foi deixada na secretária eletrônica.

20) Qual é o número de telefone de Andrea Hann?

Áudio: audio2.wma

- a. 619-555-2437
- b. 609 -555-2427
- c. 609- 555-3437

d. 609 –555–2477

21) You are in a meeting with a client and you need a repetition about an explanation.

What would

you say to the client?

- Repeat the explanation.
- Could you repeat the explanation?
- Do you repeat the explanation, please?
- I want you to repeat the explanation.

22) After 2 hours studying English with a friend, you are exhausted. How would you

suggest that you

need a break?

- We must have a break now.
- I can have a break now.
- We should have a break now.
- Stop now.

23) Choose the best option to fill the gap:

Hi, Pedro.

Just a short note to let you know that my printer is not working.

_____ by email.

Could you please print it for me?

Thanks, Rosa.

- I am sending the report
- I would send you the report
- I had sent the report
- I was sending the report

24) Read the text below and choose the best alternative.

All about Alice

Alice is an innovative 3D programming environment that makes it easy to create an animation for telling a story, playing an interactive game, or a video to share on the web.

Alice is a teaching tool for introductory computing. It uses 3D graphics and a drag-and-drop

interface to facilitate a more engaging, less frustrating first programming experience.

(from <http://www.alice.org/>)

Alice is a 3D programming environment created to

- teach programmers how to make 3D graphics.
- teach people how to use computers in an easy way.
- teach people how to program experiences.
- teach advanced students how to work with 3D graphics.

25) Choose the best option to complete the instructions about a tour to the museum.

The bus will leave at 9:00 and get to the museum by 11:00. We plan to return at 4:00 in the

afternoon and be back here at 6:00. _____ or enough money to get something to

eat.

- a. You mustn't bring a snack
- b. You might bring a snack
- c. You can't bring a snack
- d. You should bring a snack

26) Read the paragraph below and answer questions 26 and 27.

Choose the best alternative to complete the text.

When conducting a study on computers, whether you are working on the data for a class or

for market research, you may wish to find out the number of personal computers per household. As computers have become more affordable, their use _____ more widespread. Determining the number of PCs per household can help you understand the current state of the consumer market.

(from http://www.ehow.com/how_7602206_determine-number-pcs-per-household.html)

- a. had become
- b. has becoming
- c. became
- d. has become

27) "Determining the number of PCs per household can help you understand the current state of the

consumer market" (lines 4, 5 and 6) means

- a. The current state of the consumer market can be determined by the number of PCs per household.
- b. Families buy PCs to help the consumer market.
- c. You can determine how many families use PCs in their houses with the help of the current consumer market.
- d. The number of PCs per household is determined by the help of the current state of the market consumer.

28) As questões de 28 a 30 referem-se a uma gravação.

Áudio: audio3.wma

Choose the correct alternative.

- a. Mr. James wants to go to Taipei tomorrow.
- b. Mr. James wants to leave a message.
- c. Mr. James is calling for a new internet connection.
- d. Mr. James wants to rent a car for 5 days.

29) When will James leave the hotel?

Áudio: audio3.wma

- a. June 30
- b. June 28
- c. June 24
- d. June 22

30) The dialogue is probably between:

Áudio: audio3.wma

- a. A receptionist and a guest.
- b. A husband and a wife.
- c. A boss and a secretary.
- d. A saleswoman and a client.

31) Your boss is travelling on business and is on the phone now. She needs some information that is in the report of the last meeting. What does she say to you?

- a. Could you send me the report?
- b. Must you send me the report?
- c. Would you like me to send the report?
- d. Should you send me the report?

32) Choose the best alternative to complete the sentence.

I think that smoking cigarettes _____ illegal because it can hurt people around you.

- a. isn't allowed to be
- b. should be
- c. should to be
- d. have to be

33) Choose the best option to complete the sentence.

_____ they were extremelly tired, they went to school.

- a. On the other hand
- b. However
- c. Although
- d. But

34) Read the policy below on the use of calculators in university examinations and then choose the only sentence which is true:

i) No programmes or information stored in any medium may be brought into the examination room.

ii) All calculators should be reset prior to entering the examination room.

iii) No calculator may be used which would disturb another candidate.

- a. Students shouldn't use calculators in university examinations.
- b. Students can't use calculators in university examinations.
- c. Students must use programmes stored in the calculators during their examinations.
- d. Students can use calculators in the classroom if they follow certain rules.

35) Complete the statement of a negotiation below with a suitable ending.

This deal is very important for the company _____

- a. then we shouldn't consider it.
- b. so let's analyze it carefully.
- c. then we shouldn't worry about it.
- d. so let's postpone it.

36) Which sentence best completes the text?

I don't know if I like the idea of online shopping.

_____ **Can I
return the item and get my money back?**

- a. Why don't we shop from home?
- b. What if I like the product?
- c. Why don't we buy the item on the internet?
- d. What if the item I receive is damaged?

37) What's the idea of the text above?

- a. The safety of shopping online.
- b. Doubts about shopping online.
- c. The price for shopping online.
- d. Advantages of shopping online.

38) As questões de 38 a 40 referem-se a uma gravação.

Áudio: audio4.wma

What is the problem mentioned by Rick?

- a. He has financial problems.
- b. The computer is not working.
- c. Problems with the printer.
- d. Internet problems.

39) Louise has offered some suggestions related to:

Áudio: audio4.wma

- a. Printer models
- b. Computer programs
- c. Computer store addresses
- d. Prices

40) The dialogue is probably between:

Áudio: audio4.wma

- a. Salesman and buyer
- b. Mother and son
- c. Two friends
- d. Two salesmen

41) Choose the best alternative to complete the sentence:

In response to growing demand for energy in recent years, Brazil

_____ **and exporting ethanol.**

www.gov.uz

- a. stopped producing
- b. is decreasing production
- c. is planning to produce
- d. started producing

42) You are discussing a new business opportunity and your boss says:
 “This negotiation is too important to lose, _____”
 Choose the best alternative to complete the sentence.

- because we shouldn't make a mistake.
- as long as we don't make a mistake.
- however, we shouldn't make a mistake.
- so we shouldn't make a mistake.

43) Choose the best alternative to complete the sentence.
**Trade between Africa and Brazil _____ by 415% since 2002 when
 President Lula**

_____ the Presidency of Brazil.

<http://europafrika.net/2009/06/19>

- has increased – assumed
- increased – has assumed
- is increasing – assumes
- will increase – has assumed

44) _____ the presidents of Brazil and the US came from different sides of the
 political
 spectrum, they had a good relationship.

Choose the best alternative to complete the sentence:

- As long as
- Although
- As
- Because

45) Choose the best alternative to complete the text.

Nano Technology has tremendous potential and a lot of young people take a keen
 interest in

this field. In India, _____ the economic slowdown, science & technology has done
 well and

has shown steady growth.

adapted from www.nanotech-now.com/news

- however
- in spite of
- according to
- as a result of

46) Choose the best alternative to complete the sentence.

It was almost eleven p.m. and I _____, but I couldn't go home because I had to
 finish the
 report.

- have been tired
- am tired
- am going to be tired
- was getting tired

- 47) You work in the human resources department of a large corporation. You are writing a report explaining the decision for choosing one particular candidate. Which sentence *would not* be appropriate to start the document:
- Because the position required computer skills we decided to choose Mr. Silva.
 - How the position required computer skills we decided to choose Mr. Silva .
 - As the position required computer skills, we decided to choose Mr. Silva.
 - As long as the position required computer skills we decided to choose Mr. Silva .

- 48) Listen to the extract of a meeting in a company that produces frozen food and answer questions
48 - 50:

Áudio: audio5.wma

The aim of the meeting is to...

- discuss the results of a recent market research.
- talk about healthy food and its trends.
- decide how to improve the line of products.
- discuss what kind of new products the company should develop.

- 49) According to Donna's market research:

Áudio: audio5.wma

- Consumers don't have much interest in healthy desserts.
- Consumers have a great interest in buying healthy food.
- Less than 50% of consumers buy healthy food some or all of the time.
- Consumers buy more vegetarian food than healthy food with meat or fish.

- 50) What are the consumer's options in order of preference, according to Donna's research?

Áudio: audio5.wma

- fat free meat and fish, sugar free desserts, fruit.
- low fat meat and fish, fruit and low sugar desserts, vegetarian dishes.
- healthy food with little salt, diet desserts and fruit, vegetarian dishes.
- healthy food without salt, fruit, vegetables.

- 51) Which statement *is false* according to the table?

Table 1. Brazilian Ethanol Production

Year*	Million Gallons
2003/04	3,910
2004/05	4,068
2005/06	4,174
2006/07	4,719
2007/08	5,916
2008/09**	7,054

* marketing year = May – April

** forecast

Source: GAIN report BR8013, USDA Foreign Agricultural Service, 2008.

- a. Ethanol production is expected to grow.
- b. Ethanol production has increased significantly in the last six years.
- c. Ethanol production has expanded in recent years.
- d. Ethanol production dropped in 2007.

52) Choose the best sentence to complete the gap.

Effective meetings that produce results begin with meeting planning. First, identify if other employees are needed to help you plan the meeting, then decide

_____ .
Source: humanresources.about.com

- a. if you are going to organize it or not
- b. what the actual results will be
- c. if you can make all the decisions by yourself
- d. what you hope to accomplish by holding the meeting

53) Choose the best alternative:

- a. Brazilian exports fell sharply between 2003 and 2006.
- b. Brazilian exports doubled between 2002 and 2004.
- c. Brazilian exports have increased continuously since 2002.
- d. Brazilian exports were steady in the period of 2002 and 2006.

54) According to the text below:

About a third of the fuel Brazilians use in their vehicles is ethanol, known in Brazil as 'alcohol.'

That compares with three percent in the United States. All gasoline sold in Brazil contains at least 26% ethanol, but motorists driving flexible-fuel cars have the option of filling up with pure ethanol, or E100, which currently is selling for about half the price of the blend.

The trend in Brazil has far-reaching implications for environmental policy, trade and economic development in poor countries that may have a bright future producing crops that can be easily turned into fuels.

<http://www.researchandmarkets.com/report>

- a. Poor countries have environmental problems with ethanol.
- b. In Brazil it is possible to choose from pure gasoline and pure ethanol.
- c. Brazil produces 26% more ethanol than gasoline.
- d. Many more cars use ethanol in Brazil than in the USA.

55) Choose the best alternative for a suitable answer:

You are the manager of a supermarket. You are angry with your supplier because they didn't deliver the products you bought on time. The supplier tries to calm you down.

You - We would like to cancel the order.

Supplier- What if we offer you a special discount?

You - If you offered me a 50% discount _____.

- a. I want to reconsider my decision.
- b. I reconsider my decision.
- c. I would reconsider my decision.
- d. I will reconsider my decision.

56) Choose the best alternative to report Obama's opinion about Brazil:

Obama: "Brazil is an economic power and a key player in the international scenario".

<http://www.rio2016.org.br/en/Noticias>

- a. Obama asked if Brazil was an economic power.
- b. Obama spoke that Brazil was an economic power.
- c. Obama said that Brazil was an economic power .
- d. Obama told Brazil is an economic power.

57) According to the text below, if we work from home _____

Advantages and disadvantages of working from home

Working from home can have many advantages. You may find it helps you to:

- arrange work around family commitments;
- get help from family, e.g. they may help you with filing or general administration tasks;
- save on time and money spent traveling to work;
- avoid noise and distractions of the workplace;

However, there are disadvantages from home working too. You may find you have trouble with:

- separating home and work life;
- the initial expense of setting up;
- domestic distractions and interruptions;
- isolation/loneliness.

<http://www.businesslink.gov.uk>

- a. we can make it easier to maintain team work.
- b. we can avoid distractions and interruptions.
- c. it's impossible to keep home and work life apart.
- d. we can spend less money and time travelling to work.

58) Two colleagues are talking about moving to a new office building. Listen to the dialogue and answer questions 58 - 60:

Áudio: audio6.wma

Choose the best alternative:

- a. The second speaker does not want to move to the new building.
- b. The second speaker is not convinced of the benefits of the new office.
- c. The first speaker thinks the old office building was perfect.
- d. Both the first and second speaker believe the new office will be more comfortable.

59) Which advantage of the *new building* is mentioned in the conversation?

Áudio: audio6.wma

- a. It has a lot of space for all the staff.
- b. It has a good bus service.
- c. It is near restaurants and shops.
- d. It is near downtown.

60) What does the *second speaker* suggest at the end of the conversation?

Áudio: audio6.wma

- a. The employees should hire a company bus service.
- b. They should talk to the manager about their problems.
- c. They should go to work by car.
- d. The manager should look for a new location.

Gabarito

1	d	11	b	21	b	31	a	41	d	51	d
2	d	12	b	22	c	32	b	42	d	52	d
3	a	13	a	23	a	33	c	43	a	53	a
4	c	14	d	24	b	34	d	44	b	54	d
5	b	15	a	25	d	35	b	45	b	55	c
6	b	16	d	26	d	36	d	46	d	56	c
7	a	17	c	27	c	37	b	47	b	57	d
8	a	18	b	28	d	38	c	48	d	58	d
9	c	19	b	29	b	39	a	49	b	59	a
10	c	20	a	30	c	40	a	50	b	60	b

ANEXO 2

Respostas dos questionários III – Professoras

Experts Post-Test Questionnaire

This questionnaire is part of the validation process of the Placement Test (henceforth – PT) designed for the Curso Superior em Tecnologia – (henceforth CST). The aim of PT is to improve the academic quality and efficiency of the CST in order to establish its content validity. The questionnaire is entirely anonymous.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. PT meets the requirements of an English placement test for the needs of CST students.				X	
2. PT accurately represents the content of CST English Language Programme at the various levels.			X		
3. The grammatical points represent the structure content of the respective levels.			X		
4. The lexical points represent the vocabulary content of the respective levels.			X		
5. The reading comprehension items represent the level of understanding of reading texts of the respective levels.			X		
6. The topics are appropriate for the respective levels.			X		
7. The topics are interesting.			X		
8. The instructions are clear.				X	
9. The test reflects settings test-takers would come across in their studies.		X			
10. The test reflects settings and contexts in English speaking settings.			X		
11. The test reflects the setting where test-takers will be studying.			X		
12. There is sufficient variety in test types, appropriate for the respective levels.			X		
13. PT items generally target the appropriate English level.			X		
14. The kinds of activities are appropriate.			X		
15. The activity types cover variety of learning styles.			X		
16. The time given to the test-takers to take the test is appropriate.		X			
17. There is time flexibility to cater for test-takers' difference in time needs.				X	
18. Item selection criteria are appropriate.				X	
19. The scoring system is satisfactory.				X	
20. The test results are appropriately reported to the test-takers.				X	
21. The instructions help test-takers familiarize themselves with the test, the test format, the test layout and the test items.		X			
22. The computer screen is clear.		X			
23. The activity layout is clear.		X			
24. The visuals are appropriate.		X			
25. The students are given opportunities to control the test (Review items, going backwards and towards, etc.).		X			

Turn over

26. Tell what you did not like about the PT?

I GUESS THERE SHOULD BE A KIND OF "MOCK TEST" TO GET THE STUDENTS AWARE OF THE EXERCISES, SYSTEM, TIME. BY DOING SO MANY QUESTIONS THAT POP UP DURING THE EXAM WOULD BE SOLVED BEFOREHAND.

27. Tell what you liked most about PT (if anything)?

IT'S INCREDIBLE TO THINK THAT STUDENTS CAN BE EVALUATED THROUGH AN ONLINE TEST THAT TEACHERS CAN MANAGE TO OPEN AND CLOSE FROM TIME TO TIME. IT IS AN EXAM SIMILAR TO THE INTERNATIONAL ONES (TELP, FOR EXAMPLE).

28. Tell how PT can be improved?

I GUESS THE TEST SHOULD NOT DEPEND ON THE INTERNET, STUDENTS SHOULD BE ABLE TO GO BACK AND FORTH IN THEIR QUESTIONS AS WELL.

29. Do you have problems with students that aren't placed correctly?

AS I HAVE BEEN WORKING AT THE INSTITUTION FOR LESS THAN A YEAR, I HAVE LITTLE EXPERIENCE WITH THIS. BUT, AS FAR AS I'M CONCERNED, THE EXAMS I HAVE HELD SHOWED THAT THE SPEAKING ABILITIES OF THE STUDENTS DO NOT MATCH THEIR WRITING SKILLS,

30. Do you have problems with students that don't have to attend classes for one or several terms and they have to go back later? Which are the problems?

AT MANY TIMES.

BECAUSE OF MY LACK OF EXPERIENCE THERE ARE ONLY A FEW POINTS I'D LIKE TO SHARE.

STUDENTS THAT COME BACK TO CLASSES AFTER BEING DISMISSED GENERALLY HAVE THEIR LISTENING SKILLS COMPROMISED.

ANOTHER POINT I COULD NOTICE IS THAT THE STUDENT FEELS LIKE AN "OUTSIDER" IN THE CLASSROOM FOR A WHILE.

Thank you

for accepting to contribute to the content validation of PT.

Experts Post-Test Questionnaire

This questionnaire is part of the validation process of the Placement Test (henceforth - PT) designed for the Curso Superior em Tecnologia - (henceforth CST). The aim of PT is to improve the academic quality and efficiency of the CST in order to establish its content validity. The questionnaire is entirely anonymous.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. PT meets the requirements of an English placement test for the needs of CST students.				X	
2. PT accurately represents the content of CST English Language Programme at the various levels.		X			
3. The grammatical points represent the structure content of the respective levels.				X	
4. The lexical points represent the vocabulary content of the respective levels.				X	
5. The reading comprehension items represent the level of understanding of reading texts of the respective levels.				X	
6. The topics are appropriate for the respective levels.				X	
7. The topics are interesting.				X	
8. The instructions are clear.				X	
9. The test reflects settings test-takers would come across in their studies.					X
10. The test reflects settings and contexts in English speaking settings.				X	
11. The test reflects the setting where test-takers will be studying.		X			
12. There is sufficient variety in test types, appropriate for the respective levels.		X			
13. PT items generally target the appropriate English level.				X	
14. The kinds of activities are appropriate.					X
15. The activity types cover variety of learning styles.					X
16. The time given to the test-takers to take the test is appropriate.			X		
17. There is time flexibility to cater for test-takers' difference in time needs.					X
18. Item selection criteria are appropriate.					X
19. The scoring system is satisfactory.					X
20. The test results are appropriately reported to the test-takers.					X
21. The instructions help test-takers familiarize themselves with the test, the test format, the test layout and the test items.					X
22. The computer screen is clear.			X		
23. The activity layout is clear.				X	
24. The visuals are appropriate.				X	
25. The students are given opportunities to control the test (Review items, going backwards and towards, etc.).	X			X	

Turn over

26. Tell what you did not like about the PT?

The fact that it focuses too much on business English, but teaching business English is not a generally accepted position in the States.

27. Tell what you liked most about PT (if anything)?

The grammatical structure of the different levels is appropriate.

28. Tell how PT can be improved?

By including more general English and by making all students take the test (another edition) again after one year and a half.

29. Do you have problems with students that aren't placed correctly?

A few students claim that they didn't take the test seriously and were placed in the wrong level. This is not a general complaint.

30. Do you have problems with students that don't have to attend classes for one or several terms and they have to go back later? Which are the problems?

I've had problems with students who have been able to skip one or two semesters at most. When they come back they sometimes disturb the "flowing" of the class, because they are not used to it or because they have problems in their oral performance. With students who have been dismissed for longer periods, I have never had problems.

Thank you

for accepting to contribute to the content validation of PT.

